Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G

Takahide Kouno1,2, Elizabeth M Luengas3, Megumi Shigematsu1,4, Shivender M D Shandilya2, JingYing Zhang1,4, Luan Chen1,4, Mayuko Hara1,4, Celia A Schiffer2, Reuben S Harris3 & Hiroshi Matsuo1

The human APOBEC3G (A3G) DNA cytosine deaminase restricts and hypermutates DNA-based parasites including HIV-1. The viral infectivity factor (Vif) prevents restriction by triggering A3G degradation. Although the structure of the A3G catalytic domain is known, the structure of the N-terminal Vif-binding domain has proven more elusive. Here, we used evolution- and structure-guided mutagenesis to solubilize the Vif-binding domain of A3G, thus permitting structural determination by NMR spectroscopy. A smaller N-terminal domain is bound by Vif; this enabled mutagenesis and biochemical experiments, which identified a unique Vif-interacting surface formed by the α1-β1, β2-α2 and β4-α4 loops. This structure sheds new light on the Vif-A3G interaction and provides critical information for future drug development.

The human APOBEC3 (A3) proteins are single-stranded DNA cytosine deaminases that are part of a larger network of innate immune effector proteins that serve to limit the replication of many diverse parasitic elements including the AIDS virus HIV-1 (refs. 1–3). Up to seven family members, A3A, A3B, A3C, A3D, A3F, A3G and A3H, can be expressed in a single human cell4. A combination of overexpression, knockdown, knockout and separation-of-function studies have shown that A3D, A3F and A3H contribute to HIV-1 restriction, knockdown, knockout and separation-of-function studies have shown that A3D, A3F, A3G and A3H contribute to HIV-1 restriction and hypermutation in CD4+ T lymphocytes (refs. 5–7 and references therein). A3D, A3F and A3H catalyze C-to-U deamination in 5′-TC dinucleotide motifs, whereas A3G catalyzes deamination in 5′-CC motifs. These viral cDNA-deamination events lead to G-to-A mutations in hallmark GA and GG dinucleotide motifs, respectively. A3G has two structural domains: an N-terminal pseudocatalytic domain (NTD) and a C-terminal catalytic domain (CTD)8,9. The CTD alone dictates the observed local dinucleotide preference10–12. The NTD is catalytically inert but still strongly binds RNA and single-stranded DNA and forms a ribonucleoprotein complex with the viral Gag protein, thus enabling encapsidation (examples in refs. 13–17).

HIV-1 encodes Vif as a counterdefense mechanism against A3-mediated restriction18,19. Vif interacts with core-binding factor β (CBF-β), CUL5, ELOB, ELOC and RBX2 proteins, to form an E3 ubiquitin ligase complex20,21. As a consequence, A3G and other potentially restrictive A3 enzymes are polyubiquitinated and degraded by the proteasome before being encapsidated (examples in refs. 18,19,22–25). The interaction between Vif and A3G has been studied extensively, and key regions for the interaction have been identified, which include 126-FWDPDYQ-132 of A3G, and 22-KSLVK-26, 40-YRHHY-44 and 69-YWGL-72 of Vif26–29. However, insolvability has limited the acquisition of structural information for the A3G–Vif complex. A crystal structure of the Vif–CBF-β–ELOB–CUL5 complex was reported recently30. Vif has two structural domains, called α/β and α3G-binding residues, including K22, K26, 40-YRHHY-44 and W70, are located in the α/β domain30. Although the structure of catalytically active A3 proteins A3A and A3C, as well as the catalytic domains A3G CTD and A3F CTD, were previously solved by NMR and X-ray crystallography12,31–38, a structure of the N-terminal pseudocatalytic Vif-binding domain of A3G has thus far proven elusive. To address this major gap in knowledge, we have generated a soluble A3G NTD variant (sNTD) that binds the Vif–CBF-β–ELOB (VCBC) complex. The sNTD has 80% amino acid sequence identity to wild-type A3G NTD. In this paper, we report the three-dimensional structure of sNTD and mutationally interrogate the Vif-binding regions.

RESULTS

Engineering a soluble A3G Vif-binding domain

The wild-type A3G NTD variant was insoluble in Escherichia coli. To overcome this technical hurdle, we compared the amino acid sequences of A3G NTD with those of several A3s that were soluble and amenable to structural studies: A3A, A3B CTD and A3G CTD12,31–34,37. A3A, A3B CTD and A3G CTD were catalytically active but did not interact...
with Vif, and they belong to the same phylogenetic subgroup, the Z1 group39 (Fig. 1a). An alignment of A3A, A3B CTD and A3G CTD allowed us to generate a consensus Z1 amino acid sequence (Supplementary Fig. 1a). We expressed and purified the consensus Z1 protein from E. coli and found that it was soluble and folded. We used wild-type A3G NTD amino acid residues to replace consensus Z1 residues one at a time and then tested all of the resulting constructs for protein expression, solubility and stability. We used NMR to evaluate soluble derivatives for signal dispersion and heterogeneity. Ultimately, we achieved 80% amino acid sequence identity to wild-type NTD without substantially compromising the solubility or quality of NMR spectra (Fig. 1b,c). Additional wild-type A3G amino acids were not tolerated (Supplementary Fig. 1b,c). We henceforth refer to this soluble A3G NTD construct as sNTD.

**Solution structure of sNTD**

We generated a series of $^{15}$N, $^{13}$C and/or $^2$H isotope-labeled protein samples to determine the solution structure of sNTD with state-of-the-art NMR techniques combined with amino acid–specific labeling techniques31,40 (representative NMR spectra in Supplementary Fig. 2). We subjected each sample to size-exclusion chromatography for purification and evaluated the monomeric status by the elution profile. sNTD eluted with a major peak corresponding to a molecular weight of 17 kDa and a minor peak corresponding to 47 kDa (Fig. 1b); the theoretical molecular weight of sNTD is 21 kDa. sNTD exhibited fair-quality NMR spectra with most amide proton signals resolved in a transverse relaxation-optimized spectroscopy (TROSY) spectrum, although signal intensities were inhomogeneous, thus suggesting diverse dynamics throughout sNTD (Supplementary Fig. 3). Using conventional triple-resonance NMR spectra40, we assigned backbone signals ($^{1}$HN, $^{13}$N, $^{1}$Hα, $^{13}$Cα and $^{13}$C”) for 146 residues (81.1%) of sNTD. 24 residues were missing from TROSY spectra, owing to severe line broadening, thus suggesting conformational exchange. NOE signals and a $^1$H/$^2$H-exchange experiment indicated that sNTD contains a β-sheet of five β-strands arranged in the order of β2(54–56)–β1(36–42)–β3(85–92)–β4(113–120)–β5(148–152), with β2–β1–β3 in an antiparallel orientation and β3–β4–β5 in a parallel orientation (Supplementary Fig. 4). $^1$H–$^1$H NOE patterns indicated that β2–β1 form an antiparallel β-sheet; however, only Y37 NH from β1 showed slow $^2$H exchange with solvent, whereas F54 NH and G56 NH from β2 exchanged completely with solvent within 1 h (Supplementary Fig. 4 and data not shown). Thus, hydrogen bonds between F54 NH and V39 CO, and between G56 NH and Y37 CO were weaker than that between Y37 NH and G56 CO, results indicating that the structure and motions of β2 are likely to be less restrained than in the other β strands.

We determined the sNTD solution structure with 1,609 distance and 106 dihedral-angle restraints (structural statistics summarized in Table 1). The structure of sNTD has six α-helices, including α1(16–19), α2(68–75), α3(98–108), α4(131–140), α5(154–164) and α6(178–194), organized around the β-strands described above (Fig. 2). Secondary structures were well defined, although α1, β2 and α2 were less converged, as compared to other secondary structures (Fig. 2a). The poorer convergence for β2 was indicative of weaker hydrogen-bonding with β1. The loop between β1 and β2 (β1–β2 loop) was disordered because of insufficient NOE restraints. Residues located in the β1–β2 loop (E38, V39, K40, K42, G43, S45 and R46) and the α2–β3 loop (Q84, E85 and Y86) each showed two weak signals in TROSY spectra (Supplementary Fig. 3). Because the β1–β2 and α2–β3 loops face each other in the sNTD structure, these two loops may exchange between two conformations (Fig. 2b).

**Table 1 NMR and restraint statistics**

| NMR distance and dihedral constraints |  |
|--------------------------------------|--|
| Distance constraints |  |
| Total NOE | 1.543 |
| Intraresidue | 676 |
| Inter-residue |  |
| Sequential ($i$–$j$ = 1) | 338 |
| Medium range ($i$–$j$ < 4) | 183 |
| Long range ($i$–$j$ > 5) | 346 |
| Intermolecular | 0 |
| Hydrogen bonds | 33 (66) |
| Total dihedral-angle restraints |  |
| φ | 106 |
| ψ | 0 |

**Structure statistics**

Violations (mean ± s.d.)

| Distance constraints (Å) | 0.07 ± 0.06 |
| Dihedral-angle constraints (*) | 0.90 ± 0.77 |
| Max. dihedral-angle violation (*) | 3.70 |
| Max. distance constraint violation (Å) | 0.30 |
| Deviations from idealized geometry |  |
| Bond lengths (Å) | 0.004546 ± 0.000087 |
| Bond angles (*) | 0.6737 ± 0.0076 |
| Improper s (*) | 0.4877 ± 0.0153 |
| Average pairwise r.m.s. deviation (Å) |  |
| Heavy | 1.57 ± 0.17 |
| Backbone | 0.82 ± 0.17 |

*aPairwise r.m.s. deviation was calculated with residues 35–42, 53–58, 68–74, 84–93, 101–121, 131–161 and 172–194 of ten refined structures.*
The overall structure of sNTD was similar to that of other A3 proteins, including A3A, A3C, A3F CTD and A3G CTD. However, the sNTD structure also had several unique features. First, the start of α2 is residue 68 of sNTD, one residue after the conserved glutamate residue (E67), whereas in α2 of catalytically active A3 domains, the helix starts one residue prior to the conserved glutamate. This structural difference substantially affected the spatial positioning of E67, by causing the residue to be located too far from the Zn$^{2+}$ ion to participate in catalysis (Fig. 3b). Second, the α2 N terminus was tilted toward the C terminus of α3, an orientation not seen in any reported A3 catalytic-domain structures (Fig. 3c). These two features provided a likely explanation for prior studies that assigned catalytic activity exclusively to the CTD.

Second, sNTD possessed catalytic activity, by using one-dimensional NMR spectra. For a positive control, we incubated the catalytically inactive A3G sNTD with the VCBC complex (Fig. 4a, lane 8), and this interaction became undetectable when we used Vif H43A Y44A instead of wild-type Vif (Fig. 4a, lane 9). The location of A126 appeared to be in the middle of the loop between β4 and α4, where it was exposed to solvent (Fig. 4b). This surface location was important because the A126F substitution was unlikely to alter either secondary- or tertiary-structural elements.

sNTD is catalytically inactive in vitro

Because our solubilization strategy started with the consensus sequence of several active deaminases, we next asked whether sNTD possessed catalytic activity, by using one-dimensional $^1$H NMR spectra. For a positive control, we incubated the catalytically active A3G CTD variant (CTD 2K3A) with substrate DNA (5’-ATTCCCAATT-3’) and saw a signal corresponding to the uridine H5 proton emerge within 30 min at 5.72 p.p.m. (Supplementary Fig. 5). This result indicated that CTD 2K3A had catalyzed the first deamination reaction (5’-CCU to 5’-CUU), within 24 h, CTD 2K3A completed the second deamination reaction, 5’-CUU to 5’-CCU, thus resulting in two H5 proton signals, one at 5.74 p.p.m. and the other at 5.81 p.p.m. (Supplementary Fig. 5c). By contrast, sNTD did not cause changes to the NMR spectrum of the substrate DNA even after 24 h (Supplementary Fig. 5f). Thus, sNTD lacks detectable deaminase activity, consistently with prior studies demonstrating the catalytic inactivity of this domain.

sNTD binds Vif

We used GST pulldown experiments to monitor the interaction between sNTD and VCBC coexpressed in E. coli. sNTD bound to the VCBC complex (Fig. 4a, lane 10), and H43A and Y44A substitutions in Vif reduced binding to undetectable levels (Fig. 4a, lane 11). Because F126 of wild-type NTD had been substituted to an alanine in sNTD, in order to increase solubility, we reversed this mutation and tested binding to the VCBC complex. sNTD F126 bound more strongly than the original sNTD construct to the VCBC complex (Fig. 4a, lane 8), and this interaction became undetectable when we used Vif H43A Y44A instead of wild-type Vif (Fig. 4a, lane 9). The location of A126 appeared to be in the middle of the loop between β4 and α4, where it was exposed to solvent (Fig. 4b). This surface location was important because the A126F substitution was unlikely to alter either secondary- or tertiary-structural elements.

F126 is important for Vif-mediated degradation

We created a construct in which sNTD was fused to wild-type CTD (A3G sNTD-CTD) and tested whether this construct behaves similarly to wild-type full-length A3G in a Vif-mediated degradation assay (Fig. 4c). We used wild-type A3G as a positive control, which, as expected, demonstrated reduced protein levels in HEK293T cells when coexpressed with Vif (Fig. 4c, lane 3) than when expressed alone (Fig. 4c, lane 1). The A3G protein level was restored by addition of protease inhibitor MG132 (Fig. 4c, lane 4). We used a double mutation of H43A and Y44A in Vif (Fig. 4c, lane 5) and D128K in A3G (Fig. 4c, lane 6) as negative controls because these mutations have been shown to impair Vif-mediated degradation of A3G. Interestingly, A3G sNTD-CTD protein levels were not reduced upon coexpression with Vif (Fig. 4c, lane 15). Because sNTD F126 bound to the VCBC complex more strongly than sNTD in vitro (Fig. 4a, lane 8), we prepared a construct with sNTD F126 fused to wild-type CTD (sNTD F126-CTD) and tested its protein stability in a Vif-mediated degradation assay. sNTD F126-CTD exhibited reduced protein levels in the presence of Vif (Fig. 4c, lane 21), which were restored by coexpression with Vif H43A Y44A (Fig. 4c, lane 23) or by treatment with MG132 (Fig. 4c, lane 22). Furthermore, adding the Vif-antagonizing
single–amino acid substitution D128K protected the sNTD F126 D128K–CTD construct from degradation by Vif (Fig. 4c, lane 27). Prior studies have suggested the importance of F126 in Vif binding and in Vif–mediated degradation of A3G26,28,48. Thus, these data indicate that a single–amino acid reversal (F126) can restore Vif susceptibility to the A3G sNTD–CTD construct and that the overall mechanism of Vif–mediated degradation is preserved in this construct.

**sNTD lacks antiviral activity**

We compared wild-type A3G, sNTD–CTD and sNTD F126–CTD constructs in HIV-1 restriction experiments (Fig. 4d). As reported previously, wild-type A3G is packaged into viwf–deficient HIV-1 particles, and it restricts viral infectivity in a dose–dependent manner (Fig. 4d, lanes 1–4). Antiviral activity was strongly counteracted by Vif, as evidenced by less A3G in cells and viral particles, and mostly restored viral infectivity (Fig. 4d, lanes 5–8). In contrast, sNTD–CTD resisted Vif–mediated degradation and failed to package substantially into viral particles, despite near-wild-type A3G levels of cellular expression (Fig. 4d, comparison of lanes 17–20 and 21–24). However, as described above in coexpression studies, the single–amino acid substitution A126F (sNTD F126–CTD), restored the susceptibility of this construct to Vif–mediated degradation but did not restore packaging and restriction activities (Fig. 4d, comparison of lanes 9–12 and 13–16). These data confirmed that the sNTD F126–CTD construct is functional for Vif degradation but is compromised for packaging and restriction. Because packaging requires the formation of a ribonucleoprotein complex with Gag (examples in refs. 13–17), and similar regions of A3G have been implicated in both Vif interaction and RNA binding, the sNTD F126 variant appeared to separate these two activities by preserving the capacity to be bound and degraded by Vif and by losing the capacity to bind RNA (which also manifests as packaging and restriction deficiencies).

**Vif binding to sNTD differs in A3C and A3F**

We used sNTD variants, E. coli coexpression and GST copurification as described above to identify amino acid residues required for interaction of sNTD with the VCBC complex. Substitutions Y19 at the C terminus of α1; I26, L27 and W34 in the α1–β1 loop; V58 and Y59 in the β2–α2 loop; and Y124, W127 and D128 in the β4–α4 loop decreased binding to the VCBC complex by >50% (Fig. 5a and representative SDS–PAGE gel images in Supplementary Fig. 5g). These residues substantially expanded the size of the A3G NTD surface area required for Vif interaction (Fig. 5b). Interestingly, using a different experimental approach, Lavens et al. implicated a similar surface area (ref. 49 and Fig. 5c). Thus, the A3G NTD Vif–binding surfaces appeared to be in distinct areas from those of the Vif–binding surfaces of A3C and A3F CTD, both of which are formed by residues located in the α2 and α3 helices28 (Fig. 5d,e).

**DISCUSSION**

We used sequence conservation and restorative mutagenesis to engineer a soluble variant of the A3G NTD, which is 80% identical to the wild-type enzyme (Fig. 1a). This variant, like wild-type A3G NTD, is not catalytically active but still binds Vif and succumbs to Vif–mediated degradation (Fig. 4 and Supplementary Fig. 5f). This engineering approach, although labor intensive, is applicable to other systems that have evaded structural characterization because of protein solubility issues.

We used NMR spectroscopy to solve the structure of sNTD, generating what is, to our knowledge, the first structure of an APOBEC3 noncatalytic domain and the first structure of the Vif–binding domain of A3G. Our results suggest a mechanistic basis for the lack of......
catalytic activity. Zn$^{2+}$ is coordinated by residues H65, C97 and C100 at the protein interior (Fig. 3b), and it contacts the hydrophobic side chains of L35, W90, I92 and M104 (Supplementary Fig. 6a). These interactions reduce accessibility for a putative target cytosine because the pocket volume of the Zn$^{2+}$-coordinating region is reduced in comparison to that of the A3G CTD (Supplementary Fig. 6b). The active site pocket volume of the A3G CTD is ~400 Å$^3$, whereas that of sNTD is ~100 Å$^3$, which is smaller than the volume of a cytosine nucleobase. This result supports the general trend observed for catalytically active and inactive domains$^{25,26}$. In addition, conserved E67 of sNTD is not in a position in which Zn$^{2+}$ coordination is possible, thus preventing E67 from having a catalytic role (ref. 51, Fig. 3b and Supplementary Fig. 6a). This altered position can be attributed to changes at the N terminus of the α2 helix, where A3G NTD has a proline at residue 66 just before E67 that shortens this helix so that it begins at M68. In all catalytically active A3 domains, the helix begins at A66.

To make a soluble NTD, we had to introduce four amino acid substitutions in α2 (F71L, H72S, W73L and F74V), although these are conserved mutations among APOBEC3 proteins (Supplementary Fig. 6c). Residues 71 and 74 interact with F70 (α2), W90 (β3) and F107 (α3), and F70, W90 and F107 interact with Y37 (β1), V39 (β1) and V88 (β3). Together, these residues form a hydrophobic core that stabilizes the β1-α2-β3-α3 structure (Supplementary Fig. 6d). All known A3 protein structures have similar hydrophobic cores involving corresponding residues$^{12,21-38}$. Residues 72 and 73 are not involved in the hydrophobic cores of known A3 structures, and we have shown that these residues can be substituted with alanine without compromising the catalytic activity of the A3G CTD$^{52}$. Thus, substitutions at residues 71–74 are unlikely to be responsible for the unique structural features of α2 and α3 of sNTD.

The known Vif H43A Y44A$^{26}$ variant disrupts wild-type A3G binding and also does not permit binding to sNTD, thus suggesting that contacts made through H43 and Y44 are maintained between sNTD
and Vif. However, sNTD fused to wild-type CTD was largely resistant to Vif-mediated degradation. We found that A126 was responsible for this reduction because replacement of this alanine with endogenous phenylalanine increased interaction with the VBCB complex and restored susceptibility to Vif-mediated degradation. These results confirmed the importance of A3G F126 in Vif binding and, together with additional mutant data, demonstrated that the sNTD construct is a good structural mimic of wild-type A3G NTD.

A caveat of sNTD function is that neither sNTD-CTD nor sNTD F126–CTD was able to incorporate into viral particles or restrict HIV-1 infection. Encapsidation requires interaction with RNA and Gag protein (examples in refs. 13,14,53–55). The failure of sNTD variants to encapsidate is most probably due to an abrogated interaction with RNA because we could not detect interaction between sNTD and RNA in a phosphate buffer containing 100 mM NaCl (data not shown). This loss may be due to deletion and/or substitution of basic residues, including K2, R6, R11, R14, R78, K113, K141, R142, R146, R169 and K180 in sNTD (Supplementary Fig. 1). Future studies may be able to use sNTD as a starting point to determine residues essential for encapsidation and RNA binding.

The Vif-binding regions of sNTD identified by this study, including α1–β1, β2–α2 and β4–α4 loops, provide a structural binding site for Vif and validate previous studies with wild-type A3G (Fig. 5b,c). The β4–α4 loop contains residues 124–YYFWD–128, which are known to be important for Vif-mediated A3G degradation. Structures of the A3C and A3F Vif-binding domains have been solved, and their Vif-binding surfaces have been mapped by mutagenesis, communoprecipitation and degradation assays to a different surface. In these cases, Vif was proposed to bind a continuous surface located on the α2 and α3 helices of A3F CTD and A3C, away from the catalytic zinc (Fig. 5d,e). A3G residues F74, L80, Y86 and F107 are conserved in A3C, and they were previously found to be critical for A3C binding to Vif but, consistently with our results, not to be important for Vif-mediated degradation of A3G. Previous studies have shown that a negatively charged region around D128 appears to be important for binding to Vif (Supplementary Fig. 6b) because substitution of this amino acid abrogates binding to Vif26–29. The A3G CTD has a more positively charged profile in the corresponding area that probably interferes with binding to Vif30 (Supplementary Fig. 6b). In summary, we report the structure of the Vif-binding domain of A3G (sNTD), which further elaborates a new mode of Vif binding and provides a substrate for future structural studies of the entire A3G–Vif ligase complex as well as for drug discovery.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 2MZZ. NMR constraints have been deposited in the Biological Magnetic Resonance Bank under entry 25509.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by grants from the US National Institutes of Health (AI073167 to H.M. and GM091743 to R.S.H.). Salary support for T.K. was provided in part by a Takeda Biotechnology Foundation Fellowship. The University of Minnesota Supercomputing and NMR center (US National Science Foundation, BIR-961477) provided NMR instrumentation. We thank C. Kojima (Institute for Protein Research, Osaka University) for the pCold vector, K. Strelbel (US National Institutes of Health) for the pcDNA-hVif plasmid, M. Katahira for advice on structure calculations, Y. Iwatani for advice on degradation assays, Y. Xia for advice on NMR experiments, Y. Hong for supporting preparation of expression vectors, F. Liu for advice on human-cell experiments and K. Walters for editing the manuscript.

AUTHOR CONTRIBUTIONS

T.K. designed the strategy for searching for soluble A3G NTD mutants and all A3G NTD mutants; performed harvest assays, Vif binding assays in E. coli, Vif-dependent degradation assays and deamination assays; prepared NMR samples; and determined the solution structure. E.M.L. performed single-cycle infectivity assays. M.S. optimized human-cell experiments. S.M.D.S. analyzed...
surface-charge distribution and binding-pocket volumes of the NTD structure. IZ provides support for one or two I.C and M.H. processors with preparation of expression plasmids and with the Vif binding assay in E. coli. T.K., E.M.L., S.M.D.S., C.A.S., R.S.H. and H.M. contributed to writing and editing the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
ONLINE METHODS

Plasmids, antibodies and reagents. DNA fragments encoding the A3 consensus protein (Supplementary Fig. 1a) or wild-type A3G NTD (residues 1–196) were synthesized with codons optimized for expression in E. coli (GenScript). All DNA fragments encoding A3G NTD variants were prepared with a PCR-based method with the synthesized DNAs encoding the consensus protein or wild-type A3G NTD as templates. The prepared DNA fragments were inserted into the NdeI/Xhol site of the pCold-GST vector64. The DNA fragment encoding soluble A3G CTD variant (A3G CTD 2K3A containing residues 191–384 with mutations L234K C234A F310K C321 A356A) was obtained with a PCR-based method with the previously prepared expression vector31 and was inserted into the NdeI/Xhol site of the pCold-GST vector64. For expression of Vif (Vif-176, residues 1–176) and CBF-β (residues 1–141) in E. coli, codon-optimized DNA fragments encoding those proteins (GenScript) were inserted into the BamHI/HindIII and NdeI/Xhol sites of the pRSFDuet vector (Novagen), respectively. The Vif-176 H43A Y44A expression vector was prepared in the same manner, with a PCR-based method with Vif-176. For expression of elongin B (residues 1–112) and elongin C (residues 17–112) in E. coli, DNA fragments encoding those proteins were inserted into the BamHI/HindIII and NdeI/Xhol sites of the pACYCDuet vector (Novagen), respectively. DNA fragments encoding wild-type A3G NTD (residues 1–196), CTD (residues 197–384) or sNTD (residues 1–180) were synthesized with codons optimized for expression in human cells (GenScript). D218K substitution was introduced into the A3G NTD construct with a PCR-based method. Similarly, sNTD variants, including sNTD D128K, sNTD F126 and sNTD F126 D128K, were prepared with a PCR-based method with the DNA fragment encoding sNTD as a template. Each DNA fragment encoding NTD and its variants was combined with the wild-type A3G CTD with a PCR-based method to yield a full-length construct including sNTD-CTD, sNTD F126–CTD and sNTD F126 D128K–CTD. All full-length constructs were inserted into the Nhel/HindIII site of the pcDNA3.1 myc-His(-) vector (Invitrogen). The pcDNA-HVif expression plasmid was provided by K. Strebel (NIH). A double mutation, H43A and Y44A, was introduced into pcDNA-HVif with a PCR-based method to generate pcDNA-HVif H43A Y44A. DNA sequences of all constructs were verified by DNA sequencing. DNA oligomers for PCR and deamination assays were purchased from Integrated DNA Technologies. Mouse monoclonal anti-c-myc antibody (clone 9E10) was purchased from Sigma-Aldrich. Mouse monoclonal anti-Vif antibody (clone 319) and mouse monoclonal anti-α-tubulin antibody (clone DM1A) were purchased from Abcam. Anti-mouse IgG antibody conjugated with a fluorescent dye was purchased from Rockland Immunochemicals (cat. no. 610-152-121). Validation of antibodies used for this experiment are provided on the manufacturer’s websites. All stable isotope-labeled reagents, including [15N]ammonium chloride, [15N]glucose, [15N]deuterium oxide, [15N]phenylalanine, [15N]-leucine, [15N]-lysine, [15N]-arginine, [15N]-tyrosine, [15N]-alanine, [13C]-phenylalanine, [13C]-isoleucine, [13C]-lysine, [13C]-methionine, [13C]-arginine, [13C]-threonine, [13C]-tyrosine, [13C]-valine, [2H, 13C, 15N]-labeled algal amino acid mixture and [2H]choline o-sulfate, were purchased from Cambridge Isotope Laboratories.

Protein expression and purification. The consensus protein (Supplementary Fig. 1a), sNTD and their variants were prepared in a similar manner to methods described previously31. BL21 (DE3) cells (Invitrogen) were transformed with a prepared pCold-GST expression vector64–based plasmid and grown at 37 °C in M9 medium containing 100 μg ml–1 of ampicillin until reaching an optical density at 600 nm of 0.6–0.8. After the addition of 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 10 μM ZnCl2, the cells were further incubated at 16 °C overnight. The bacterial cells were harvested by centrifugation and resuspended with a lysis buffer (50 mM sodium phosphate, pH 7.3, 150 mM NaCl, 0.005% Tween-20 and 0.5 mM Tris (2-carboxyethyl)phosphine (TCEP)). The lysed cells were disrupted by sonication and then centrifuged. The supernatant was applied to glutathione-immobilized resin (GE Healthcare Life Sciences), and the resin was washed with the lysis buffer. The proteins bound to the resin were digested by HRV 3C protease (GE Healthcare Life Sciences) overnight. The digested proteins were applied to a Superdex-75 column (GE Healthcare Life Sciences) equilibrated with buffer (10 mM sodium phosphate, pH 7.3, 100 mM NaCl, 0.005% Tween-20, 0.5 mM TCEP and 0.01% NaN3). The fractions including the protein of interest were collected, and the protein was concentrated to 100–200 μM with Amicon Ultra filters (Millipore). For NMR experiments, 0.5 M choline o-sulfate, 5% deuterium oxide and 20 mM 2,2-dimethyl-2-silapentane-5-sulfonic sodium salt (DSS) were added to the protein solution. To prepare isotope-labeled proteins, M9 medium was prepared with 60% or 85% deuterium oxide with [13C]ammonium chloride and [15N]glucose or [1H, 13C]glucose. When the cells were grown in 85% deuterium oxide, 1 l−1 [1H, 13C, 15N]-labeled algal amino acid mixture was added before the IPTG induction to enhance the protein harvest. To prepare proteins with selectively [13C]- or [15N]-labeled amino acids, an amino acid uniformly labeled with [13C] or [15N] was added to the medium containing 85% deuterium oxide before the IPTG induction. Similarly, to prepare proteins with selectively protonated amino acids, an amino acid without any isotope labeling was added to the medium containing 85% deuterium oxide before the IPTG induction.

NMR spectroscopy. NMR spectra were measured with Bruker AV700, AV850 and AV900 spectrometers equipped with cryogenic probes at 298 K (MMNR center at University of MN). Sequential assignments of the backbone [1H, 13C] and [15N] resonances of sNTD were achieved by triple-resonance NMR techniques combined with amino acid selective labeling60–64. Assignments of side chain [1H] and [15N] resonances were achieved with a combination of [1H, 13C] HSQC spectra of amino acid selectively [13C]-labeled sNTD (Supplementary Fig. 2) and [1H, 13C, 15N] NOESY spectra of amino acid selectively protonated sNTD. [1H] chemical shifts were directly referenced to DSS, whereas [13C] and [15N] chemical shifts were indirectly referenced by the absolute frequency ratios [13C/H] = 0.251449530 and [15N/H] = 0.101329118 with internal DSS67. All NMR spectra were processed and analyzed with NMRPipe68 and Sparky (http://www.cgl.ucsf.edu/home/sparky), respectively.

Structure calculations. The solution structures of sNTD were calculated with a distance geometry–simulated annealing protocol with X-PLOR-NIH69. The structure calculations and NOE peak assignments were performed in an iterative and manual manner. For distance restraints, the lower limit was set to 1.8 Å, and the upper limit was set to 2.5, 3.0, 3.5, 4.0, 4.5 or 5.0 Å according to the NOE cross-peak intensity. Additionally, pseudoatom corrections were applied66. In total, 1,543 distance restraints were collected from [13C] and [15N]-edited NOESY spectra and a series of [1H, 15N] NOESY spectra with amino acid–selectively protonated sNTD. A total of 106 backbone dihedral angles (φ) were estimated with TALOS, on the basis of 13Cα, 13Cβ and 13C′ chemical shifts34. A total of 33 hydrogen bonds were identified from an H/D-exchange experiment, and the structures were calculated in the early stage. The hydrogen-bond restraints, 2.3–3.3 Å for N-O and 1.3–3.3 Å for HN-O, were introduced into the late stage of calculations. Additionally, Zn2+ ion and empirical restraints between Zn2+ ion and sNTD residues H65, C97 and C100 were introduced into the late stage of the calculation32. A total of 100 structures were calculated, and the ten lowest-energy structures were selected. These ten structures were used for the further calculation of an energy-minimized average structure. The quality of obtained structures was analyzed with MOLMOL71 and PROCHECK72. Structural figures were generated with MOLMOL and MacPyMOL software (Schrödinger).

Protein harvest assay. To monitor expression level and relative solubility of A3G NTD variants, the aforementioned protein expression and purification procedures were applied. The volumes of medium (400 ml), lysis buffer (25 ml) and glutathione-immobilized resin (100 ml) were strictly fixed. The resin bound to the protein was resuspended with lysis buffer to adjust the total volume to 200 μl. 5 μl of the resuspension was applied to an SDS-PAGE gel that was stained with Coomassie brilliant blue and quantified for measured protein band density by Imagej, provided by the US National Institutes of Health. Each variant was independently cultured three times to calculate mean and s.d.

DNA deamination assay. Synthesized DNA fragment d(S′-ATTCCCAATT-3′) was dissolved in buffer (10 mM sodium phosphate, pH 7.3, 100 mM NaCl, 0.005% Tween-20 and 0.5 mM TCEP) at a concentration of 100 μM. Purified A3G CTD 2K3A or sNTD was added to the DNA solution at a final concentration of 20 μM. The mixture was incubated at 25 °C. One-dimensional 1H NMR spectra were measured 0.5, 3 and 24 h after the addition of protein.
**E. coli coexpression and GST pulldown assay.** BL21(DE3) cells were transformed with a pCold-GST-NTD variant expression plasmid, prSEF-Duet-Vif-176-CBF-β and pACYC-Duet-ELOBC, and were grown at 37 °C in 400 ml of Luria-Bertani medium containing 100 µg ml⁻¹ of ampicillin, 25 µg ml⁻¹ of kanamycin and 30 µg ml⁻¹ of chloramphenicol until reaching an optical density at 600 nm of 0.6–0.8. After the addition of 1 mM IPTG, the cells were further incubated at 16 °C overnight. The harvested cells were applied to the protein-purification procedure described above. The proteins were collected with a 100-µl volume of glutathione-immobilized resin. The resin bound to the protein was resuspended with lysis buffer to adjust the total volume to 200 µl of the resuspension was separated by SDS-PAGE, which was stained with Coomassie brilliant blue and quantified for protein band density. For each variant, the assay was repeated in three independent cell cultures.

**Degradation assay with human cells.** Human embryonic kidney (HEK293T) cells in 12-well plates were cotransfected with pcDNA3.1-based A3G or sNTD expression vector (0.66 µg) and pcDNA-HVif, pcDNA-HVif H43A Y44A or empty pcDNA3.1 vector (1.32 µg) with FuGENE HD (Promega). After incubation for 24 h, 2 µM MG132 or dimethylsulfoxide was added to the medium, and the cells were further incubated for 24 h. Resulting cell lysates were separated by SDS-PAGE and immunoblotted with mouse monoclonal anti-c-myc antibody (9E10) (Sigma-Aldrich). The protein bands were visualized and scanned with an Odyssey imaging system (LI-COR).

**HIV single-cycle assays.** Single-cycle infectivity experiments were conducted in triplicate as previously described.35,71. HEK293T cells were maintained in Dulbecco’s modified Eagle medium supplemented with 10% FBS and 0.5% pen/strep. Cells were plated at 250,000 cells per well of a six-well plate and 24 h later were transfected (TransIT-LT1; Mirus Bio) with 1 µg of previously reported Vif-proficient or Vif-deficient HIV-1 IIIB proviral expression constructs72 and 0 ng, 25 ng 50 ng or 100 ng of A3G-myc expression constructs. DNA quantities were equalized with empty pcDNA3.1 (Life Technologies). 48 h after transfection, the producer cells and the virus-containing supernatants were harvested. A portion of the virus-containing supernatants was used to infect a CEM-GFP reporter cell line for 48 h, after which infectivity was determined by GFP flow cytometry. The remainder of the virus-containing supernatants was centrifuged through a 20% sucrose cushion to partially purify viral particles for analysis by immunoblotting. Cell and viral pellets were lysed in 2.5× Laemmli reducing sample buffer for immunoblotting. Expression of A3G-myc in cell and viral lysates was detected with an anti-c-Myc antibody (Sigma-Aldrich; C9396), and expression of Vif in cell lysates was detected by an anti-Vif antibody (NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 Vif monoclonal antibody 319 from M.H. Malam). Tubulin and p24 served as loading controls for cell and viral lysates, respectively, and were detected with an anti-α-tubulin antibody (Covance; MMS-407R) and an anti-p24 antibody (NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 p24 hybridoma 183-H12-5C from B. Chesebro). Anti-mouse horseradish peroxidase (HRP)-conjugated secondary antibody (Jackson ImmunoResearch; 715-035-150) was used to detect the mouse primary antibodies, and an anti-rabbit HRP-conjugated secondary antibody (Jackson ImmunoResearch; 111-035-144) was used to detect the rabbit primary antibody. All of the antibodies used for this experiment have been validated, as can be found on the manufacturers’ websites.

**Surface-charge and pocket-volume analysis.** The sNTD NMR structure (model 1) was used as the template to generate the A3G-NTD homology model, with PRIME (Schrödinger), including energy minimization to resolve atomic clashes. Surface electrostatics were calculated in Maestro (Schrödinger). Pocket-volume calculations were performed with SiteMap (Schrödinger).