BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

APOBEC2 Is a Monomer in Solution: Implications for APOBEC3G Models

Troy C. Krzysiak,[†] Jinwon Jung,[†] James Thompson,[‡] David Baker,[‡] and Angela M. Gronenborn*,[†]

[†]Department of Structural Biology, University of Pittsburgh School of Medicine, and Pittsburgh Center for HIV Protein Interactions, 3501 Fifth Avenue, Pittsburgh, Pennsylvania 15261, United States

[‡]Department of Biochemistry, University of Washington, 1959 Northeast Pacific Street, Seattle, Washington 98195, United States

Supporting Information

ABSTRACT: Although the physiological role of APOBEC2 is still largely unknown, a crystal structure of a truncated variant of this protein was determined several years ago [Prochnow, C. (2007) *Nature 445*, 447−451]. This APOBEC2 structure had considerable impact in the HIV field because it was considered a good model for the structure of APOBEC3G, an important HIV restriction factor that abrogates HIV infectivity in the absence of the viral accessory protein Vif. The quaternary structure and the arrangement of the monomers of APOBEC2 in the crystal were taken as being representative for APOBEC3G and exploited in explaining its enzymatic and anti-HIV activity. Here we show, unambiguously, that in contrast to the findings for the crystal, APOBEC2 is monomeric in solution. The nuclear magnetic resonance solution structure of full-length APOBEC2 reveals that the N-terminal tail that was removed for crystallization resides close to strand β2, the dimer interface in the crystal structure, and shields this region of the protein from engaging in intermolecular contacts. In addition, the presence of the N-terminal region drastically alters the aggregation propensity of APOBEC2, rendering the full-length protein highly soluble and not prone to precipitation. In summary, our results cast doubt on all previous structure−function predictions for APOBEC3G that were based on the crystal structure of APOBEC2.

APOBEC2 (apolipoprotein B mRNA editing catalytic-polypeptide-like 2; A2) is a muscle specific family member of the APOBEC/AID (activation-induced deaminase) family of cytidine deaminases.[1] This family includes AID, a protein responsible for C-U deamination at distinct regions in immunoglobin genes, resulting in antibody diversification;[2] APOBEC1, which catalyzes the C-U conversion of cytidine 6666 in apolipoprotein B RNA, leading to the translation of a truncated form of the protein;[3,4] the APOBEC3s (A−H) that play a variety of roles in the restriction of retroelements and retroviruses via C-U deamination of the target genome, leading to genome hypermutation;[5−12] and the newly identified APOBEC4 protein, whose functional role is currently unknown.[13]

A2 is a highly conserved protein among vertebrates[1] whose biological function is largely unknown. In muscle tissue, protein expression is correlated with orderly aging and the maintenance of correct fiber ratios, although gene loss is not lethal.[14−16] There is a growing body of evidence to suggest that A2 may act at the transcriptional level: A2 expression is upregulated during early stem cell differentiation,[17] and expression in *Xenopus* mesoderm is correlated with the orderly development of the left−right body axis.[18] Furthermore, simultaneous overexpression of A2 as well as various APOBEC/AID family members, a thymidine glycosylase, and Gadd45α results in gene demethylation,[19−21] a phenomenon intimately related to transcriptional regulation.[22−25] The coupling of APOBEC-driven methlyated cytidine deamination to thymidine, followed by base excision repair via a thymidine glycosylase, has been proposed as a possible maintenance mechanism.[19−21] Unfortunately, the field is currently ambivalent with regard to the deaminase activity of A2: in vivo studies in higher organisms suggest that A2 deaminates cytidines in DNA,[19,26] whereas *Escherichia coli* mutator studies as well as in vitro deamination assays, using purified recombinant A2 protein, have not been able to detect any deamination activity.[14,27,28]

While numerous APOBEC/AID family members are difficult to purify in a biochemically pure and physically homogeneous form, a truncated version of A2 was purified and crystallized by Prochnow et al.[29] The X-ray structure revealed a novel, extended V-shaped homotetramer, contrasting earlier findings for other deaminases that exhibited square/rectangular homodimeric or homotetrameric assemblies with the Zn-coordinated active sites of all subunits at the center.[30−32] Because little was known about the function of A2, rendering structure−function studies not feasible, the major value of the A2 structure was its utility as a surrogate for other APOBEC/AID family members. All APOBEC/AID family members exhibit a high degree of primary sequence similarity, for one-domain and two-domain variants.[33] Among the APOBECs, A3G has received a great deal of attention given its ability to restrict HIV in a ΔVif background.[5,34] Both monomeric and

Received: January 7, 2012
Revised: February 16, 2012
Published: February 17, 2012

dimeric models of APOBEC3G (A3G)[35,36] were created, based on the A2 structure. In the A2 crystal structure, each wing of the V is comprised of two A2 monomers and thus two deaminase domains. As a result, the majority of researchers in the A3G field have adopted the view that the double domain, monomeric A3G, can be represented by the dimeric wing of the A2 structure and that the functional form, the A3G dimer,[37] can be represented by the A2 tetramer.

In this report, nuclear magnetic resonance (NMR) spectroscopy and light scattering measurements have been used unambiguously to demonstrate that purified A2 is a monomer in solution. This holds true for both the full-length A2 protein and the N-terminal truncation construct that was used for crystallography. In addition, the N-terminal region in full-length A2 contributes to its thermodynamic stability and suppresses aggregation, although it does not alter the fold of the catalytic domain. These findings suggest that, in the absence of a bona fide high-resolution A3G structure, the current views with respect to A3G structure−function relationships that are based on the A2 structure need to be re-examined.

EXPERIMENTAL PROCEDURES

Experimental Conditions. For all experiments unless otherwise noted, the experimental conditions were 20 mM HEPES (pH 7.0), 50 mM NaCl, and 5 mM DTT (A2 buffer) at 25 °C.

Cloning and Purification. Full-length human A2 (amino acids 1−224) [A2(1−224)] and the same truncation construct as the one crystallized by Prochow et al.,[29] A2 (amino acids 41−224) [A2(41−224)], were cloned into a modified pET41a vector and used to transform BL21(DE3) (Invitrogen). For ^{15}N-labeled samples, cells were grown at 37 °C in modified M9 medium, supplemented with [^{15}N]ammonium chloride until the OD_{600} reached 0.8, induced with 500 μM IPTG, and grown at 18 °C for 16−18 h. For triply labeled samples, cells were grown in modified M9 medium, containing [^{15}N]ammonium chloride, [^{13}C]glucose, and 99% D_2O, and induced at an OD_{600} of 0.4. Cells were harvested by centrifugation (4600*g* for 10 min at 4 °C), resuspended in 20 mM HEPES (pH 7.0), 1 M NaCl, and 5 mM DTT, and lysed using a microfluidizer (Microfluidics). DNase (80 μg/mL) and RNase (64 μg/mL) were added to the lysate, and the reaction mixture was incubated at 4 °C with stirring for 2 h. The lysate was clarified by centrifugation (38000*g* for 1 h at 4 °C) and applied to a GSTrap column (GE Life Sciences). Bound protein was eluted with 20 mM HEPES (pH 8.0), 1 M NaCl, and 40 mM reduced glutathione. GST-A2 was separated from any contaminating proteins by gel filtration over a Superdex 200 26/60 column, equilibrated in 20 mM HEPES (pH 7.5), 150 mM NaCl, and 5 mM DTT. Digestion with TEV protease, followed by ion exchange using Q Sepharose, was used to remove the fusion tag. Fusion tag free A2 was finally passed over a Superdex 75 16/60 column, equilibrated in A2 buffer for buffer exchange and to remove the last traces of contaminants. The final protein purity was estimated to be >99% by sodium dodecyl sulfate−polyacrylamide gel electrophoresis. All A2 proteins contain bound zinc based on the analytical method of Kornhaber et al.[38]

NMR Spectroscopy. Backbone chemical shift assignments for full-length A2 and the truncated A2(41−224) were obtained using standard two-dimensional (2D) HSQC and three-dimensional (3D) HNCACB, HN(CO)CACB, HNCA, HN(CO)CA, HNCO, and HN(CA)CO experiments, recorded on a Bruker AVANCE 600 MHz NMR spectrometer, equipped with a *z*-axis gradient cryoprobe at 37 °C. For 3D experiments, 32, 32, 16, 16, 16, and 64 scans were used, respectively. Spectra were processed using NMRPipe[39] and analyzed using CARA.[40] All experiments were performed using 1 mM ^{2}H-, ^{13}C-, and ^{15}N-labeled protein [both A2(1−224) and A2(41−224)] in A2 buffer. Differences in combined ^{1}H and ^{15}N chemical shifts between the ^{1}H−^{15}N HSQC spectra of A2(1−224) and A2(41−224) were calculated according to

$$\Delta\delta = \sqrt{\Delta\delta_{HN}{}^2 + (\Delta\delta_N/6.51)^2}$$

where $\Delta\delta_{HN}$ and $\Delta\delta_N$ represent the ^{1}HN and ^{15}N chemical shift differences, respectively, in parts per million (ppm).[41]

^{1}H−^{15}N residual dipolar couplings (RDCs) were measured for 0.4 mM A2(41−224) in A2 buffer, using either 5% $C_{12}E_5$/hexanol or 15 mg/mL Pf1 for alignment.[42,43] IPAP ^{1}H−^{15}N HSQC spectra in Pf1 were recorded on a Bruker AVANCE 600 MHz NMR spectrometer at 25 °C, equipped with a *z*-axis gradient cryoprobe, while spectra in $C_{12}E_5$/hexanol were recorded on a Bruker AVANCE 700 MHz NMR spectrometer, equipped with a 5 mm triple-resonance, three-axis gradient probe.[44] All spectra were processed using SPARKY version 3.115,[45] and the RDC data were analyzed using PALES.[46] Ninety-two and ninety RDCs were obtained for the $C_{12}E_5$/hexanol and Pf1 samples, respectively, with eight unique to one data set or the other.

To estimate the oligomeric state of A2(41−224), backbone relaxation R_1 and R_2 data as well as ^{1}H−^{15}N heteronuclear NOEs were recorded for 0.4 mM A2(41−224) in A2 buffer on a Bruker AVANCE 700 MHz NMR spectrometer at 25 °C, equipped with a 5 mm triple-resonance, three-axis gradient probe. For R_1, the relaxation delays were 20, 200, 400, 700, 1000, 1500, and 2000 ms. For R_2, the CPMG mixing times were 16, 32, 48, 64, 80, 96, and 112 ms with 2048 × 512 complex points. For error estimation, R_1 experiments with 200 and 400 ms delays and R_2 experiments with 32 and 48 ms mixing times were repeated. For the ^{1}H−^{15}N heteronuclear NOE reference spectrum, a 6 s interscan delay was used, and during the last 3 s of the interscan delay, a 120° pulse train with 5 ms spacing was inserted. The rotational correlation time of A2(41−224), τ_C, was calculated according to

$$\tau_C \approx 1/(4\pi\nu_N)\sqrt{6(R_2/R_1) - 7}$$

where ν_N represents the ^{15}N resonance frequency in hertz.[47] The molecular mass of A2(41−224) was estimated on the basis of the empirical relationship for monomeric proteins:

$$\text{molecular mass} \approx \tau_C/0.6$$

Generation of a Model Structure of Full-Length APOBEC2. A model structure of full-length A2 was generated using HN, N, Cα, and Cβ chemical shifts for 199 of the 214 non-proline residues in A2(1−224) as well as RDCs in the two independent alignment media (Pf1 and $C_{12}E_5$/hexanol) of A2(41−224). The latest version of CS-Rosetta, CS-HM Rosetta (Thompson, Sgourakis, and Baker, personal communication), that incorporates distance constraints from homologous proteins to enhance convergence properties and improves structure determination of proteins up to 25 kDa was employed. The murine A2(46−224) NMR structure [Protein Data Bank (PDB) entry 2RPZ] was utilized as the homology template.

dx.doi.org/10.1021/bi300021s | *Biochemistry* 2012, 51, 2008−2017

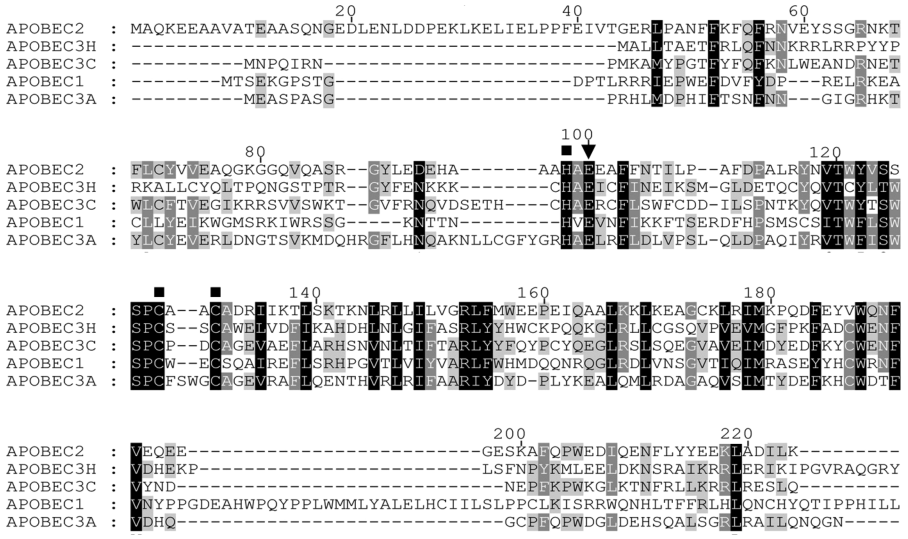


Figure 1. Sequence alignment between A2 and all single-domain APOBEC family members. Amino acid sequences for A2 and A1, A3A, A3C, and A3H were aligned using ClustalX2. The arrowhead denotes the conserved catalytic glutamic acid, while the squares denote the conserved histidine and two cysteines that coordinate the zinc ion at the active site.

Light Scattering Measurements. Size exclusion multi-angle light scattering (SEC-MALS) measurements were performed using a Superdex 200 10/300 column (GE Life Sciences) with in-line multiangle light scattering (HELEOS, Wyatt Technology), UV (Agilent 1100, Agilent Technology), and refractive index (OptilabrEX, Wyatt Technology) detectors. Data were analyzed using ASTRA software, version 5.3.1.4 (Wyatt Technology). In all experiments, 100 μL of protein sample in sterile-filtered, degassed A2 buffer was injected into the system. Dynamic light scattering measurements were performed using a Wyatt Technology DynaPro Plate Reader, and the data were analyzed using Dynamics software, version 7.01. For each protein concentration, the measurements were performed in triplicate with 10 acquisitions per sample.

Differential Scanning Calorimetry (DSC). DSC measurements for both full-length A2 and A2(41−224) were performed using a MicroCal VP-DSC microcalorimeter. For both proteins, 0.5, 1, and 1.5 mg/mL protein in A2 buffer were scanned from 20 to 100 °C with a scan rate of 80 °C/h. The data was analyzed using Origin software, version 7.

RESULTS

Chemical Shift Assignment of APOBEC2. Compared to the other single-domain APOBECs, the amino acid sequence of A2 contains a unique N-terminal extension (Figure 1). Given that the previous structural work was conducted with an N-terminal deletion product of A2, it appeared appropriate to evaluate whether any structural differences exist between the deletion construct and full-length A2. The equivalent set of NMR spectra was collected, namely, 2D ^{1}H−^{15}N HSQC and 3D HNCACB, HN(CO)CACB, HNCA, HN(CA)CO, HN(CO)CA, and HNCO spectra for human full-length A2(1−224) and A2(41−224). Both proteins were soluble at room temperature and below, could be concentrated to >1 mM without immediate aggregation, and yielded well-resolved spectra. A superposition of ^{1}H−^{15}N HSQC spectra for A2(41−224) and A2(1−224) is provided in Figure 2A and reveals nearly perfectly matched resonances in both spectra, with only a few resonances exhibiting small chemical shift differences. Of the non-proline resonances that can be assigned, 89% of 177 were assigned for A2(41−224) and 93% of 214 for A2(1−224). It is interesting to point out the unusual chemical shift for E100, the conserved catalytic glutamic acid. The amide proton resonance for E100, in both spectra (Figure 2A, top left inset), is located at very low field (11.8 ppm), significantly different from the average position for a glutamic acid at 8.34 ppm (BMRB). Such strong deshielding is indicative of hydrogen bonding and seems to be present for the equivalent amide group in all available APOBEC spectra. In murine A2, the E100 amide is located even slightly further downfield at 11.932 ppm (BMRB), and the amide of the corresponding catalytic glutamic acid, E259, in the C-terminal catalytic domain (CTD) of A3G, is located slightly less downfield at 10.955 ppm.[48] The most likely source for this strong deshielding is the carboxyl group of the E100 side chain; no other electron-withdrawing groups are present in the vicinity of the amide proton in either the NMR structure of murine A2 or the available NMR structures of the A3G C-terminal domain.[48−50]

The conformational propensity of the first 40 amino acids of A2 can be qualitatively evaluated from the ^{1}H−^{15}N HSQC spectrum of full-length A2(1−224) (Figure 2A, large left inset). Almost all resonances that belong to residues in this region of the protein are very narrow and cluster around 8 ppm (^{1}H), suggesting that this stretch of polypeptide is highly mobile. Secondary structure prediction with TALOS+[51] also suggests that this region is very flexible and lacks any well-structured elements. Although the absence or presence of the first 40 amino acids of A2 did not significantly alter the ^{1}H−^{15}N HSQC spectrum of the well-folded domain of A2 (see the superposition of the spectra in Figure 2A), indicating that no gross structural changes are induced in the deaminase domain, closer examination of the combined ^{1}H−^{15}N chemical shift differences between the two ^{1}H−^{15}N HSQC spectra revealed a distinct pattern (Figure 2B,C). Resonances of residues with chemical shift differences greater than the average plus one standard deviation are located predominantly along helix α1, strand β2, and loops neighboring these elements. This is not too surprising because helix α1 and the short preceding loop are the structural elements in the A2 crystal structure closest to the

dx.doi.org/10.1021/bi300021s | *Biochemistry* 2012, 51, 2008−2017

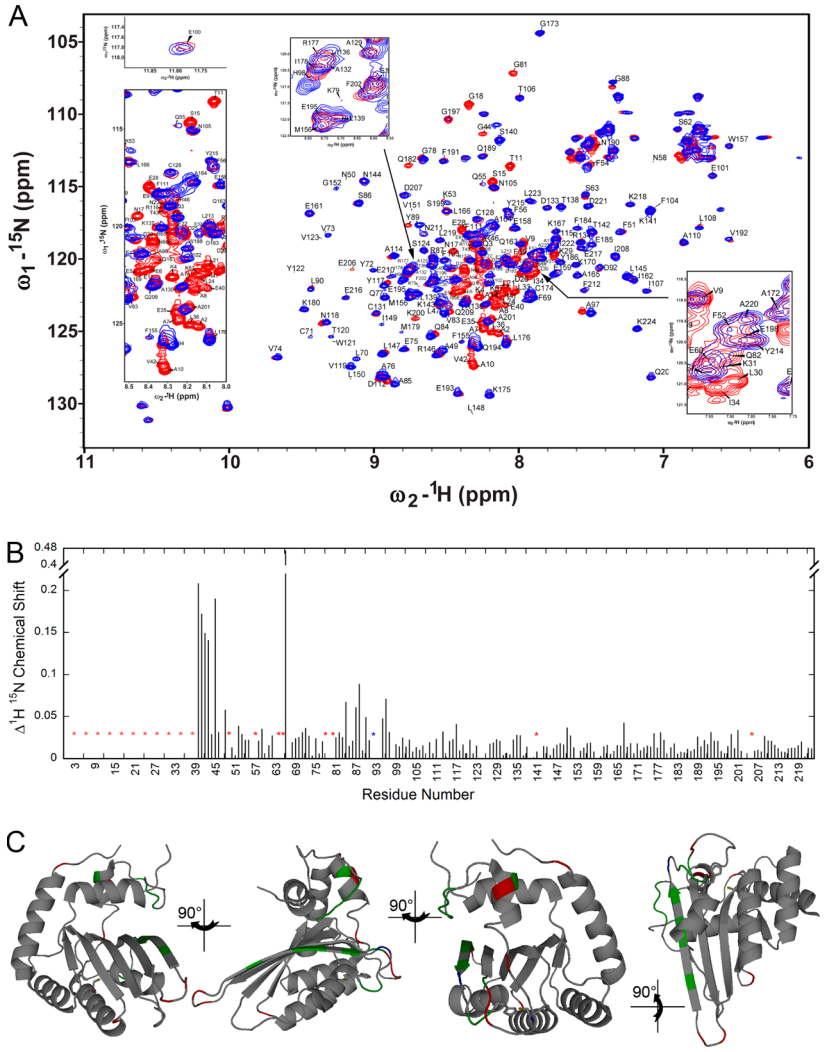


Figure 2. Comparison between full-length and truncated A2. (A) The ^{1}H−^{15}N HSQC spectrum of A2(1−224) (red) is superimposed onto that of A2(41−224) (blue). The amide proton of the putative catalytic E100 exhibits an unusual downfield shift, and its cross-peak is highlighted in the small left-hand corner inset. The amide protons of the first 40 amino acids of A2 produce intense cross-peaks and are all clustered between 8 and 8.5 ppm, indicating that they reside in a flexible, rapidly exchanging region (large left inset). (B) Combined ^{1}H−^{15}N chemical shift differences between A2(1−224) and A2(41−224). Residues assigned in only A2(1−224) are labeled with a red asterisk and those assigned in only A2(41−224) with a blue asterisk. (C) The A2(41−224) crystal structure (PDB entry 2NYT) monomer, residing at the tetramer interface, was used as the template to display the location of residues that were observed in only the A2(1−224) spectrum (red) and the A2(41−224) spectrum (blue). Positions of residues that exhibited a combined chemical shift difference greater than the average +1 standard deviation are shown in green.

N-terminus (residue 41) and would undoubtedly experience different chemical environments in the presence and absence of the N-terminal tail. Interestingly, the differences in shifts observed for the stretch of residues from position 85 to 97, comprising strand β2 and the preceding loop, suggest that the N-terminal tail of A2 interacts with the region that forms the dimerization interface in the A2(41−224) crystal structure. Thus, the first 40 amino acids, while not significantly altering the structure of the deaminase domain, may have a substantial effect on the oligomerization properties of the protein.

APOBEC2 Solution Dynamics. If in solution a similar tetrameric arrangement as in the crystal would have been present, more than one set of resonances should have been present, given the two different chemical environments for the monomeric subunits. For example, the stretch of residues from F59 to G67 is in two distinct conformations in the human A2 crystal structure, forming either a loop at the tetramer interface or a discontinuous β-strand in the monomer that forms the dimer.[29] No indication for such a situation was found in the spectra. However, the most surprising finding was the quality of the spectra in terms of line widths, considering that a dimer or tetramer would possess a molecular mass of 44 or 88 kDa for A2(41−224) and a molecular mass of 52 or 104 kDa for A2(1−224). Proteins of this size generally exhibit very broad resonances and necessitate TROSY-type experiments. Given the unexpected high spectral quality, ^{1}H−^{15}N heteronuclear NOE, R_1, and R_2 relaxation experiments for characterizing the backbone dynamics were conducted. The observed average R_2/R_1 ratios for A2(41−224) yielded a rotational correlation time of 17 ns, compatible with a 28 kDa protein and not a 44 or 88 kDa protein (Figure S1 of the Supporting Information).

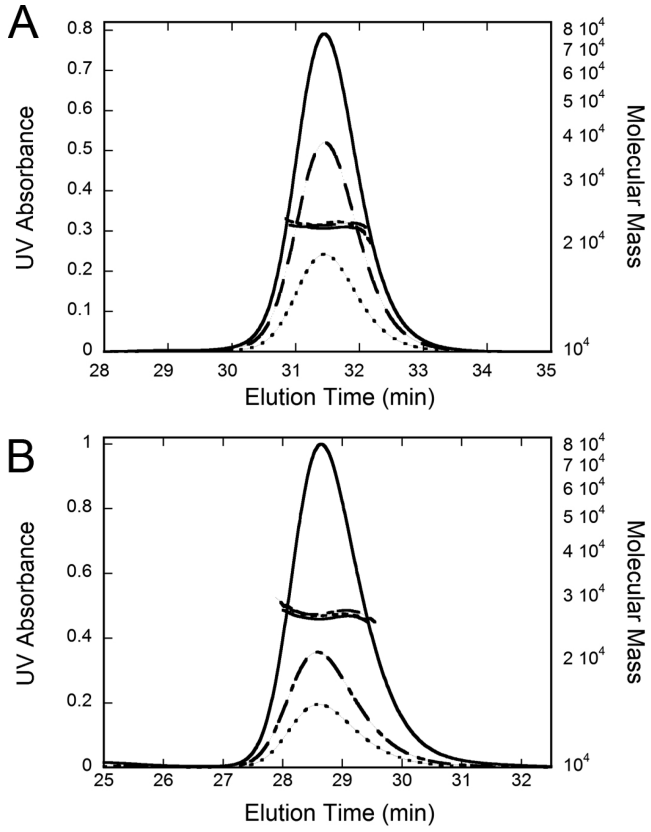
RDCs were measured for the A2 deaminase domain, A2(41−224), using two different alignment media, Pf1[43] and $C_{12}E_5$/hexanol[42] (Figure S2 of the Supporting Information). To evaluate whether the structure of A2(41−224) in solution is identical or very close to monomeric or dimeric units of the

dx.doi.org/10.1021/bi300021s | *Biochemistry* 2012, 51, 2008−2017

tetramer in the crystal, experimental RDCs measured in both media were compared to those calculated on the basis of the different monomer units in the tetramer as well as the dimeric arm of the V-shaped tetramer. The experimental RDCs in both alignment media agreed well with the structure of the central monomer in the tetramer (in Pf1, $R = 0.933$ and $Q_{Saupe} = 0.345$; in $C_{12}E_5$/hexanol, $R = 0.937$ and $Q_{Saupe} = 0.248$). To compare the measured RDCs with those predicted on the basis of the dimeric unit of the A2 crystal structure, the experimental RDC data set was duplicated and compared to those predicted for the dimer or the experimental RDC data set was arbitrarily split in half and compared to the predicted set. Comparison of the measured RDCs in Pf1 revealed that the agreement between experimental and predicted values was inferior to those predicted for the monomer unit ($R = 0.904$ and 0.891, and $Q_{Saupe} = 0.325$ and 0.353). The experimental RDCs measured in $C_{12}E_5$/hexanol exhibited even more profound differences from those predicted on the basis of the dimer, with R values of 0.744 and 0.750 and Q_{Saupe} values of 0.663 and 0.714. Therefore, consistent with the $^1H-^{15}N$ heteronuclear NOE, R_1, and R_2 relaxation data, the RDC measurements also supported the finding that the quaternary structure of A2 in solution is monomeric.

Quaternary State of APOBEC2. The NMR solution dynamics and RDC data eliminated a dimer or tetramer from being the dominant oligomeric state of A2 in solution but could not rule out the possibility that oligomeric species existed in solution. In the $^1H-^{15}N$ HSQC spectra, subpopulations of dimeric and tetrameric species might not have been observed due to (1) the large intensity of the monomer signals, (2) broad peaks of dimeric or tetrameric signals due to the large molecular mass, (3) essentially identical chemical shifts, given the similar chemical environments for equivalent residues in the monomer and putative higher-order oligomers, or (4) all of the above. To unambiguously determine the quaternary state of full-length A2(1−224) and truncated A2(41−224) in solution, multiangle light scattering (SEC-MALS) experiments were performed. The data in Figure 3 demonstrate unequivocally that at protein concentrations of 0.5−2 mg/mL, only a single state is observed for both protein constructs, with molecular masses corresponding to the monomer state for A2(41−224) of 22.4 ± 0.055 kDa ($MW_{theoretical} = 21.7$ kDa) and for A2(1−224) of 26.5 ± 0.026 kDa ($MW_{theoretical} = 26.1$ kDa). We also tested the protein by dynamic light scattering (DLS), and both A2 proteins at concentrations of 1 mg/mL yielded average particle sizes of 2.25 ± 0.04 nm, corresponding to a molecular mass of 24.5 ± 1.08 kDa for A2(41−224), and 2.71 ± 0.03 nm, corresponding to a molecular mass of 34.7 ± 0.9 kDa for A2 (1−224). In both cases, the solutions were monodisperse with peak polydispersity of <10%.

Model Structure of Full-Length APOBEC2. A2 is a highly conserved protein among vertebrates,[52] with the murine and human forms exhibiting 90% sequence identity. The NMR structure of murine A2(46−224) had been determined by the RIKEN Structural Genomics/Proteomics Initiative (PDB entry 2RPZ), and because the N-terminal tail of A2 did not drastically change the chemical shifts of residues in the deaminase domain (Figure 2), it was deemed reasonable to calculate the NMR solution structure of human A2 using CS-HM Rosetta (Thompson, Sgourakis, and Baker, personal communication). CS-HM Rosetta allows for the determination of protein structures up to 25 kDa by combining traditional NMR data, such as chemical shifts, RDCs, and NOE constraints, with



C

	Molecular Radius	Molecular Mass	Polydispersity
A2 41-224	2.25 ± 0.04nm	24.5 ± 1.08kDa	9.08 ± 0.7%
A2 1-224	2.71 ± 0.03nm	34.7 ± 0.9kDa	9.94 ± 0.44%

Figure 3. Light scattering analysis and the oligomeric state of A2. Light scattering, both SEC-MALS and dynamic, was used to determine the oligomeric state of A2(41−224) (A and C) and A2(1−224) (B and C) in solution. SEC-MALS of A2(41−224) (A) and A2(1−224) (B) at injection concentrations of 0.5 (···), 1 (−−−), and 2 mg/mL (—) predicts molecular masses of 22.4 ± 0.055 and 26.5 ± 0.026 kDa, respectively. (C) The average particle size, predicted molecular mass of the two constructs, and polydispersity of the solution from dynamic light scattering are listed.

distance constraints from previously determined structures of homologous proteins. Using the chemical shifts of 199 of the 214 non-proline resonances of full-length A2, RDCs for 90 and 92 residues of A2(41−224) in Pf1 and $C_{12}E_5$/hexanol, respectively, and the mouse 2RPZ structure as the homology template, the solution structure of human A2 was calculated (Figure 4A). In the CS-HM Rosetta A2 model, the N-terminal tail is found as a set of three helices, positioned along the face of the deaminase domain where strand $\beta 2$ resides. As discussed above, the N-terminus is highly mobile and does not adopt a unique, stable conformation. To highlight the flexibility of this region, eight conformers with the lowest energy were selected for its representation (Figure 4B). The helices in all conformers are positioned along only one face of the protein, essentially forming a curtain that shields this surface. Interestingly, this shielded region of the structure constitutes the dimerization interface in the A2(41−224) crystal structure (Figure 4C). As can be appreciated, the position of the N-terminus in the solution structure is incompatible with A2 dimerization via strand $\beta 2$. Furthermore, while A2 is a negatively charged

dx.doi.org/10.1021/bi300021s | Biochemistry 2012, 51, 2008−2017

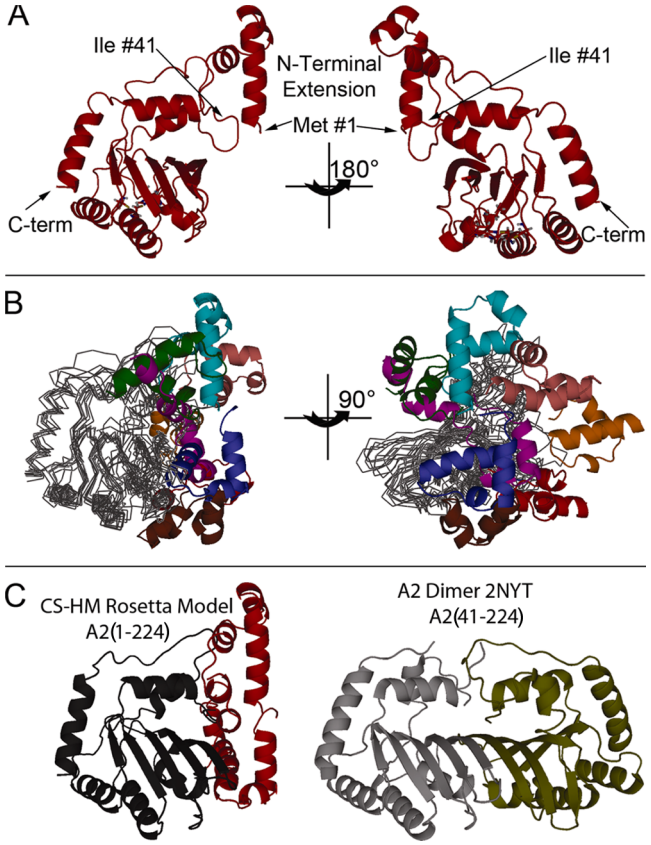


Figure 4. CS-HM Rosetta model of full-length A2. (A) One model selected from the set of lowest-energy structures of A2(1−224). (B) The N-terminal tail is highly flexible, and its positions, as seen in eight of the lowest-energy models, are indicated by the colored cartoons relative to the black ribbon representations of the deaminase domain. (C) The N-termini of three of the lowest-energy models of A2(1−224) are colored red, along with the deaminase domain (black). The A2(41−224) dimer of the crystal structure (PDB entry 2NYT) is provided for comparison. The monomer residing at the tetramer interface in the crystal structure is colored gray and that at the dimer interface is shown in olive.

protein, with a predicted charge of −13 at pH 7.0, the N-terminal extension is exceptionally rich in acidic residues (−9 at pH 7.0), and dimerization similar to that seen in the A2 crystal structure would bring two highly negative regions together, clearly an unfavorable electrostatic arrangement.

Although CS-HM Rosetta models the N-terminus as three helices, those helices are highly fluctuating. Backbone Cα chemical shifts generally serve as a good predictor of the helicity of a polypeptide region.[53] Upon comparison of the Cα chemical shift for a particular residue with that found in a random coil conformation, a chemical shift difference of >0.7 ppm indicates that this residue exhibits helical ϕ and ψ angles. On the basis of the analysis of Cα chemical shifts exhibited by the N-terminus of A2, we note that no stable helical segments are suggested by these data (Figure S3 of the Supporting Information). Thus, the varying conformations for the N-terminal tail predicted by CS-HM Rosetta also indicate a high degree of flexibility. Indeed, the identity of the helical residues varies in each model, suggesting that particular amino acids exchange between unique ϕ and ψ angles. Although $^{1}H-^{15}N$ HSQC spectra of A2(1−224) strongly suggest that the N-terminus is predominatly in a random coil conformation, we were able to observe most amide resonances, even at 37 °C, suggesting some degree of protection from exchange. Missing resonances in the $^{1}H-^{15}N$ HSQC NMR spectra of A2(41−224), however, often were detectable in the NMR spectra of A2(1−224) (Figure 2B), with the associated residues found predominantly in helical stretches proximal to the strand β2 face of the protein. Transient H-bonding as well as steric occlusion may play a role in this difference. Therefore, taken together, these data suggest that the N-terminus of A2 exhibits a highly fluctuating structure that may contain nascent helices as modeled by CS-HM Rosetta.

As expected, the deaminase domain of the model of full-length A2 is nearly identical to that of an individual domain in the human A2(41−224) crystal structure and the mouse A2(46−224) solution NMR structure, although two areas of distinct differences can be discerned (Figure 5). In particular, the long loop connecting helix α1 with strand β1 exhibits different conformations (Figure 5D). In the murine structure, the loop points away from the putative active site, while in the crystal structure of human A2, this loop obscures the active site. In the model of full-length A2, this loop is also close to the active site, but like the N-terminus, this region is conformationally highly variable and is either covering the active site or leaving it completely exposed. This region also exhibits two different conformations in the crystal structure. In the monomers that are involved in the dimer interface, this stretch forms a short β-strand (Figure 5C,D), while it is a loop in those monomers that are involved in forming the tetramer interface (Figure 5B,D). This difference was attributed to the destabilization of the β-strand through contacts in the tetramer interface.[29] In our solution structure, this region is clearly disordered. A second difference pertains to helix α1. Like in the murine NMR solution structure, our model shows this helix at a 76° angle with respect to helix α1 of the crystal structure, nearly perpendicular to the rest of the fold. Because this helix is the first structured element in the deaminase domain, it could act as a lever for positioning of N-terminal tail with respect to the overall structure (Figure 5D).

Stability of A2(41−224) and A2(1−224). During the course of the collection of NMR data, there was a distinct difference in the long term behavior of full-length A2 as compared to that of the deletion construct. A2(41−224) readily precipitated at 37 °C, while no precipitate was seen in the A2(1−224) samples, even after 1 week. In principle, precipitation can occur via nonspecific aggregation of the folded or unfolded protein. We, therefore, assessed the thermodynamic stability of both proteins by DSC (Figure 6). Thermal unfolding of both proteins was not reversible, and full-length A2 exhibited a T_m of 63.77 ± 0.01 °C. The deletion construct, on the other hand, exhibited a complex thermal melting pattern with an apparent early unfolding event (T_m ∼ 52 °C) and visible precipitation. Therefore, reliable thermodynamic parameters could not be extracted for A2(41−224). Precipitation of A2(41−224) also occurred during the NMR experiments, and the structure that was determined represents only that of the soluble fraction of this protein.

DISCUSSION

Implication of the A2 Model on A3G. The greatest impact of the A2 crystal structure has been on predictions of the A3G structure, a two-domain member of the APOBEC/AID family.[29,35,36] Prior to the A2 crystal structure, all other known tetrameric cytidine deaminase structures were square-

dx.doi.org/10.1021/bi300021s | Biochemistry 2012, 51, 2008−2017

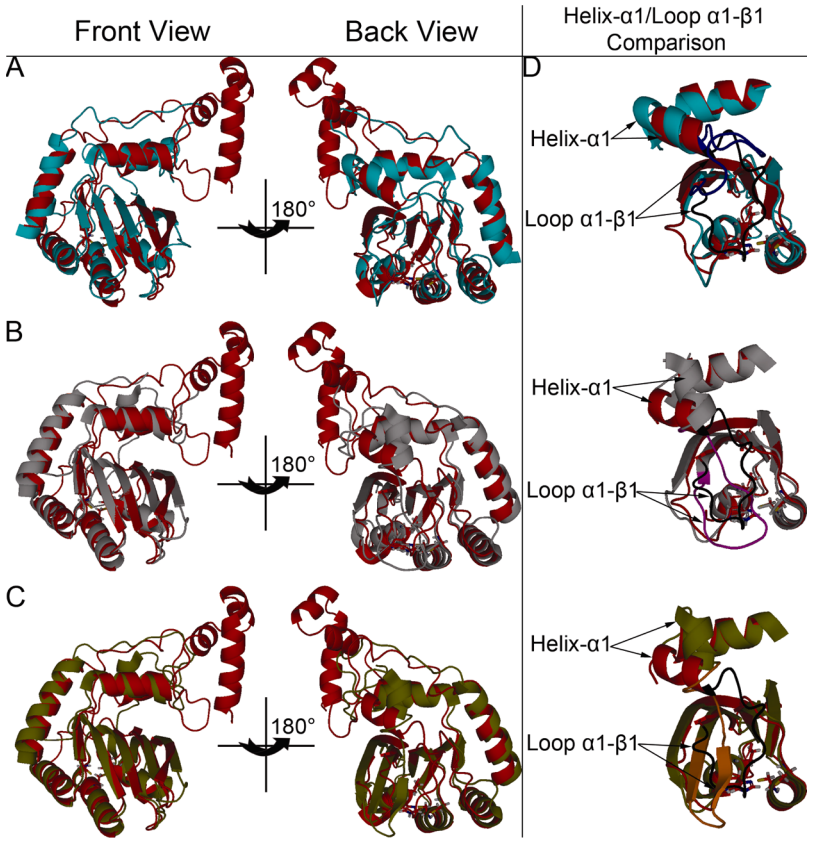


Figure 5. Comparison of the model of full-length A2 with other A2 structures. The CS-HM Rosetta model of A2(1−224) (red) is aligned with (A) the RIKEN murine A2(46−224) NMR structure (PDB entry 2RPZ) (cyan), (B) the A2 monomer comprising the tetramer interface of the human A2(41−224) crystal structure (PDB entry 2NYT) (gray), and (C) the A2(41−224) monomer not residing at the tetramer interface (olive). (D) For each pair, the similarities and differences in helix α1 and loop $\alpha 1\beta 1$ are highlighted. For the sake of clarity, the $\alpha 1-\beta 1$ loops are colored differently [black for A2(1−224), purple for the tetramer interface of the A2(41−224) monomer, and orange for the dimer interface of the A2(41−224) monomer]. In all structures, the catalytic E100 and the Zn-coordinating residues H98, C128, and C131 are shown as sticks.

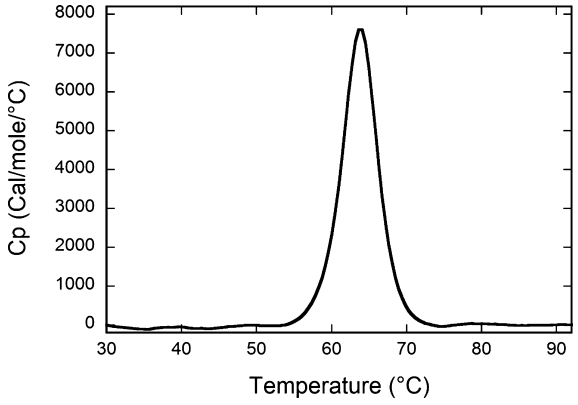


Figure 6. DSC analysis of full-length A2. A2(1−224) (1.5 mg/mL) in A2 buffer was heated from 20 to 100 °C. The protein A2(1−224) displays a single melting transition with a T_m of 63.77 ± 0.01 °C.

shaped, with each zinc-coordinating active site located at the center of the molecule.[30−32] The A2 tetramer displayed a novel type of oligomerization in which two monomers pair up via their β2 strands, forming a dimer with a contiguous β-sheet, and two dimers engage in head-to-head interactions via helix α6, the AC loop and loop 7, resulting in an elongated V-shaped structure. In A3G, the N-terminal nucleic acid binding domain and the C-terminal catalytic deaminase domain are very similar in sequence and have arisen through gene duplication.[33] The intimate nature of the A2 dimer comprised of a central β-sheet was taken as evidence that a single A3G molecule could undoubtedly be represented by this arrangement of A2 monomers. Furthermore, while there still is debate in the APOBEC field regarding the oligomeric state of the functional A3G molecule,[54] the A3G dimer is generally regarded as the functional unit.[37] Although this structure is compelling, the major problem with this notion has always been that in a V-shaped structure for dimeric A3G, the catalytic deaminase domains point away from each other and from the nucleic acid binding site. This creates a topological dilemma, and although RNA binding and deamination are functionally separated on two different domains,[55−58] the extended A2 structure implied for A3G that a more complex way for organizing nucleic acids between seemingly independent active sites would be necessary. A more recent model by Harjes et al.[49] proposed a slightly different type of fold for the A3G monomer that positions the active sites on opposite faces of the monomer. However, all current full-length A3G models are still speculative in the absence of an experimentally determined full-length A3G structure.

Our finding that A2 exists as a monomer erodes the basis for the previous assumption that double-domain APOBEC family members exhibit the organization found in the A2 crystal. Both the truncated form used for crystallization, A2(41−224), and full-length A2 are monomeric in solution, and even at a concentration of 1 mM, soluble A2 was not seen to

dx.doi.org/10.1021/bi300021s | *Biochemistry* 2012, 51, 2008−2017

oligomerize. Full-length A2 did not exhibit any tendency to aggregate or precipitate, even throughout our prolonged data collection. Visible precipitation was not observed after several months at 4 °C, or for the NMR sample that was in the spectrometer at 37 °C for 1 week. Even after thermal unfolding, it seems that the unfolded polypeptide is also soluble. This is in stark contrast to the behavior seen for the A2(41−224) truncation. The majority of this protein precipitates after a few minutes in the spectrometer at 37 °C or storage at 4 °C for 1 month, and thermal unfolding results in quantitative precipitation.

Extensive intermolecular contacts between the β2 strands of two A2(41−224) monomers in the crystal[29] suggested that this face of the protein might serve as a nucleation point for aggregation. Our model of the physiologically relevant full-length protein indicates that the interface at strand β2 is not readily accessible to other A2 molecules. The flexible A2 N-terminal tail, although not in a unique conformation, clearly shields the entire face of the protein in which strand β2 resides (Figure 4B). In addition, the preponderance of negatively charged amino acids, located at the N-terminus of A2, is expected to cause significant electrostatic repulsion between A2 molecules, preventing dimerization via strand β2. Ultimately, only the determination of a high-resolution structure of A3G or another double-domain APOBEC, preferably with a subsrate mimic, will reveal how such molecules orient the two domains to conduct cytidine deamination.

Structural Regulation of Deamination. Comparing the full-length A2 model with the other APOBEC structures suggests that active deamination by APOBEC/AID family members involves at least two regulatory features. (1) A2 possesses a long loop that connects helix α1 to strand β1. In the A2 crystal structure, this loop conceals the catalytic E100; however, in the murine A2 NMR structure, this loop is located far from the active site. In our full-length human A2 structure, this loop is flexible and can either cover the active site or not cover it. The homologous loop in A3G is also seen in varying positions in all known structures.[48,49,59−61] Thus, different conformations of this loop clearly could influence accessibility to the active site and thereby regulate catalytic activity. (2) The amide resonances of the catalytic E100 in A2 (Figure 2A), the catalytic E259 in A3G,[48] and the catalytic E72 in A3A (I.-J. Byeon, private communication) display unusual downfield shifts in the proton dimension, indicative of their possible involvement in hydrogen bonding. The relevant H-bond is most likely one that involves its own side chain carboxylate group and may keep this glutamic acid carboxylate in an inactive state. Substrate binding could trigger release of the carboxylate side chain from this inactive conformation, rendering it capable of engaging in hydrogen transfer. In that regard, it is interesting to point out that in *E. coli* mutator assays, A2 has not been found to be active as a cytidine deaminase, with A3G a relatively poor deaminase, compared to other family members, such as A1 and A3A.[27,28] It will be interesting to see whether the degree of deshielding of the catalytic glutamic acid amide resonance correlates with the robustness of deaminase activity.

Potential Implications for A2 Transcriptional Regulation. Finally, while the physiological role of A2 has yet to be identified, its potential involvement in transcriptional activation is intriguing. The existing model suggests that a cytidine deaminase converts 5′-methylcytidine to thymine, such that in a subsequent step a glycosylase can recognize the DNA mismatch and excise the incorrect base.[19,21] The abasic site is then repaired by the rest of the excision repair (BER) machinery.[62,63] Interestingly, direct interaction between the glycosylase and the deaminase has been suggested on the basis of immunoprecipitation,[19,20] with the thymidine glycosylase, MBD4, implicated in this process.[19] MBD4 is a well-structured, positively charged protein,[64] and it is tempting to speculate that the flexible, highly negatively charged, N-terminal tail of A2 could be involved in the interaction. In addition, the fact that A2 deaminates 5′-methylcytidine implies that its substrate is double-stranded DNA, and not single-stranded DNA, as is the case for A3G. In this situation, a monomeric structure would be easier to position on the DNA duplex for successful catalysis. Further structural studies involving substrate may shed light on this question.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

NMR relaxation and RDC data for A2(41−224) and a comparison of the secondary Cα chemical shifts. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Department of Structural Biology, University of Pittsburgh School of Medicine, 3501 Fifth Ave., Pittsburgh, PA 15260. Phone: (412) 648-9959. Fax: (412) 648-9008. E-mail: amg100@pitt.edu.

Funding

This work was supported by National Institute of General Medical Sciences Grants P50GM082251 to A.M.G. and F32 GM087138 to T.C.K.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Jinwoo Ahn and In-Ja Byeon for useful discussions and Mike Delk for NMR technical support.

■ ABBREVIATIONS

AID, activation-induced deaminase; APOBEC2/A2, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2; APOBEC3G/A3G, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G; BMRB, Biological Magnetic Resonance Bank; CPMG, Carr−Purcell−Meiboom−Gill pulse sequence; CTD, C-terminal domain; DLS, dynamic light scattering; DSC, differential scanning calorimetry; DTT, 1,4-dithio-D-threitol; GST, glutathione *S*-transferase; HSQC, heteronuclear single-quantum coherence transfer; MBD4, methyl-CpG binding domain protein 4; NOE, nuclear Overhauser effect; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; RDC, residual dipolar coupling; SEC-MALS, tandem size exclusion multiangle light scattering; TEV, catalytic domain of nuclear inclusion a protein from tobacco etch virus; TROSY, transverse relaxation-optimized spectroscopy.

■ REFERENCES

(1) Liao, W., Hong, S. H., Chan, B. H., Rudolph, F. B., Clark, S. C., and Chan, L. (1999) APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochem. Biophys. Res. Commun. 260*, 398−404.

(2) Neuberger, M. S., Harris, R. S., Di Noia, J., and Petersen-Mahrt, S. K. (2003) Immunity through DNA deamination. *Trends Biochem. Sci. 28*, 305−312.
(3) Teng, B., Burant, C. F., and Davidson, N. O. (1993) Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science 260*, 1816−1819.
(4) Navaratnam, N., Morrison, J. R., Bhattacharya, S., Patel, D., Funahashi, T., Giannoni, F., Teng, B. B., Davidson, N. O., and Scott, J. (1993) The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J. Biol. Chem. 268*, 20709−20712.
(5) Sheehy, A. M., Gaddis, N. C., Choi, J. D., and Malim, M. H. (2002) Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature 418*, 646−650.
(6) Bishop, K. N., Holmes, R. K., Sheehy, A. M., Davidson, N. O., Cho, S. J., and Malim, M. H. (2004) Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr. Biol. 14*, 1392−1396.
(7) Zheng, Y. H., Irwin, D., Kurosu, T., Tokunaga, K., Sata, T., and Peterlin, B. M. (2004) Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. *J. Virol. 78*, 6073−6076.
(8) Dang, Y., Wang, X., Esselman, W. J., and Zheng, Y. H. (2006) Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. *J. Virol. 80*, 10522−10533.
(9) Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G. G., and Munk, C. (2006) APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J. Biol. Chem. 281*, 22161−22172.
(10) Baumert, T. F., Rosler, C., Malim, M. H., and von Weizsacker, F. (2007) Hepatitis B virus DNA is subject to extensive editing by the human deaminase APOBEC3C. *Hepatology 46*, 682−689.
(11) Wiegand, H. L., and Cullen, B. R. (2007) Inhibition of alpharetrovirus replication by a range of human APOBEC3 proteins. *J. Virol. 81*, 13694−13699.
(12) OhAinle, M., Kerns, J. A., Li, M. M., Malik, H. S., and Emerman, M. (2008) Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe 4*, 249−259.
(13) Rogozin, I. B., Basu, M. K., Jordan, I. K., Pavlov, Y. I., and Koonin, E. V. (2005) APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle 4*, 1281−1285.
(14) Mikl, M. C., Watt, I. N., Lu, M., Reik, W., Davies, S. L., Neuberger, M. S., and Rada, C. (2005) Mice deficient in APOBEC2 and APOBEC3. *Mol. Cell. Biol. 25*, 7270−7277.
(15) Piec, I., Listrat, A., Alliot, J., Chambon, C., Taylor, R. G., and Bechet, D. (2005) Differential proteome analysis of aging in rat skeletal muscle. *FASEB J. 19*, 1143−1145.
(16) Sato, Y., Probst, H. C., Tatsumi, R., Ikeuchi, Y., Neuberger, M. S., and Rada, C. (2010) Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *J. Biol. Chem. 285*, 7111−7118.
(17) Pennings, J. L., van Dartel, D. A., Pronk, T. E., Hendriksen, P. J., and Piersma, A. H. (2011) Identification by gene coregulation mapping of novel genes involved in embryonic stem cell differentiation. *Stem Cells Dev. 20*, 115−126.
(18) Vonica, A., Rosa, A., Arduini, B. L., and Brivanlou, A. H. (2011) APOBEC2, a selective inhibitor of TGFβ signaling, regulates left-right axis specification during early embryogenesis. *Dev. Biol. 350*, 13−23.
(19) Rai, K., Huggins, I. J., James, S. R., Karpf, A. R., Jones, D. A., and Cairns, B. R. (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell 135*, 1201−1212.
(20) Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., Abramowitz, L. K., Bartolomei, M. S., Rambow, F., Bassi, M. R., Bruno, T., Fanciulli, M., Renner, C., Klein-Szanto, A. J., Matsumoto, Y., Kobi, D., Davidson, I., Alberti, C., Larue, L., and Bellacosa, A. (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell 146*, 67−79.
(21) Guo, J. U., Su, Y., Zhong, C., Ming, G. L., and Song, H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell 145*, 423−434.
(22) Frank, D., Keshet, I., Shani, M., Levine, A., Razin, A., and Cedar, H. (1991) Demethylation of CpG islands in embryonic cells. *Nature 351*, 239−241.
(23) Jost, J. P., Oakeley, E. J., Zhu, B., Benjamin, D., Thiry, S., Siegmann, M., and Jost, Y. C. (2001) 5-Methylcytosine DNA glycosylase participates in the genome-wide loss of DNA methylation occurring during mouse myoblast differentiation. *Nucleic Acids Res. 29*, 4452−4461.
(24) Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R. Z., Carmouche, R. P., Ibberson, D., Barath, P., Demay, F., Reid, G., Benes, V., Jeltsch, A., Gannon, F., and Salbert, G. (2008) Cyclical DNA methylation of a transcriptionally active promoter. *Nature 452*, 45−50.
(25) Song, F., Mahmood, S., Ghosh, S., Liang, P., Smiraglia, D. J., Nagase, H., and Held, W. A. (2009) Tissue specific differentially methylated regions (TDMR): Changes in DNA methylation during development. *Genomics 93*, 130−139.
(26) Okuyama, S., Marusawa, H., Matsumoto, T., Ueda, Y., Matsumoto, Y., Endo, Y., Takai, A., and Chiba, T. (2011) Excessive activity of apolipoprotein B mRNA editing enzyme catalytic polypeptide 2 (APOBEC2) contributes to liver and lung tumorigenesis. *Int. J. Cancer 130*, 1294−1301.
(27) Harris, R. S., Petersen-Mahrt, S. K., and Neuberger, M. S. (2002) RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell 10*, 1247−1253.
(28) Lada, A. G., Krick, C. F., Kozmin, S. G., Mayorov, V. I., Karpova, T. S., Rogozin, I. B., and Pavlov, Y. I. (2011) Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. *Biochemistry (Moscow, Russ. Fed.) 76*, 131−146.
(29) Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F., and Chen, X. S. (2007) The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature 445*, 447−451.
(30) Kumasaka, T., Yamamoto, M., Furuichi, M., Nakasako, M., Teh, A. H., Kimura, M., Yamaguchi, I., and Ueki, T. (2007) Crystal structures of blasticidin S deaminase (BSD): Implications for dynamic properties of catalytic zinc. *J. Biol. Chem. 282*, 37103−37111.
(31) Teh, A. H., Kimura, M., Yamamoto, M., Tanaka, N., Yamaguchi, I., and Kumasaka, T. (2006) The 1.48 Å resolution crystal structure of the homotetrameric cytidine deaminase from mouse. *Biochemistry 45*, 7825−7833.
(32) Randau, L., Stanley, B. J., Kohlway, A., Mechta, S., Xiong, Y., and Soll, D. (2009) A cytidine deaminase edits C to U in transfer RNAs in Archaea. *Science 324*, 657−659.
(33) Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., and Navaratnam, N. (2002) An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics 79*, 285−296.
(34) Lecossier, D., Bouchonnet, F., Clavel, F., and Hance, A. J. (2003) Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science 300*, 1112.
(35) Zhang, K. L., Mangeat, B., Ortiz, M., Zoete, V., Trono, D., Telenti, A., and Michielin, O. (2007) Model structure of human APOBEC3G. *PLoS One 2*, e378.
(36) Chelico, L., Prochnow, C., Erie, D. A., Chen, X. S., and Goodman, M. F. (2010) Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. *J. Biol. Chem. 285*, 16195−16205.
(37) Chelico, L., Pham, P., Calabrese, P., and Goodman, M. F. (2006) APOBEC3G DNA deaminase acts processively 3′ → 5′ on single-stranded DNA. *Nat. Struct. Mol. Biol. 13*, 392−399.
(38) Kornhaber, G. J., Snyder, D., Moseley, H. N., and Montelione, G. T. (2006) Identification of zinc-ligated cysteine residues based on $^{13}C\alpha$ and $^{13}C\beta$ chemical shift data. *J. Biomol. NMR 34*, 259−269.
(39) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277−293.

(40) Keller, R. L. J. (2004) *The Computer Aided Resonance Assignment Tutorial*, CANTINA Verlag, Goldau, Switzerland.
(41) Mulder, F. A., Schipper, D., Bott, R., and Boelens, R. (1999) Altered flexibility in the substrate-binding site of related native and engineered high-alkaline *Bacillus* subtilisins. *J. Mol. Biol. 292*, 111−123.
(42) Ruckert, M., and Otting, G. (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc. 122*, 7793−7797.
(43) Hansen, M. R., Mueller, L., and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol. 5*, 1065−1074.
(44) Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson. 131*, 373−378.
(45) Goddard, T. D., and Kneller, D. G. (2004) *SPARKY*, version 3.3.110, University of California, San Francisco.
(46) Zweckstetter, M. (2008) NMR: Prediction of molecular alignment from structure using the PALES software. *Nat. Protoc. 3*, 679−690.
(47) Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry 28*, 8972−8979.
(48) Chen, K. M., Harjes, E., Gross, P. J., Fahmy, A., Lu, Y., Shindo, K., Harris, R. S., and Matsuo, H. (2008) Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature 452*, 116−119.
(49) Harjes, E., Gross, P. J., Chen, K. M., Lu, Y., Shindo, K., Nowarski, R., Gross, J. D., Kotler, M., Harris, R. S., and Matsuo, H. (2009) An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *J. Mol. Biol. 389*, 819−832.
(50) Furukawa, A., Nagata, T., Matsugami, A., Habu, Y., Sugiyama, R., Hayashi, F., Kobayashi, N., Yokoyama, S., Takaku, H., and Katahira, M. (2009) Structure and real-time monitoring of the enzymatic reaction of APOBEC3G which is involved in anti-HIV activity. *Nucleic Acids Symp. Ser.*, 87−88.
(51) Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR 44*, 213−223.
(52) Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K., and Neuberger, M. S. (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol. Biol. Evol. 22*, 367−377.
(53) Wishart, D. S., and Sykes, B. D. (1994) The ^{13}C chemical-shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR 4*, 171−180.
(54) McDougall, W. M., Okany, C., and Smith, H. C. (2011) Deaminase activity on ssDNA occurred in vitro when APOBEC3G forms homotetramers and higher-order complexes. *J. Biol. Chem. 286*, 30655−30661.
(55) Hache, G., Liddament, M. T., and Harris, R. S. (2005) The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *J. Biol. Chem. 280*, 10920−10924.
(56) Navarro, F., Bollman, B., Chen, H., Konig, R., Yu, Q., Chiles, K., and Landau, N. R. (2005) Complementary function of the two catalytic domains of APOBEC3G. *Virology 333*, 374−386.
(57) Iwatani, Y., Takeuchi, H., Strebel, K., and Levin, J. G. (2006) Biochemical activities of highly purified, catalytically active human APOBEC3G: Correlation with antiviral effect. *J. Virol. 80*, 5992−6002.
(58) Chen, K. M., Martemyanova, N., Lu, Y., Shindo, K., Matsuo, H., and Harris, R. S. (2007) Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G. *FEBS Lett. 581*, 4761−4766.
(59) 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) Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature 456*, 121−124.
(60) Furukawa, A., Nagata, T., Matsugami, A., Habu, Y., Sugiyama, R., Hayashi, F., Kobayashi, N., Yokoyama, S., Takaku, H., and Katahira, M. (2009) Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO J. 28*, 440−451.
(61) Shandilya, S. M., Nalam, M. N., Nalivaika, E. A., Gross, P. J., Valesano, J. C., Shindo, K., Li, M., Munson, M., Royer, W. E., Harjes, E., Kono, T., Matsuo, H., Harris, R. S., Somasundaran, M., and Schiffer, C. A. (2010) Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure 18*, 28−38.
(62) Ma, D. K., Guo, J. U., Ming, G. L., and Song, H. (2009) DNA excision repair proteins and Gadd45 as molecular players for active DNA demethylation. *Cell Cycle 8*, 1526−1531.
(63) Hajkova, P., Jeffries, S. J., Lee, C., Miller, N., Jackson, S. P., and Surani, M. A. (2010) Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science 329*, 78−82.
(64) Zhang, W., Liu, Z., Crombet, L., Amaya, M. F., Liu, Y., Zhang, X., Kuang, W., Ma, P., Niu, L., and Qi, C. (2011) Crystal structure of the mismatch-specific thymine glycosylase domain of human methyl-CpG-binding protein MBD4. *Biochem. Biophys. Res. Commun. 412*, 425−428.