Single-stranded DNA Scanning and Deamination by APOBEC3G Cytidine Deaminase at Single Molecule Resolution*5•

Received for publication, January 17, 2012, and in revised form, February 22, 2012 Published, JBC Papers in Press, February 23, 2012, DOI 10.1074/jbc.M112.342790

Gayan Senavirathne§1, Malgorzata Jaszczur§1, Paul A. Auerbach§1, Thomas G. Upton§, Linda Chelico§2, Myron F. Goodman§3 and David Rueda§4

From the §Department of Chemistry, Wayne State University, Detroit, Michigan 48202 and the •Departments of Biological Sciences and Chemistry, University of Southern California, Los Angeles, California 90089

Background: Apo3G, an ssDNA-dependent C deaminase, inactivates HIV-1 in T cells by C to T hypermutation.

Results: smFRET is used to detect Apo3G scanning and C-deamination on ssDNA.

Conclusion: Apo3G scans ssDNA randomly and bidirectionally, favoring nonrandom 3′ to 5′ deamination.

Significance: This smFRET study describes a broadly applicable approach to visualize motion and catalysis in real time by an enzyme that scans ssDNA.

APOBEC3G (Apo3G) is a single-stranded (ss)DNA cytosine deaminase that eliminates HIV-1 infectivity by converting C → U in numerous small target motifs on the minus viral cDNA. Apo3G deaminates linear ssDNA in vitro with pronounced spatial asymmetry favoring the 3′ → 5′ direction. A similar polarity observed in vivo is believed responsible for initiating localized C → T mutational gradients that inactivate the virus. When compared with double-stranded (ds)DNA scanning enzymes, e.g. DNA glycosylases that excise rare aberrant bases, there is a paucity of mechanistic studies on ssDNA scanning enzymes. Here, we investigate ssDNA scanning and motif-targeting mechanisms for Apo3G using single molecule Förster resonance energy transfer. We address the specific issue of deamination asymmetry within the general context of ssDNA scanning mechanisms and show that Apo3G scanning trajectories, ssDNA contraction, and deamination efficiencies depend on motif sequence, location, and ionic strength. Notably, we observe the presence of bidirectional quasi-localized scanning of Apo3G occurring proximal to a 5′ hot motif, a motif-dependent DNA contraction greatest for 5′ hot > 3′ hot > 5′ cold motifs, and diminished mobility at low salt. We discuss the single molecule Förster resonance energy transfer data in terms of a model in which deamination polarity occurs as a consequence of Apo3G binding to ssDNA in two orientations, one that is catalytically favorable, with the other unfavorable.

Incidental deamination of dC → dU occurs frequently, especially on single-stranded (ss)DNA (1), which is a potential source of spontaneous C → T mutations. The elimination of U-G mismatches by base excision repair ensures that deamination-initiated mutations are minimized in genomic DNA (1, 2). However, dC deaminations also occur enzymatically, as a regulated function of the immune system (3, 4). For the HIV-1 host restriction factor, APOBEC3G (Apo3G), this entails deaminating C residues on newly reverse transcribed HIV-1 cDNA (minus strand DNA), most often in 5′-CCC motifs, mainly at 3′-C (underlined), although occasionally at the middle C in the motif (5, 6). The virus can be neutralized by the catalytic action of Apo3G in at least two ways. The presence of U might induce degradation of the HIV-1 minus strand by concerted action of uracil DNA glycosylase and apurinic/apyrimidinic endonuclease (2, 7, 8), or G → A mutations that occur following synthesis of the plus strand may destroy viral infectivity (5, 6, 9). In the event that Apo3G does not induce sufficient mutagenesis to result in HIV-1 inactivation, it may contribute to HIV-1 evolution and development of drug-resistant quasi-species (10). It is important to establish the biochemical mechanisms of how Apo3G induces multiple mutations in the HIV-1 genome per se, and also to support ongoing efforts to develop Apo3G-based HIV therapies (11).

Our objective is to examine how Apo3G locates and then deaminates C motifs on an ssDNA substrate. Apo3G contains two deaminase domains. The N-terminal domain (CD1) is not catalytically active, but it is required for HIV-1 virion encapsidation and can bind DNA and RNA (12, 13). The C-terminal domain (CD2) contains the active site (14). Previous “bulk” biochemical studies showed that Apo3G catalyzes processive deaminations on linear ssDNA prior to acting on another substrate molecule while displaying a pronounced catalytic polarity (15, 16). Trinucleotide motifs are deaminated with increased

© 2012 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.
efficiency when located nearer to the 5'-end of the DNA, and the terminal 33 nt at the 3'-end of an ssDNA substrate are barely deaminated, creating a deamination “dead zone” (16). The current model for the Apo3G scanning mechanism that has resulted from bulk biochemical studies has yet been unable to answer several questions regarding the scanning mechanism. These questions are addressed here using single molecule FRET (smFRET). From a biological perspective, it has been suggested that Apo3G deamination polarity offers a likely explanation for the presence of localized mutational gradients in the HIV-1 genome that appear to be important in viral inactivation (17).

Using smFRET, we visualize Apo3G movement and catalysis on linear ssDNA in real time. We investigate how motif location and identity, i.e. “hot” versus “cold” motifs, influence Apo3G motion, C deamination efficiency, and DNA contraction. The smFRET data provide new insights into a model in which Apo3G is presumed to scan ssDNA in a symmetric, bidirectional manner, yet causes spatially polar deaminations by binding asymmetrically to ssDNA in two orientations, one that is catalytically active and another that is essentially inactive (18), with pseudo-localized scanning in the vicinity of a 5’ hot motif.

EXPERIMENTAL PROCEDURES

Apo3G Labeling with Cy5 Fluorescent Dye—To remove surface-exposed cysteine residues (confirmed by bioinformatic structural modeling of Apo3G), we used a QuikChange site-directed mutagenesis protocol to mutate cysteine 139, 243, and 308 to leucine, alanine, and leucine, respectively. Only one surface-exposed residue (Cys-356) is available for labeling with Cy5-maleimide. A baculovirus-expressed mutated GST-Apo3G protein variant was prepared as described previously (15, 16). The Cy5 maleimide mono-reactive dye kit (GE Healthcare) protocol was used to achieve 10–20% labeling efficiency, to prevent nonspecific labeling of protein side chains. Free label was removed by gel filtration on Bio-Gel P-6 (Bio-Rad). Protein concentration was determined by Bradford assay and BSA protein standardization. Labeling efficiency was determined by UV (280 and 552 nm). Mass spectrometry analysis of Cy5-Apo3G confirmed that the predicted surface-exposed Cys-356 was labeled (Harvard University Mass Spectrometry and Proteomics Core). We have further verified by photobleaching that Apo3G has a single Cy5-label. Cy5-Apo3G retained activity and deamination properties of native Apo3G. We have also verified, based on previous atomic force microscopy data (16, 18), that Apo3G is predominantly monomer under smFRET conditions.

Pfu DNA Polymerase Expression, Purification, and Labeling—Two rounds of site-directed mutagenesis (QuikChange site-directed mutagenesis kit, Stratagene) of Pfu DNA polymerase expression vector pET-30a-PFU (Dr. Stephen Bell, Cambridge, UK) were performed (D125A, I48C) to make an exo-Pfu suitable for labeling. The resulting vector (pPFUE-I48C) was transformed into BL21 Rosetta(DE3) pLysS (Novagen). Cells were grown at 37 °C until an A590 of ~0.5 was reached. Isopropyl-1-thio-β-D-galactopyranoside was added (1 mM final concentration), and cells were grown at room temperature (~25 °C) overnight and then centrifuged at 4 °C (4,000 rpm for 15 min) and resuspended in 25 mM HEPES, pH 7.8, 500 mM NaCl, 5 mM imidazole, 1 mM phenylmethanesulfonyl fluoride, protease inhibitor mixture, 50 mg/ml lysozyme. After sonication on ice (10 × 15-s pulses), 20 units of DNase I and 10 μg/ml RNase A were added, and cells were incubated at 37 °C for 30 min and then boiled for 15 min. Samples were centrifuged twice (17,000 × g for 15 min, 4 °C), the supernatant was loaded onto a 5-ml nickel column (nickel-nitrioltriacetic acid His-Label resin, Novagen) and washed four times with 25 mM HEPES, pH 7.8, 500 mM NaCl, 5 mM imidazole. Protein was eluted with 25 mM HEPES, pH 7.8, 500 mM NaCl, 200 mM imidazole, and fractions were analyzed by SDS-PAGE and Coomassie Blue staining. Protein concentration of fractions (~90 kDa) was determined by near-UV absorbance (A280). Pfu D125A/I48C was labeled using the Cy5 maleimide mono-reactive dye kit (GE Healthcare) according to provided protocols.

smFRET—Quartz microscope slides were passivated with methoxy-PEG (Mw = 5,000; Laysan Bio Inc.) and 1% biotin-PEG (Mw = 3,400; Laysan Bio Inc.) to minimize nonspecific binding (19, 20), and ssDNAs were surface-immobilized by using a common anchor DNA (supplemental Table S1) with 5’-Cy3 and 3’-biotin (see Fig. 2A). ssDNAs (2 μM) were annealed to the anchor DNA (1 μM) in standard buffer (10 μl, 50 mM MOPS, pH 7.4, 5 mM MgCl2, 60 mM NaCl, 2% (v/v) 2-mercaptoethanol, and 2 μM Trolox) by heating to 90 °C for 45 s and cooled to room temperature over 15 min. Solution was serially diluted for surface immobilization (25–50 pm) of partial double-stranded DNA in standard buffer. Excess DNA was washed out, Cy5-labeled Apo3G (1 nm) was introduced in a standard buffer, immediately followed by data acquisition on a home-built prism-based total internal reflection single molecule fluorescence microscope at 1-s or 30-ms time resolution (21). Movies were recorded for ~10 min in five different areas of the reaction channel.

Hidden Markov Model and TDP Analysis—Long binding (from 25 s to 10 min), single molecule FRET time trajectories were analyzed using a hidden Markov model, as described (22). Transitions were divided into 10 virtual macrocanonical states separated by 0.1 FRET units, which do not correspond to specific conformational states, but rather to an ensemble of states with the protein located at a certain distance from the 5’-end. FRET trajectories were compiled into transition density plots (TDPs) showing the number of transitions observed between a given initial and final FRET value. All initial binding events originate from zero FRET and are located on the y axis.

Exponential Extension Length Analysis—The excursion distance is calculated using the initial binding FRET value (~0.2) and Förster’s equation $E_{\text{FRET}} = 1/[1 + (R/R_0)^6]$, where $R$ is the observed distance and $R_0$ is the Förster distance corresponding to 50% energy transfer (55 Å for Cy3, Cy5, assuming $k^2 = 2/3$). Final FRET values in the TDP upper diagonal represent the location of the protein after each transition, which is converted into a distance using Förster’s equation. The TDP provides the fraction of transitions at this distance. The fraction of transitions is plotted as a function of the distance to the initial binding site to estimate the distance scanned by the protein, which decays exponentially, indicating random excursions. Because we are calculating differences in distances, the $x^2 = 2/3$ approximation only minimally affects our results.
Post-synchronization Analysis—To synchronize the long binding events with respect to their initial bindings, we analyzed the long binding FRET trajectories from 10 s before the initial binding (threshold 0.15 FRET) until 10 s after the dissociation (0 FRET). Using a MATLAB script kindly provided by Jody Puglisy (Stanford University), we binned the trajectories (0.05 FRET bins, 3-s time bins) to determine how many trajectories are located at a given FRET and time value and then generate the histograms.

smFRET Deamination Assay—DNA substrates were surface-immobilized as described above using deamination buffer (50 mM MOPS, pH 7.4, 125 mM NaCl). Cy5-labeled Pfu exo− (50 nM) and unlabeled Apo3G (1 nM) were introduced on to the surface-immobilized substrate DNA immediately followed by data acquisition with 1-s time resolution (21) for 10 min each. For each substrate tested, we analyzed 100-200 single molecule trajectories (N), and for each substrate, we found the number of DNA molecules (n) showing any specific Cy5-labeled Pfu exo− binding (t_{on} \approx 1 s), and the bound percentage was calculated as n/N \times 100%. A small background (~10%) of nonspecific binding (measured in the absence of Apo3G) was subtracted for each substrate.

smFRET DNA Conformational Dynamics Experiments—To study the Apo3G-induced ssDNA conformational dynamics, we immobilized ssDNAs with 3′-Cy5- to the 5′-Cy3-labeled anchor DNA as described above. Data acquisition from five different areas of the reaction channel was done in the absence and presence of Apo3G (1 nM) with 1-s time resolution (21) for 10 min each. The smFRET histograms were obtained by time-binning <100 time trajectories and fit to a Gaussian distribution, y = A \exp \left\{ -\frac{(FRET - FRET_0)^2}{2\sigma^2} \right\}

using Igor (WaveMetrics, Lake Oswego, OR).

RESULTS

Processive Apo3G-catalyzed C Deamination on ssDNA—Apo3G was incubated with phage M13mp2 circular DNA containing a series of in-frame 5′-aaaCCaaa hot motifs embedded in lacZα reporter sequence located within a single-stranded gapped region of M13 dsDNA (Fig. 1A). Following transfection of the DNA in Ung− Escherichia coli, Apo3G-catalyzed deaminations are identified as C \rightarrow T mutations in DNA isolated and sequenced from individual mutant phage clones (23). M13 mutant phage (white plaques) comprise \approx 2% of the total plaques, mutant and wild type (blue plaques), so that virtually all of the individual DNA clones were deaminated by at most one Apo3G molecule, in accord with Poisson statistics (24, 25). The mutations occur as singletons and in clusters containing 2-5 consecutively deaminated motifs, with deaminations observed predominantly at C in 5′-aaaCaaa (Fig. 1B). The data suggest that a single Apo3G molecule scans ssDNA pro-
cessively, deaminating C → U haphazardly, reminiscent of the processive stochastic deamination patterns observed for activation-induced deoxycytidine deaminase (23), which is also an Apobec family protein (3, 4). We have purified native Apo3G to apparent homogeneity and have previously obtained a high-resolution x-ray structure for the catalytic CD2 domain (14). We now examine the scanning behavior of native full-length Apo3G at single molecule resolution using smFRET.

Apo3G ssDNA Scanning Observed by smFRET—We investigated the scanning behavior of Apo3G on a surface-immobilized ssDNA labeled at the 5'-end with Cy3, containing an ata-CCCaaa hot motif (15) located near the 5'-end (pdT 5' hot) and used a Cy5-labeled Apo3G to scan the DNA (Fig. 2A, supplemental Table S1). Based on previous atomic force microscopy data (16, 18), we have verified that Apo3G is predominantly a monomer under the smFRET conditions. The smFRET trajectories (Fig. 2B) show Apo3G binding and scanning motion as changes in apparent FRET efficiency, binding as an abrupt increase from zero, and dissociation as a sharp decrease to zero. Motions toward 5' and 3' directions are observed as increases and decreases in FRET, respectively. Representative traces, including anticorrelated FRET trajectories, are presented as supplemental material (supplemental Fig. S1).

The FRET trajectories reveal two populations, consistent with bulk experiments (16). About half show short Apo3G binding, <25 s (FRET ~0.2), with apparent binding and dissociation pseudo-first order rate constants $k_{on} = 0.01 \pm 0.01 \text{ s}^{-1}$ and $k_{off} = 0.23 \pm 0.04 \text{ s}^{-1}$, respectively (supplemental Fig. S2). Long binding trajectories (≥25 s) exhibit more complex dynamics indicating rapid scanning (Fig. 2B, Scanning trajectories). The FRET efficiencies oscillate rapidly between ~0.2 and 0.8 for up to 10 min, with an average bound time of 110 s (Fig. 2B, Scanning trajectories; supplemental Fig. S3; supplemental Table S2). Because specific FRET states could not be identified, a hidden Markov Model (HMM) (22) was used to determine FRET densities corresponding to the locations of Apo3G relative to the 5'-end. Thus, each “FRET state” corresponds to an ensemble of configurations characterized by the distance between the protein and the 5'-end. The HMM is used to generate a TDP showing the number of transitions observed between initial and final FRET states (Fig. 2B, TDP). All initial binding events stem from zero FRET on the y axis, with most occurring at ~0.2 FRET, indicating that Apo3G binds preferentially away from the tethered 5'-end, confirmed by switching the tethered DNA to the 3'-end, in which case favored binding still occurs near the free end (i.e. 5'-end) (supplemental Fig. S4). Movements toward the 5'- and 3'-ends appear as peaks above and below the diagonal, respectively. The transitions are symmetric in both directions (Fig. 2B, TDP), indicating that Apo3G scans ssDNA without directional preference. Analysis of the dwell times between FRET states shows that scanning transitions occur with similar rate constants ~1 s⁻¹, measured at 33 frames/s (supplemental Fig. S5; supplemental Table S3).

An estimate of scanning distance in either direction is obtained using the number of transitions in the TDP and Förster’s equation. The number of transitions decays exponentially from the initial binding site, consistent with random excursions (Fig. 2B, Scanning lengths) with a half-distance $l = 12 \text{ Å} (~9 \text{ nt})$ (26). To eliminate data blurring from asynchronous binding, long binding trajectories were post-synchronized by alignment at the initial binding event (27). The post-synchronization histogram (PSH) shows an initial FRET increase to ~0.2 (Fig. 2B, PSH), consistent with the initial binding observed in the TDP. The FRET ratio increases to ~0.4 followed by oscillations between 0.2 and 0.7. Therefore, Apo3G moves in a bidirectional random manner yet hovers in the vicinity of the 5' hot motif, as shown by trajectories that remain synchronized up to 150 s (Fig. 2B, PSH).

Moving the hot motif closer to the 3'-end shows that scanning is strongly influenced by motif location, where Apo3G goes to high FRET values (>0.4) less often for the 3' hot motif (Fig. 2C). The number of transitions from the initial binding site decays exponentially with a half-distance $l = 7 \text{ Å} (~5$-nt excursions) (Fig. 2C, Scanning lengths), about half when compared with the 5' hot motif. The post-synchronization histogram for the 3' hot trajectories confirms initial binding at ~0.2 FRET followed by oscillations between 0.2 and 0.6 FRET, but with more transitions in the 0.2–0.4 range and fewer in the 0.4–0.6 range (Fig. 2C, PSH). The trajectories remain synchronized up to 50 s (Fig. 2C, PSH) when compared with ~150 s for the 5' hot motif.

The local sequence surrounding the target CCC influences scanning as shown by replacing the 5' hot (ataCCCaaa) motif with a 5' cold (tttCCCaCttt) motif (15) (Fig. 2D). The initial binding nearer the 3'-end (~0.2 FRET), excess transitions toward the 5'-end (FRET>0.4), and excursion half-distance 11 Å (Fig. 2D, Scanning lengths) are similar to the 5' hot motif. However, the trajectories remain synchronized for only about half as long, 75 s when compared with 150 s for the 5' hot motif (Fig. 2D, PSH), suggesting that Apo3G hovers in the vicinity of the 5' hot motif (~0.4 FRET), about twice as long when compared with the 5' cold motif, because for the cold motif, much more of the scanning occurs at lower FRET states (0.2–0.3) away from the target motif (Fig. 2D, PSH). Scanning toward the 5'-end (FRET>0.4) remains when the hot 5'-CCC is replaced by the 5'-CCU deaminated product, and the scanning excursion (9 Å) is intermediate (Fig. 2E, Scanning lengths). However, the synchronized trajectories, reflecting hovering, near the 5’-CCU are 75 s (Fig. 2E, PSH), which is half as long as the 5' hot motif, similar to the 5' cold motif. In the absence of a deamination target (poly(dT), Fig. 2F), the scanning is similar to the 3' hot DNA (pdT 3' hot) with respect to traces, TDP, and displacement half-distance from the initial binding site.

Notably, however, the histogram for poly(dT) (pdT) shows that there is no longer a discernible spatial localization of Apo3G on the DNA (Fig. 2F, PSH). In other words, Apo3G hovering is absent when there is no deamination motif. The long binders exhibit similar average residence times for each of the constructs (supplemental Table S2). In summary, the data show that Apo3G scans ssDNA bidirectionally over the entire molecule, favoring movement in the vicinity of a 5' hot motif.

Apo3G Contracts ssDNA in a Deamination Motif-dependent Manner—With fluorescent probes located at 5'- and 3'-ends, we can use smFRET to detect DNA contraction by measuring end-to-end distances in the presence of unlabeled Apo3G (Fig. 3A). In the absence of Apo3G, the histogram for pdT 5' hot has...
a narrow distribution centered at 0 (Fig. 3B), indicating that the DNA is in an extended conformation with its ends separated by >90 Å. In the presence of Apo3G, the FRET distribution shifts to 0.5 and broadens (0.2–0.8), indicative of contraction (Fig. 3C). The broad distribution indicates the presence of rapid conformational dynamics exceeding our time resolution (1 s) (28). Pronounced differences are observed for the pdT 3′ hot, where the FRET distribution is centered at 0.4 and the width narrows (0.2–0.6) (Fig. 3D). A further shift to 0.3 and narrowing (0.1–0.4) occurs with pdT 5′ cold (Fig. 3E), 5′-CCU product (Fig. 3F),
and pdT (Fig. 3G). These data show that Apo3G contracts ssDNA in a motif-dependent manner. Contraction is most pronounced during long excursions from the initial binding site in the vicinity of a 5'-hot motif (Fig. 2).

**Distinguishing between Apo3G Scanning and ssDNA Contraction**—The FRET fluctuations shown in Fig. 2, with the FRET donor situated at the 5'-tethered end of the ssDNA, are interpreted as Apo3G scanning the ssDNA. However, a possible alternative explanation for these fluctuations could be that Apo3G binds at a fixed position on the DNA and causes DNA conformational changes, e.g., by contracting the DNA, which would also result in distance changes between the FRET donor on the 5'-ssDNA end and acceptor on Apo3G, which binds preferentially near the un tethered 3'-end (Fig. 2).

To establish that Apo3G does not remain bound at a fixed position, but instead moves along the ssDNA, we relocated the FRET donor to the 3'-end of the ssDNA substrate (Fig. 4A). If Apo3G were to bind nearer the 3'-tail and contract the ssDNA without scanning, then the FRET ratio should remain approximately constant. However, the observed smFRET trajectories and TDP analysis for the pdT 5'-hot ssDNA (Fig. 4B) show rapid, bidirectional FRET changes (in the range of FRET 0–1), independently indicating that Apo3G moves along the entire ssDNA.

Another important distinction between Apo3G ssDNA scanning and contraction was obtained by repeating smFRET scanning measurements using poly(dA) with a 5'-hot motif (Fig. 5, pdA 5'-hot). Because of significantly greater base stacking, pdA is considerably stiffer than pdT and therefore should decrease contraction while maintaining scanning. Indeed, when unlabeled Apo3G binds pdA 5'-hot that is labeled on each end (Fig. 5A), ssDNA contraction is significantly reduced (FRET ~0.2). Fig. 5B (Scanning trajectories) shows representative FRET time trajectories of Cy5-labeled Apo3G moving along the pdA 5'-hot ssDNA. Similar FRET fluctuations are observed as on pdT 5'-hot (Fig. 2B), indicating that Apo3G also scans the stiffer ssDNA (Fig. 5B). The resulting TDP shows symmetric FRET oscillations between ~0.2 and ~0.8, indicating that Apo3G also scans poly(dA) 5'-hot randomly and bidirectionally (Fig. 5B).

Once again, to confirm that Apo3G scans the DNA and does not gain access to different DNA regions by contracting the DNA while remaining bound at a fixed position, the Cy3 donor label was placed was the 3'-end of ssDNA. Here again, the FRET trajectories and TDP analysis show that Apo3G randomly and bidirectionally scans the entire ssDNA (FRET 0–1; Fig. 5C). These data confirm that the FRET fluctuations observed in the time trajectories reflect the scanning behavior of the enzyme on the DNA for the less flexible pdA 5'-hot (Fig. 5B) as it did for the more flexible pdT 5'-hot (Fig. 2B).

**Reduced Apo3G Mobility on ssDNA at Low Salt**—The catalytically inactive N-terminal CD1 domain has a predicted large net positive charge (+11), in contrast to the catalytically active CD2 domain (~4.5), and is likely to govern the mobility of Apo3G on ssDNA, which should depend on metal ion concentration. When the experiments for 5'-hot and poly(dt) DNA were repeated in "low" salt conditions (30 mM NaCl, 0 mM MgCl2), when compared with "high" salt (60 mM NaCl, 5 mM MgCl2), there was a marked reduction in scanning speed, accompanied by lengthy pauses in the FRET trajectories (Fig. 6). Although bidirectional motion still occurs over the entire ssDNA at low salt, there are far more transitions confined to low FRET values near the initial binding site (Fig. 6, A and B,
Single Molecule FRET Studies of Apo3G Scanning ssDNA

FIGURE 4. Apo3G binds and scans the entire ssDNA. A, single molecule FRET setup for analyzing scanning of Apo3G on ssDNA relative to the free 3′-end. 3′-Cy3-labeled (reverse Fl-labeled) ssDNA (92 nt) is annealed to a surface-immobilized anchor DNA. Binding and scanning of Cy5-labeled (A = acceptor) Apo3G result in FRET changes. B, representative smFRET scanning trajectory shows the FRET ratios (y axis, black) with the HMM fit (red) over a given time (x axis). C, the TDP was prepared by HMM analysis of the long binding FRET transitions (n = 66); the intensity bar relates to the frequency of FRET transitions at specific states (blue, low frequency; yellow, high frequency). Scanning trajectories and TDP show that Apo3G scans the entire DNA and does not remain bound to the 3′-end.

FIGURE 5. Apo3G scans stiffer ssDNA with reduced contraction. Apo3G-induced ssDNA contraction for pdA 5′-hot DNA was monitored using a setup similar to Fig. 3A. A, single molecule FRET histograms in the presence (panel 1) and absence (panel 2) of Apo3G reveal less protein-induced DNA contraction as a small change in the center of the FRET distribution and width. B, representative long binding smFRET time trajectory and TDP for pdA 5′-hot DNA scanning (n = 68) show similar scanning to the pdT 5′ hot (setup as in Fig. 2A). C, representative long binding smFRET time trajectory and TDP for scanning 3′-Cy3-labeled (reverse Fl-labeled) pdA 5′-hot DNA (n = 50) show that Apo3G scans this stiffer DNA and does not remain bound to the 3′-end.

TDP, and the average excursion distance for the pdT 5′ hot is reduced to 8 Å when compared with 12 Å at high salt (Figs. 2B and 6A, Scanning lengths). The scanning transition rate constants are reduced 10–100-fold at low salt (supplemental Table S4), and the localized motions of Apo3G near the 5′ hot motif at high salt (Fig. 2B, PSH) are absent at low salt (Fig. 6A, PSH). The same restricted motion is observed for the pdT substrate, with a 6 Å excursion distance (Fig. 6B, Scanning lengths). Low salt also diminishes DNA contracting, as shown in histograms having narrower distributions centered at lower FRET values for the pdT 5′ hot (Fig. 6C). These nonphysiological salt concentrations were chosen to illustrate changes in Apo3G scanning mobility. The large increase in the mobility of Apo3G on ssDNA at higher salt is likely caused by a partial shielding of the electrostatic interactions between the strongly positively charged CD1 domain and the negatively charged DNA phosphate backbone, possibly augmented by an increased DNA base stacking (29) that could further facilitate Apo3G movement by providing a more ordered ssDNA backbone.

Observing Apo3G C Deamination Polarity with smFRET—Apo3G exhibits a marked deamination polarity favoring the 5′ direction on linear ssDNA measured in bulk solution (15, 16) (Fig. 7A). Both native and Cy5-labeled forms of Apo3G favor deamination of the 5′-CCC motif located toward the 5′-end of the ssDNA substrate by about 2-fold when compared with 5′-CCC situated nearer the 3′-end. Both enzymes also show correlated double deaminations, indicative of enzyme processivity (15, 16), as also illustrated by the clustered deamination patterns observed in Fig. 1B. In contrast, the C-terminal catalytically active CD2 domain is neither catalytically polar nor processive in the absence of the catalytically inactive CD1 domain (18) (Fig. 7A).

Native Apo3G polarity can be explored directly by smFRET using Cy5-labeled Pfu DNA polymerase to bind to U on ssDNA (30, 31) following Apo3G conversion of 5′-CCC → CCU (Fig. 7B). In the presence of Apo3G, Cy5-Pfu binds 18% of 5′-hot DNA (Fig. 7C, panel 1; FRET 0.4; Fig. 7D). No measurable binding of Cy5-Pfu occurs in the absence of Apo3G (Fig. 7C, panel 3) or U (supplemental Fig. S6). Placing the hot motif toward the 3′-end results in a 2-fold reduction in the fraction of DNA bound by Cy5-Pfu (8%) accompanied by a lower FRET increase (Fig. 7C, panel 2; FRET 0.3; Fig. 7D). Therefore, Apo3G deaminates the 5′ hot motif more efficiently than the 3′ hot motif, as observed in bulk experiments (15) (Fig. 7A). Replacing 5′ hot with a 5′ cold motif results in a 3.6-fold decrease in the fraction of DNA bound by Cy5-Pfu (5%). Placing the hot motif 18 nt from the 3′-end within a 33-nt dead zone (16) results in marginal DNA binding by Cy5-Pfu (2%, Fig. 7D, “dead”). Therefore, Apo3G catalyzes asymmetric deamination favoring the 3′ → 5′ direction. It is important to point out that smFRET dynamics (Fig. 2) could depend on prior- and post-reaction Apo3G target interactions.

DISCUSSION

Apolobec family proteins are classified in two distinct groups, having either two deaminase domains, such as Apo3G, or...
one deaminase domain, e.g., activation-induced deoxycytidine deaminase (3, 4). When present as a monomer, Apo3G is composed of a catalytically active C-terminal domain, CD2, with a negative charge (\(\sim H11002 4.5\)), and a catalytically inactive CD1 domain, with a predicted large positive charge (\(\sim H11001 11\)), needed for DNA and RNA binding (12, 13, 18). A model to explain deamination polarity suggests that Apo3G binds ssDNA in either an active or an inactive orientation, with equal probability. The active orientation occurs when CD2 faces the 5’-ssDNA end, whereas the inactive orientation has CD2 facing the 3’-end (18) (Fig. 8).

The smFRET data for the ssDNA substrate speak directly to the asymmetric deamination model. The C deamination efficiency is about 2.3-fold higher for the 5’ hot motif when compared with the 3’ hot motif (Fig. 7D). The TDPs are symmetrical (Fig. 2), showing that Apo3G scans along ssDNA with equal probability toward 5’ and 3’ directions. Therefore, the deamination polarity cannot be caused by motion favoring the 5’ direction. Instead, the likely source of the 3’→5’ deamination polarity suggests that Apo3G binds ssDNA in either an active or an inactive orientation, with equal probability. The active orientation occurs when CD2 faces the 5’-ssDNA end, whereas the inactive orientation has CD2 facing the 3’-end (18) (Fig. 8).

The smFRET data for the ssDNA substrate speak directly to the asymmetric deamination model. The C deamination efficiency is about 2.3-fold higher for the 5’ hot motif when compared with the 3’ hot motif (Fig. 7D). The TDPs are symmetrical (Fig. 2), showing that Apo3G scans along ssDNA with equal probability toward 5’ and 3’ directions. Therefore, the deamination polarity cannot be caused by motion favoring the 5’ direction. Instead, the likely source of the 3’→5’ polarity is that Apo3G is active mainly when bound with CD2 facing the 5’-end of the ssDNA (see e.g. Fig. 8). The presence of a dead zone, a 33-nt region at the 3’-end of the DNA in which deaminations are barely detectable (16) (Fig. 7D), would occur whenever catalytically inactive CD1 binds at the 3’-end, thereby precluding access by CD2. Binding in the opposite direction, with CD1 facing the 5’-end, would allow unrestricted access of CD2 to the 3’-end (dead zone), but in a catalytically inactive orientation (Fig. 8). It is the 3’-C of the 5’-CCC-3’ motif that is almost always deaminated, whereas the 5’-C is never deaminated (6, 16). Based on an x-ray structural analysis of CD2 (14), it was proposed that the 3’-C is properly positioned nearby the catalytic zinc hydroxylation site only when CD2 is facing the 3’-C of the 5’-CCC motif (see the legend for Fig. 8), in agreement with the asymmetrical deamination model derived from the biochemical data (18).

The catalytic asymmetry is likely a consequence of the double domain structure of Apo3G. In the absence of CD1, deaminations catalyzed by CD2 have no dead zone (18) and are nonpolar when acting on linear ssDNA (18) (Fig. 7A). Activation-induced deoxycytidine deaminase, which has just a single catalytically active domain, also has no dead zone and catalyzes deaminations with equal efficiencies in 5’ and 3’ directions (16) (supplemental Fig. S7). Notably, it is the use of a linear ssDNA construct that imposes the type of end constraint that facilitates identification of dual catalytic orientations, active and inactive. As predicted, there was no polarity observed on circular ssDNA (16). However, annealing a complementary DNA to a circle restored the polarity (16), which may explain the presence of localized regions showing 3’→5’ deamination polarity that are observed on the HIV-1 cDNA during reverse transcription in vivo (17), where cDNA synthesis and RNA template degradation occur concurrently, leaving ssDNA for Apo3G to act on proximal to multiple RNA/DNA hybrid regions.

There is a paucity of dynamic data for enzymes that scan ssDNA. The smFRET data provide an initial picture describing ssDNA scanning by Apo3G. The magnitude of the temporal

FIGURE 6. Low salt restricts Apo3G scanning. A and B, representative low salt long binding smFRET time trajectories, TDP, scanning length analysis, and PST, for 5’ hot DNA and poly(dT) DNA (n = 69 and 67, respectively), as indicated. Experimental set up is as in Fig. 2A. Decreasing the salt concentration increases the number and duration of pauses between transitions and reduces the amount of high FRET transitions (TDP), the average scanning length (± error from the fitting), and the pseudo-localized motion, especially for 5’ hot DNA, indicating that low salt slows scanning down. C, single molecule FRET histograms reveal less ssDNA contraction in low salt, but still in a motif-dependent manner.

The catalytic asymmetry is likely a consequence of the double domain structure of Apo3G. In the absence of CD1, deaminations catalyzed by CD2 have no dead zone (18) and are nonpolar when acting on linear ssDNA (18) (Fig. 7A). Activation-induced deoxycytidine deaminase, which has just a single catalytically active domain, also has no dead zone and catalyzes deaminations with equal efficiencies in 5’ and 3’ directions (16) (supplemental Fig. S7). Notably, it is the use of a linear ssDNA construct that imposes the type of end constraint that facilitates identification of dual catalytic orientations, active and inactive. As predicted, there was no polarity observed on circular ssDNA (16). However, annealing a complementary DNA to a circle restored the polarity (16), which may explain the presence of localized regions showing 3’→5’ deamination polarity that are observed on the HIV-1 cDNA during reverse transcription in vivo (17), where cDNA synthesis and RNA template degradation occur concurrently, leaving ssDNA for Apo3G to act on proximal to multiple RNA/DNA hybrid regions.

There is a paucity of dynamic data for enzymes that scan ssDNA. The smFRET data provide an initial picture describing ssDNA scanning by Apo3G. The magnitude of the temporal
indicated. Each binding event results in a FRET increase (time trajectories of Cy5-fragment. A 72-nt-long ssDNA with two 5'-ataCCaaa target motifs by unlabeled and Cy5-labeled Apo3G and Apo3G-CY-C. DNA is internally labeled with fluorescein (F). The Apo3G to ssDNA ratio is 1:20 for Apo3G and Apo3G-CY and 1:1 for Apo3G-CY-C. A 5'-deamination results in a 54-nt fragment, a 5'-deamination results in a 40-nt fragment, and double deaminations result in a 22-nt fragment. S' C'/C' ratios are indicated below each lane. The deaminase assay is carried out under the previously described single hit conditions (15); no more than 15% of substrate is used. B, single molecule FRET setup for detection of Apo3G-catalyzed deamination for Cy3-labeled (D = donor) ssDNA and Cy5-labeled (A = acceptor) Pfu exo − polymerase. C, representative smFRET time trajectories of Cy5-Pfu binding to uracil on various ssDNA substrates, as indicated. Each binding event results in a FRET increase (arrows). D, background-corrected percentages of Pfu bound DNA with different deamination motifs: 5' hot (18 ± 3%, n = 281 molecules), 5' cold (8 ± 1%, n = 148 molecules), and "dead" (2 ± 2%, n = 243 molecules). Error bars (±) were calculated based on n, as 100/(n)1/2%. The data show that a 5' hot motif is deaminated preferentially, a hot motif is favored for deamination over a cold motif, and a region near the 3'-end containing a hot motif is almost refractory to deamination.

From a biological perspective, Apo3G deamination polarity has been suggested as a likely explanation for the presence of localized mutational gradients in the HIV-1 genome that appear to have an important functional role in viral inactivation (17). The demonstration that random scanning of ssDNA generates nonrandom catalysis supports a model in which Apo3G binds in an asymmetric catalytically active orientation (18). Acknowledgments—We thank Jody Puglisi and Alfonso Brenlla for assistance generating the post-synchronization histograms.

REFERENCES
1. Barnes, D. E., and Lindahl, T. (2004) Repair and genetic consequences of endogenous DNA base damage in mammalian cells. Annu. Rev. Genet. 38, 445–476
2. Friedman, J. I., and Stivers, J. T. (2010) Detection of damaged DNA bases by DNA glycosylase enzymes. Biochemistry 49, 4957–4967
3. Pelot, J. U., Kuang, F. L., Iglesias-Ussel, M. D., Roa, S., Kalis, S. L., Goodman, M. F., and Scharff, M. D. (2008) The biochemistry of somatic hypermutation. Annu. Rev. Immunol. 26, 481–511
4. Chelioco, L., Pham, P., Petruska, J., and Goodman, M. F. (2009) Biochemical basis of immunological and retroviral responses to DNA-targeted cytosine deamination by activation-induced cytidine deaminase and APOBEC3G. J. Biol. Chem. 284, 27761–27765
5. Zhang, H., Yang, B., Pomerantz, R. J., Zhang, C., Aunachalam, S. C., and Gao, L. (2003) The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature 424, 94–98
6. Yu, Q., König, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J. M., and Landau, N. R. (2004) Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome.
Single Molecule FRET Studies of Apo3G Scanning ssDNA

7. Yang, B., Chen, K., Zhang, C., Huang, S., and Zhang, H. (2007) Virion-associated uracil DNA glycosylase-2 and apurinic/apyrimidinic endonuclease are involved in the degradation of APOBEC3G-edited nascent HIV-1 DNA. J. Biol. Chem. 282, 11667–11675

8. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L., and Xie, X. S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. Proc. Natl. Acad. Sci. U.S.A. 103, 5752–5757

9. Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003) Broad antiretroviral defense by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 424, 99–103

10. Jern, P., Russell, R. A., Pathak, V. K., and Coffin, J. M. (2009) Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. PLoS Pathog. 5, e1000367

11. Chiu, Y. L., and Greene, W. C. (2008) The APOBEC3 cytidine deaminases: an innate defensive network opposing exogeneous retroviruses and endogenous retroelements. Annu. Rev. Immunol. 26, 317–353

12. Navarro, F., Bollman, B., Chen, H., König, R., Yu, Q., Chiles, K., and Landa, N. R. (2005) Complementary function of the two catalytic domains of APOBEC3G. Virology 333, 374–386

13. Blanchard, S. C., Gonzalez, R. L., Chu, S., and Puglisi, J. D. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. Nat. Struct. Mol. Biol. 13, 419–424

14. Holden, L. G., Prochnow, C., Chang, Y. P., Bransteitter, R., Chelico, L., Sen, U., Stevens, R. C., Goodman, M. F., and Chen, X. S. (2008) Structural model for deoxycytidine deamination mechanisms of APOBEC3G. J. Biol. Chem. 283, 13780–13791

15. Blanchard, S. C., Gonzalez, R. L., Chu, S., and Puglisi, J. D. (2004) tRNA selection and kinetic proofreading in translation. Nat. Struct. Mol. Biol. 11, 1008–1014

16. Halford, S. E., and Marko, J. F. (2004) tRNA selection and kinetic proofreading in translation. Nature 430, 623–628

17. Halford, S. E., and Marko, J. F. (2004) tRNA selection and kinetic proofreading in translation. Nature 430, 623–628

18. Torella, J. P., Holden, S. J., Santoso, Y., Hohlbein, J., and Puglisi, J. D. (2003) Broad antiretroviral defense by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 424, 99–103

19. Pham, P., Bransteitter, R., Petruska, J., and Goodman, M. F. (2003) Processive AID-catalyzed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature 424, 103–107

20. Pham, P., Bransteitter, R., Park, S. J., and Goodman, M. F. (2011) Analysis of a single-stranded DNA-scanning process in which activation-induced deoxycytidine deaminase (AID) deaminates C to U haphazardly and inefficiently to ensure mutational diversity. J. Biol. Chem. 286, 24931–24942

21. Zhao, R., and Rueda, D. (2009) RNA folding dynamics by single-molecule fluorescence resonance energy transfer. Methods 49, 112–117

22. McKinney, S. A., Joo, C., and Ha, T. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. Biophys. J. 91, 1941–1951

23. Pham, P., Calabrese, P., Park, S. J., and Goodman, M. F. (2011) Analysis of a single-stranded DNA-scanning process in which activation-induced deoxycytidine deaminase (AID) deaminates C to U haphazardly and inefficiently to ensure mutational diversity. J. Biol. Chem. 286, 24931–24942

24. Pham, P., Bransteitter, R., Petruska, J., and Goodman, M. F. (2003) Processive AID-catalyzed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature 424, 103–107

25. Blanchard, S. C., Gonzalez, R. L., Chu, S., and Puglisi, J. D. (2004) tRNA selection and kinetic proofreading in translation. Nature 430, 623–628

26. Torella, J. P., Holden, S. J., Santoso, Y., Hohlbein, J., and Puglisi, J. D. (2003) Broad antiretroviral defense by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 424, 99–103

27. Halford, S. E., and Marko, J. F. (2004) tRNA selection and kinetic proofreading in translation. Nature 430, 623–628

28. Torella, J. P., Holden, S. J., Santoso, Y., Hohlbein, J., and Puglisi, J. D. (2003) Broad antiretroviral defense by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 424, 99–103

29. Halford, S. E., and Marko, J. F. (2004) tRNA selection and kinetic proofreading in translation. Nature 430, 623–628

30. Lasken, R. S., Schuster, D. M., and Rashtchian, A. (1996) Archaebacterial uracil recognition by archaeal family B DNA polymerases. Nature 384, 684–687

31. Fogg, M. J., Pearl, L. H., and Connolly, B. A. (2002) Structural basis for one-dimensional proteins find their targets? Nat. Struct. Mol. Biol. 9, H11032–H11037

32. Blanchard, S. C., Gonzalez, R. L., Chu, S., and Puglisi, J. D. (2004) tRNA selection and kinetic proofreading in translation. Nature 430, 623–628

33. Alani, E., and Greene, E. C. (2007) Dynamic basis for one-dimensional proteins find their targets? Nat. Struct. Mol. Biol. 14, 435–442