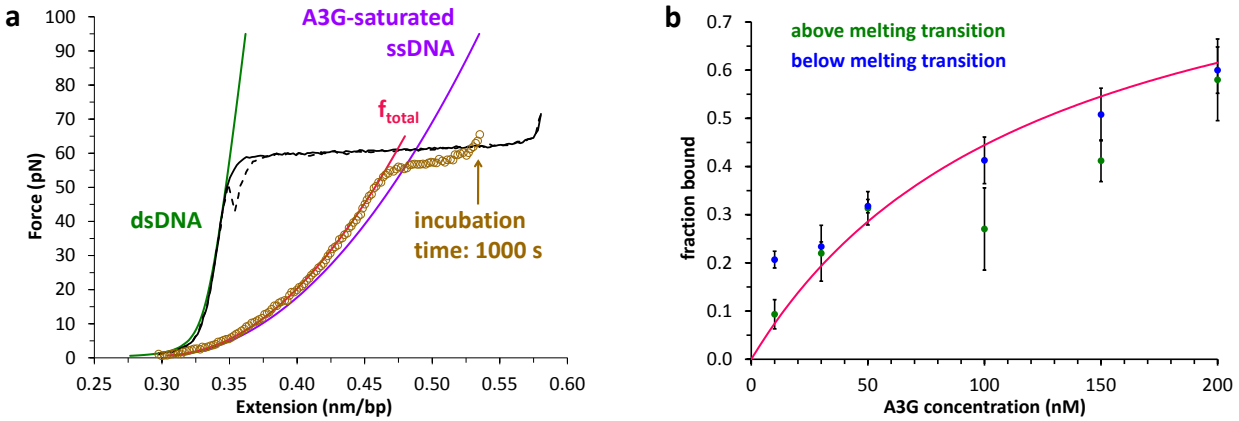
Supplementary Information

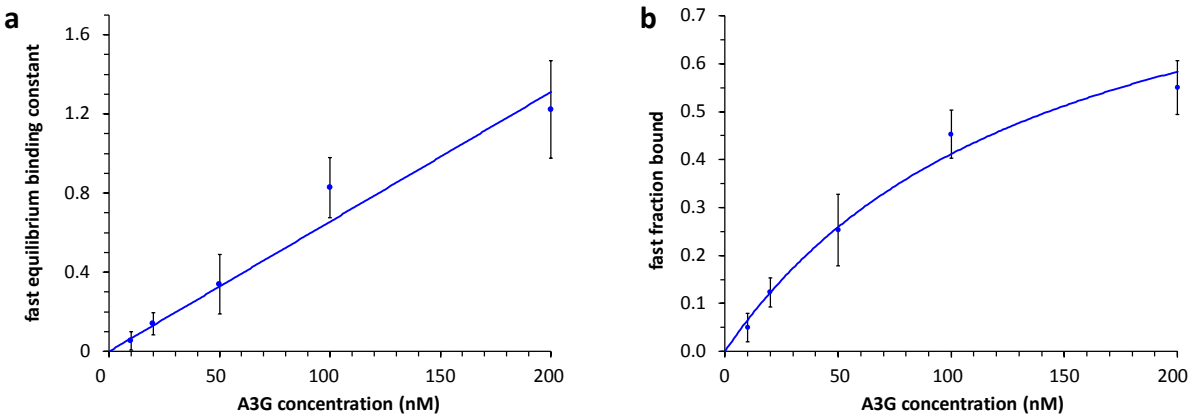
Oligomerization transforms human APOBEC3G from an efficient enzyme to a slowly dissociating nucleic acid binding protein

Kathy R. Chaurasiya, Micah J. McCauley, Wei Wang, Dominic F. Qualley, Tiyun Wu, Shingo Kitamura, Hylkje Geertsema, Denise Chan, Amber Hertz, Yasumasa Iwatani, Judith G. Levin, Karin Musier-Forsyth, Ioulia Rouzina, and Mark C. Williams

Supplementary Figure 1	DNA force-extension curves in the presence of A3G are used to quantify A3G binding to ssDNA.
Supplementary Figure 2	Concentration dependence of fast equilibrium binding constant and fast fraction bound.
Supplementary Table 1	Values for fitting parameters in the steady-state binding model.
Supplementary Table 2	Concentration-independent reaction parameters are similar across multiple methods, which supports the binding model.
Methods	Detailed supplementary methods section



Supplementary Figure 1. DNA force-extension curves in the presence of A3G are used to quantify A3G binding to ssDNA. (**a**) DNA release curve (brown circles, data also shown in Fig. 2, panel d) after stretching dsDNA in the presence of 50 nM A3G followed by 1000 s incubation at 0.53 nm/bp (indicated by the brown arrow). Below the melting transition, the force-extension curve is a linear combination of dsDNA (green, Eq. S1) and A3G-saturated ssDNA (purple, Eq. S5) because any ssDNA generated by force-induced melting that is not bound to A3G will immediately reanneal into dsDNA. The fit yields the fraction bound f_{total} = 0.90 ± 0.03 (red, Eq. S6). Uncertainty from the fits is small relative to variations across multiple measurements, so only standard errors are reported here. This data may also be fit above the melting transition as a linear combination of A3G-saturated ssDNA (purple, Eq. S5) and ssDNA (Eq. S2) because the molecule is primarily ssDNA regardless of protein binding. A3G binding during incubation shifts the release curve toward the saturated A3G curve (Eq. S7, fit not shown for clarity). (**b**) Total A3G bound without incubation (t = 50 s) as a function of concentration. Fits above (green) and below (blue) the melting transition yield similar results, as expected. Data points are averages from at least three fits, and error bars reflect standard error. The values from two methods were averaged (points and propagated error bars not shown for clarity) and fit to a simple DNA binding isotherm (solid red line, Eq. S8). The fit yields K_d = 125 ± 25 nM, which is similar to K_d = 76 ± 21 nM [18] and K_d = 160 nM[37] obtained from ensemble measurements. This measured K_d value agrees with K_{d1} = 127 ± 6 nM for fast binding obtained from K_1 (Eq. S18, Supplementary Fig. 2, panel a, see Supplementary Table 2 online).



Supplementary Figure 2. Concentration dependence of fast equilibrium binding constant and fast fraction bound. (**a**) The equilibrium binding constant K_1 is linearly dependent on A3G concentration, as expected for a bimolecular process. The solid line is a linear fit to Eq. S18 with slope k_1/k_{-1} = 6.6 (± 0.4) x 10^6 M^{-1}, which yields a dissociation constant for the fast step (K_{d1} = 127 ± 6 nM) that agrees with the one obtained from a simple DNA binding isotherm without incubation (K_d = 125 ± 25 nM, see Supplementary Fig. 1, panel b). (**b**) Fast portion of A3G binding P_{fast} saturates at high protein concentration. Data points and their uncertainties are from fits to the binding model (see Supplementary Table 1), solid line is a fit to Eq. S12 that yields k_1 = 1.4 (± 0.1) x 10^5 $M^{-1}s^{-1}$ and k_{-1} = 2.1 (± 0.9) x 10^{-2} s^{-1}. These values are fixed in the k_{slow} versus c fit (Figure 3, panel d, solid green line) to limit the fit to two free parameters.

A3G (nM)	k_{fast} (x 10^{-2} s^{-1})	k_{slow} (x 10^{-3} s^{-1})	P_{fast}	P_{total}
10	1.2 ± 0.2	0.5 ± 0.1	0.05 ± 0.03	0.90 ± 0.01
20	1.4 ± 0.2	0.7 ± 0.2	0.12 ± 0.03	0.88 ± 0.02
50	1.9 ± 0.2	1.6 ± 0.3	0.25 ± 0.07	0.95 ± 0.02
100	3.0 ± 0.2	2.9 ± 0.3	0.45 ± 0.05	0.97 ± 0.05
200	5.0 ± 0.5	5.0 ± 1.0	0.55 ± 0.06	0.99 ± 0.15

Supplementary Table 1. Values for fitting parameters in the steady-state binding model. Total binding at a given A3G concentration (f_{total}) was separated into its fast (f_{fast}) and slow (f_{slow}) fractions and monitored as a function of ssDNA-A3G incubation time, as illustrated in Fig. 2. The three curves obtained at each protein concentration (shown in Fig. 3, panel a for 50 nM A3G) were fit to Eqs. S9-S11. The parameters from each fit were used to calculate weighted averages for the observed binding rates k_{fast} and k_{slow}, and the equilibrium fractions of fast and total binding P_{fast} and P_{total}.

Method	k_1 (x 10^5 $M^{-1}s^{-1}$)	k_{-1} (x 10^{-2} s^{-1})	Method	k_2 (x 10^{-3} s^{-1})	k_{-2} (x 10^{-5} s^{-1})	Method	K_1/c (x 10^6 M^{-1})	Method	K_2
A1	2.0 ± 0.1	1.0 ± 0.1	B1	7.0 ± 1.0	3.0 ± 2.0	C1	6.6 ± 0.4	D1	166 ± 17
A2	1.4 ± 0.1	1.1 ± 0.1	B2	6.6 ± 0.7	2.8 ± 0.5	C2	12.9 ± 0.8	D2	251 ± 54
A3	1.2 ± 0.1	2.0 ± 0.2							
A4	1.4 ± 0.1	2.1 ± 0.9							
Average	1.5 ± 0.1	1.2 ± 0.1		6.7 ± 0.6	2.8 ± 0.5		7.9 ± 0.4		174 ± 16

Supplementary Table 2. Concentration-independent reaction parameters are similar across multiple methods, which supports the binding model. Weighted averages and uncertainty are reported in the main text. Methods A1 and A2 are linear fits of k_{fast}, k_1c, and k_{-1} vs. c (Fig. 3, panel a). Calculations from Eqs. S14 and S16 were averaged across protein concentration in method A3. Method A4 is the P_{fast} vs. c fit (Fig. S2, panel b), and method B1 is the k_{slow} vs. c fit (Fig. 3, panel b). Calculations from Eqs. S24 and S25 were averaged across A3G concentration in method B2. Method C1 is the K_1 vs. c fit (Supplementary Fig. 2, panel a) and method C2 uses the weighted averages of k_1 and k_{-1} in Equation S18. The weighted average of calculations from Eq. S23 over all measured protein concentrations is method D1. Method D2 uses experimentally-determined parameters in Eq. S22.

METHODS

A3G preparation and purification

Recombinant, catalytically active A3G was expressed in a baculovirus expression system with an N-terminal glutathione S-transferase (GST) tag[18]. The initial steps in purification were performed essentially as described previously[18]. Briefly, cell lysates treated with DNase I and RNase A were prepared and the soluble fraction was bound to glutathione Sepharose High Performance resin (GE Healthcare) for 3 h at 4 °C. The resin was washed as described[18] and bound GST-A3G was eluted with glutathione buffer (50 mM Tris-HCl, pH 8.0, 1 M NaCl, 40 mM reduced glutathione, 10 µM $ZnCl_2$, 10% glycerol (v/v)). The eluate was filtered through a Acrodisc 25 mm syringe filter with a 1.2 µm Versapor® membrane (Pall Corp.) and was then dialyzed overnight at 4 °C in buffer containing 20 mM Tris-HCl, pH 7.4, 1 M NaCl, 5 mM DTT, and 10% glycerol. Samples were stored at -80 °C prior to further purification. The partially purified GST-A3G was loaded onto a gel filtration column (GE Healthcare, Superdex 200 HiLoad 16/60) pre-equilibrated in a buffer containing 20 mM HEPES, pH 7.4, 200 mM NaCl, 5 mM DTT, and 10% glycerol (v/v). The protein was eluted with the same buffer and the peak corresponding to monomeric GST-A3G was collected and concentrated to an appropriate volume. The GST tag was removed using a Novagen Enterokinase Cleavage Capture Kit. After the cleavage reaction, enterokinase was removed using the capture resin provided in the kit and the GST tag was removed with glutathione Sepharose 4B (GE Healthcare). The deaminase activity of purified A3G was measured using a gel-based uracil DNA glycosylase assay with the 40-nt ssDNA substrate labeled at its 5´ end with Alexa Fluor® 488 (Integrated DNA Technologies, Coralville, IA) instead of ^{32}P[18].

For expression of dimerization-deficient recombinant protein, A3G FW, the pFastBac-GST-A3G FW plasmid was first constructed by cloning an EcoRI–HindIII fragment of pAcG2T-APO3G F126A/W127A (a gift from Myron F. Goodman)[41] into pFastBac1 (Life Technologies) and recombinant virus was obtained using the Bac-to-Bac Baculovirus Expression System (Life Technologies) as previously described[18]. Purification of the A3G FW protein was carried out using our previously reported procedure[18], with the exception of a GST tag-cleaving step and a final purification step. Briefly, Sf21 cells expressing GST-A3G FW were lysed in lysis buffer (25 mM HEPES, pH 7.0, 500 mM NaCl, 1% Triton, 10 mM $CaCl_2$, 1 mM EDTA, 5 mM 2-mercaptoethanol (2-ME), 10% glycerol (v/v), 1 mM PMSF, complete protease inhibitor (Roche Applied Sciences)) and treated with DNase I (5 U/ml, Sigma-Aldrich) and RNase A (40 µg/ml, Qiagen). Cleared cell lysates were then incubated with Glutathione Sepharose 4 Fast Flow resin (GE Healthcare Life Sciences) and subjected to a series of salt washes (500-1000 mM NaCl). The bound GST-A3G FW protein was eluted with the glutathione buffer and the eluate was dialyzed against thrombin cleavage buffer (110 mM Tris HCl, pH 7.4, 100 mM NaCl, 2 mM $CaCl_2$, 1 mM 2-ME, 0.5% TritonX-100). To obtain A3G FW without a GST tag, the dialyzed solution was incubated in the presence of recombinant Thrombin (10 U/ml) (GE Healthcare Life Technologies) at 20 °C for 16 h. The resultant A3G FW was subjected to fractionation with ion exchange chromatography using an SP Sepharose FF Column (GE Healthcare Life Technologies) and collected. The deaminase activity of purified A3G FW was determined without removing the GST tag and was measured using a gel-based uracil DNA glycosylase assay with the 40-nt ssDNA substrate labeled at its 5’ end with FAM (Carboxyfluorescein, Operon Biotechnologies, Japan).

The concentrations of the final monomeric A3G and A3G FW were determined by UV absorbance (ε_{280} = 105,260 $M^{-1}cm^{-1}$) and Bradford assays. The purified proteins were stored at -80 °C.

Single molecule experiments

Single molecule DNA stretching experiments were conducted with dual beam optical tweezers as previously described[40]. Biotin-labeled bacteriophage λ DNA (48.5 kbp, 16.5 µm contour length) was captured between two streptavidin-coated polystyrene beads (Bangs Labs), one fixed on a micropipette

tip and one in the optical trap. The flow chamber was rinsed with buffer (50 mM Na^+, 10 mM HEPES, pH 7.5), and the captured DNA molecule was then stretched and released to obtain a DNA-only force-extension curve (see Supplementary Fig. 1, panel a, black). End-to-end DNA extension is scaled by the total number of base pairs in the molecule to make force-extension curves independent of DNA length. However, the total time for a stretch-release cycle at a fixed pulling rate does depend on DNA length. Specifically, λ DNA increases approximately 10 μm in length during the force-induced melting transition, so going through the transition takes 100 s at the fixed pulling rate of 100 nm/s. DNA release takes the same amount of time.

The Worm-Like Chain (WLC) model describes dsDNA in terms of length b_{ds} at a given force F:

$$b_{ds}(F) = B_{ds}\left[1 - \frac{1}{2}\left(\frac{k_B T}{P_{ds} F}\right)^{\frac{1}{2}} + \frac{F}{S_{ds}}\right] \quad \text{S1}$$

with persistence length P_{ds}, contour length B_{ds}, and stretch modulus S_{ds}. The freely-jointed chain (FJC) model describes ssDNA in terms of length b_{ss} at a given force F:

$$b_{ss}(F) = B_{ss}\left[\coth\left(\frac{2P_{ss}F}{k_B T}\right) - \frac{1}{2}\frac{k_B T}{P_{ss} F}\right]\left[1 + \frac{F}{S_{ss}}\right] \quad \text{S2}$$

The WLC (green) and FJC (blue) polymer models are shown in Supplementary Fig. 1 with typical parameter values (B_{ds} = 0.34 nm/bp, P_{ds} = 48 nm, and S_{ds} = 1200 pN in Eq. S1, B_{ss} = 0.55 nm/bp, P_{ss} = 0.75 nm, and S_{ss} = 720 pN in Eq. S2). A linear combination of the WLC and FJC models describes the measured length b as a function of the fraction of ssDNA generated upon DNA extension f_{ss}[40]:

$$b(f_{ss}) = b_{ds}(1 - f_{ss}) + b_{ss} f_{ss} \quad \text{S3}$$

After three force-extension cycles of DNA alone, the buffer was replaced with protein dilutions ranging from 1 nM to 200 nM A3G. After the initial stretch-release cycle in the presence of A3G, the DNA molecule was stretched again. Incubation was performed at fixed DNA extension above the melting transition (Fig. 2) before the release curve was obtained. The following stretch-release cycle was a control experiment performed without incubation to ensure that the additional binding effect was no longer observed. A 30 s wait time was introduced between each DNA release and the subsequent stretch to ensure complete dissociation of the A3G component exhibiting fast binding kinetics.

Total A3G-ssDNA binding saturated at 200 nM A3G after incubation for 500 s or longer. The force versus extension curve $F(b_{A3G})$ for A3G-saturated ssDNA was phenomenologically described by a second-order polynomial:

$$F(b_{A3G}) = A b_{A3G}^2 + B b_{A3G} + C \quad \text{S4}$$

or, in terms of extension b versus force F:

$$b_{A3G}(F) = \frac{-B + \sqrt{B^2 - 4A(C - F)}}{2A} \quad \text{S5}$$

The parameters of the fit to this polynomial are A = 1713, B = 1028, C = 155 (Fig. 1). The A3G-saturated ssDNA curve was used to analyze fractional A3G binding. Below the melting force F_m = 61.0 ± 0.5 pN, the force-extension curve in the presence of A3G was fit to a linear combination of the dsDNA curve $b_{ds}(F)$ (Eq. S1) and the A3G-saturated curve $b_{A3G}(F)$ (Eq. S5), shown in Supplementary Fig. 1, panel a:

$$b(F < F_m) = b_{ds}(1 - f_{total}) + b_{A3G} f_{total} \quad \text{S6}$$

where f_{total} is fraction of A3G-bound ssDNA. Fits above the melting transition were to a linear combination of $b_{A3G}(F)$ (Eq. S5) and the ssDNA curve $b_{ss}(F)$ (Eq. S2):

$$b(F > F_m) = b_{A3G} f_{total} + b_{ss}(1 - f_{total}) \quad \text{S7}$$

Measurements at forces above the transition were used to confirm that the results of both independent methods agree within error (Supplementary Fig. 1, panel b). Values above and below the melting transition were averaged (data points with propagated error bars not shown), and fraction bound f_{total} as a function of protein concentration c was fit to a simple DNA binding isotherm:

$$f_{total} = \frac{\frac{c}{K_d}}{1 + \frac{c}{K_d}} \quad \text{S8}$$

where K_d is the equilibrium dissociation constant.

Two-step binding model

The kinetics of A3G-ssDNA binding was described as a two-step process, with an initial fast bimolecular binding step that reaches pre-equilibrium before a slow, unimolecular conversion to a more stable complex (Eq. 1). A complete analytical description of this process[48] yields explicit expressions for the fast, slow, and total fractions of A3G bound as a function of time:

$$f_{fast}(t) = P_{total} P_{fast} (e^{-k_{slow} t} - e^{-k_{fast} t}) \quad \text{S9}$$

$$f_{slow}(t) = P_{total} (1 - e^{-k_{slow} t}) \quad \text{S10}$$

$$f_{total}(t) = P_{total} (1 - P_{fast} e^{-k_{fast} t} - (1 - P_{fast}) e^{-k_{slow} t}) \quad \text{S11}$$

where k_{fast} and k_{slow} are binding rates, and P_{fast} and P_{total} are the equilibrium fractions of fast and total complex bound. P_{fast} is the ratio of the forward rate and the total rate for the first step:

$$P_{fast} = \frac{k_1 c}{k_1 c + k_{-1}} \quad \text{S12}$$

Fits to Eqs. S9-S11 (Figs. 3, panels a and b) yield the four parameters k_{fast}, k_{slow}, P_{fast} and P_{total} (reported in supplementary Table 1), which are used to determine the four elementary reaction rates $k_1 c$, k_{-1}, k_2 and k_{-2}, as well as the two equilibrium binding constants K_1 and K_2, for the first and the second binding steps, respectively.

Fast on and off rates

The fast rate k_{fast} is a sum of the forward and backward elementary reaction rates:

$$k_{fast} = k_1 c + k_{-1} \quad \text{S13}$$

where $k_1 c$ is the on rate and k_{-1} is the off rate for the first step, defined in Eq. 1. The model assumes the fast process is bimolecular, so the on rate is proportional to A3G concentration c with rate constant k_1. The data confirms that $k_{fast}(c)$ is linear (Fig. 3, panel c), with bimolecular rate constant k_1 = 2.0 (± 0.1) x 10^5 $M^{-1}s^{-1}$ and off rate k_{-1} = 1.0 (± 0.1) x 10^{-2} s^{-1} (Supplementary Table 2, method A1).

The fast on and off rates were also calculated as follows:

$$k_1 c = k_{fast} P_{fast} \quad \text{S14}$$

The slope of a linear fit to the calculated on rate versus *c* (Fig. 3, panel c) is an additional measurement of k_1:

$$k_1 c = k_1 c + b \quad \text{S15}$$

where the measured value of k_1 = 1.4 (± 0.1) x 10^5 $M^{-1}s^{-1}$ (Supplementary Table 2, method A2) is consistent with the value from the first method. As expected, the y-intercept of the fit is nearly zero (*b* = -8 (± 5) x 10^{-4} s^{-1}). The binding model also defines the off rate k_{-1}:

$$k_{-1} = k_{fast}(1 - P_{fast}) \quad \text{S16}$$

which is expected to be concentration-independent (Fig. 3, panel d). Although the linear fit:

$$k_{-1} = mc + k_{-1} \quad \text{S17}$$

yields a non-zero slope (*m* = 6.1 (± 0.3) x 10^4 $M^{-1}s^{-1}$), the linear dependence is an order of magnitude smaller than that of $k_{fast}(c)$. This measurement of k_{-1} = 1.1 (± 0.1) x 10^{-2} s^{-1} (Supplementary Table 2, method A2) agrees with the value from the first method within error.

The bimolecular rate constant calculated from Eq. S14 is concentration-independent as expected, with a weighted average of k_1 = 1.2 (± 0.1) x 10^5 $M^{-1}s^{-1}$. The k_1 values from the second and third methods agree within error. The off rates calculated from Eq. S16 are also nearly concentration-independent, and the weighted average of the five concentrations is k_{-1} = 2.0 (± 0.2) x 10^{-2} s^{-1} (Supplementary Table 2, method A3).

Fast equilibrium binding constant

The equilibrium binding constant for the bimolecular reaction step, K_1, is a ratio of the forward and backward rates:

$$K_1 = \frac{k_1 c}{k_{-1}} \quad \text{S18}$$

and was calculated from values of $k_1 c$ and k_{-1} obtained from fits. It may also be calculated from $P_{fast}(c)$:

$$K_1 = P_{fast}(1 - P_{fast}) \quad \text{S19}$$

The fast equilibrium binding constant depends linearly upon A3G concentration, as expected (Fig. S2, panel a), and the slope was used to calculate the equilibrium association constant K_1/c (Supplementary Table 2, method C1). Experimentally determined values of k_1 and k_2 were also used to calculate K_1/c directly from Eq. S18 (Supplementary Table 2, method C2). This was used to obtain the equilibrium dissociation constant K_{d1} for the first step:

$$K_{d1} = \frac{c}{K_1} = \frac{k_{-1}}{k_1} \quad \text{S20}$$

The K_d value obtained from this method (127 ± 6 nM) agrees with the K_d of the total fraction bound at *t* = 50 s of A3G exposure to ssDNA (125 ± 25 nM) fit to a simple DNA binding isotherm (Supplementary Fig. 1, panel b). This is a reasonable observation, in light of the fact that total A3G binding measured after only 50 s is primarily fast binding, which has equilibrated ($1/k_{fast}$ = 24 ± 1 s) while slow binding has not ($1/k_{slow}$ = 206 ± 20 s).

Slow on and off rates

The slow rate k_{slow} is the sum of the forward rate k_2 and backward rate k_{-2}:

$$k_{slow} = k_2 P_{fast} + k_{-2} \quad \text{S21}$$

where the effective forward rate is the product of the conversion rate k_2 and the equilibrium fraction of the fast bound complex P_{fast}. A fit of $P_{fast}(c)$ to Eq. S12 (Fig. S2, panel b) is a fourth method of measuring the fast on and off rates, k_1 = 1.4 (± 0.1) x 10^5 $M^{-1}s^{-1}$ and k_{-1} = 2.1 (± 0.9) x 10^{-2} s^{-1} (Supplementary Table 2, method A4). These values are fixed in the fit of $k_{slow}(c)$ to Eq. S21, which yields k_2 = 7.0 (± 1.0) x 10^{-3} s^{-1} and k_{-2} = 3.0 (± 2.0) x 10^{-5} s^{-1} (Supplementary Table 2, method B1).

An alternative way to calculate these elementary reaction rates involves the slow equilibrium binding constant K_2:

$$K_2 = \frac{k_2}{k_{-2}} \quad \text{S22}$$

which can also be expressed as:

$$K_2 = \frac{1 + \frac{1}{K_1}}{\frac{1}{P_{total}} - 1} \quad \text{S23}$$

Calculations from Eqs. S22 and S23 yield a weighted average of K_2 = 174 ± 16, as reported in Supplementary Table 2 (method D), suggesting that the second binding step is so highly forward-driven that it is nearly irreversible.

The value of K_2 may be used to calculate the forward rate k_2:

$$k_2 = \frac{k_{slow} K_2}{K_2 P_{fast} + 1} \quad \text{S24}$$

and the backward rate k_{-2}:

$$k_{-2} = \frac{k_{slow}}{K_2 P_{fast} + 1} \quad \text{S25}$$

The weighted average of these values is k_2 = 6.6 (± 0.7) x 10^{-3} s^{-1} and k_{-2} = 2.8 (± 0.5) x 10^{-5} s^{-1} (Supplementary Table 2, method B2), which is consistent with the estimate of these rates based on the fit of k_{slow}, as discussed above. Both on and off rates are independent of protein concentration within error (Figure 3, panel d), consistent with a model in which slow binding is unimolecular. Therefore, k_{slow} increases with A3G concentration as fast binding grows, and then saturates with on rate k_2 at high protein concentrations such as those relevant in the virus.