Published in final edited form as:

Biochemistry. 2012 March 6; 51(9): 2008–2017. doi:10.1021/bi300021s.

APOBEC2 is a Monomer in Solution: Implications for APOBEC3G Models

Troy C. Krzysiak¹, Jinwon Jung¹, James Thompson², David Baker², and Angela M. Gronenborn^{1,*}

¹Department of Structural Biology, University of Pittsburgh School of Medicine, and Pittsburgh Center for HIV Protein Interactions, 3501 Fifth Ave., Pittsburgh, PA 15261, United States

²Department of Biochemistry, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195, United States

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 since 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 was taken as representative for APOBEC3G and exploited for explaining its enzymatic and anti-HIV activity. Here we show, unambiguously, that in contrast to the findings in the crystal, APOBEC2 is monomeric in solution. The NMR solution structure of full-length APOBEC2 reveals that the N-terminal tail that was removed for crystallization, resides close to strand β 2, the dimer interface in the crystal structure, and shields this region of the protein from engaging in inter-molecular 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 onto all previous structure/function predictions for APOBEC3G that were based on the APOBEC2 crystal structure.

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 CU deamination at distinct regions in immunoglobulin 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, B, C, DE, F, G, and 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

*Address correspondence to Angela M. Gronenborn, Department of Structural Biology, University of Pittsburgh Medical School, 3501 5th Ave, Pittsburgh PA 15260 Phone#: 412-648-9959 Fax#: 412-648-9008 amg100@pitt.edu.

Supporting Information Available:

The supporting information contains NMR relaxation and RDC data for A2 (41–224), as well as a comparison of the secondary C_{α} chemical shifts.

This material is available free of charge via the internet at <http://pubs.acs.org>.

maintenance of correct fiber ratios, although gene loss is not lethal (14-16). There is increasing 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 Gadd45 α results in gene demethylation (19-21), a phenomenon intimately related to transcriptional regulation (22-25). The coupling of APOBEC-driven methylated cytidine deamination to thymidine, followed by base excision repair via a thymidine glycosylase has been proposed as a possible mechanism (19-21). Unfortunately, the field is currently ambivalent regarding the deaminase activity of A2: *in vivo* studies in higher organisms suggest that A2 deaminates cytidines in DNA (19, 26) whereas *E. 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). Since little was known about the function of A2, rendering structure-function studies unfeasible, 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 as well as two-domain variants (33). Among the APOBECs, A3G has received a great deal of attention given its ability to restrict HIV in a Δ Vif background (5, 34). Both, monomeric as well as 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 deaminase 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 spectroscopy (NMR) and light scattering measurements have been used unambiguously demonstrate that purified A2 is a monomer in solution. This holds true for both the full-length A2 protein as well as 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 20mM HEPES, pH 7.0, 50mM NaCl, 5mM DTT (A2 buffer) at 25°C.

Cloning and Purification

Full-length human A2 (aa1-224) and the same truncation construct as the one crystallized by Prochow et al. (29), A2 (aa41-224), were cloned into a modified pET41a vector and used to transform BL21(DE3)* cell line (Invitrogen). For ^{15}N labeled samples, cells were grown at 37°C in modified M9 medium, supplemented with ^{15}N -ammonium chloride till $\text{OD}_{600} = 0.8$,

induced with 500 μ M IPTG and grown at 18°C for 16-18hrs. For triple labeled samples, cells were grown in modified M9 medium, containing 15 N-ammonium chloride, 13 C-glucose, and 99% D₂O and induced at OD₆₀₀= 0.4. Cells were harvested by centrifugation (4600 $\times$; 10min; 4°C), resuspended in 20mM HEPES pH 7.0, 1M NaCl, 5mM DTT and lysed using a microfluidizer (Microfluidics). DNase (80 μ g/ml) and RNase (64 μ g/ml) were added to the lysate and the reaction was incubated at 4°C with stirring for 2 hours. The lysate was clarified by centrifugation (38,000 $\times$; 1hr/ 4°C) and applied to a GStrap column (GE Life Sciences). Bound protein was eluted with 20mM HEPES pH 8.0, 1M NaCl, 40 mM reduced glutathione. GST-A2 was separated from any contaminating proteins by gel filtration over a Superdex 200 26/60 column, equilibrated in 20mM HEPES, pH 7.5, 150mM NaCl, 5mM 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. Final protein purity was estimated at >99% by SDS-PAGE. 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 2D HSQC and 3D HNCACB, HN(CO)CACB, HNCA, HN(CO)CA, HNCO and HN(CA)CO experiments, recorded on a Bruker AVANCE 600MHz 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 1mM ^2H , ^{13}C , ^{15}N labeled protein (both A2 (1-224) and A2 (41-224)) in A2 buffer. Differences in combined ^1H and ^{15}N chemical shifts between the ^1H - ^{15}N HSQC spectra of A2 (1-224) and A2 (41-224) were calculated according to:

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

with $\Delta\delta_{HN}$ and $\Delta\delta_N$ representing the ^1HN and ^{15}N chemical shift differences in ppm, respectively (41).

^1H - ^{15}N residual dipolar couplings (RDC) were measured for 0.4mM A2 (41-224) in A2 buffer using either 5% C12E5/hexanol or 15mg/ml Pf1 for alignment (42, 43). IPAP ^1H - ^{15}N HSQC spectra in Pf1 were collected on a Bruker AVANCE 600MHz NMR spectrometer at 25°, equipped with a z-axis gradient cryoprobe, while spectra in C12E5/hexanol were recorded on a Bruker AVANCE 700MHz NMR spectrometer, equipped with a 5mm triple-resonance, three-axes gradient probe (44). All spectra were processed using SPARKY V3.115 software (45) and the RDC data were analyzed using PALES (46). Ninety-two and ninety RDCs were obtained for the C12E5/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 ^1H - ^{15}N heteronuclear NOEs were recorded on 0.4mM A2 (41-224) in A2 buffer on a Bruker AVANCE 700MHz NMR spectrometer at 25°, equipped with a 5mm triple-resonance, three-axes gradient probe. For R_1 , the relaxation delays were: 20, 200, 400, 700, 1000, 1500, and 2000ms. For R_2 , the CPMG mixing times were: 16, 32, 48, 64, 80, 96, and 112ms with 2048 $\times$ 512 complex points. For error estimation, R_1 experiments with 200 and 400ms delays and R_2 experiments with 32 and 48ms mixing times were repeated. For the ^1H - ^{15}N heteronuclear NOE reference spectrum a 6s inter-scan delay was used and during the

last 3s of the interscan delay a 120° pulse train with 5ms spacing was inserted. The rotational correlation time of A2 (41-224), τ_C , was calculated according to:

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

with ν_N representing the ^{15}N resonance frequency in Hz (47). The molecular mass of A2 (41-224) was estimated based on the empirical relationship for monomeric proteins:

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

Full-length APOBEC2 Model Structure Generation

A model structure of full-length A2 was generated using HN, N, C α , and C β 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 C12E5/hexanol) of A2 41-224. The latest version of CS-Rosetta, CSHM Rosetta, (Thompson, Sgourakis and Baker, personal communication) was employed that incorporates distance constraints from homologous proteins to enhance convergence properties and improves structure determination of proteins up to 25kDa. The murine A2 (46-224) NMR structure (2RPZ) was utilized as the homology template.

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 multi-angle light scattering (HELEOS, Wyatt Technology), UV (Agilent 1100, Agilent Technology), and refractive index (OptilabREX, Wyatt Technology) detectors. Data were analyzed using the ASTRA software V5.3.1.4 (Wyatt Technology). In all experiments, 100 μl of protein sample in sterile filtered, degassed A2 buffer was injected to the system. Dynamic light scattering measurements were performed using a Wyatt Technology DynaPro Plate Reader and the data were analyzed using Dynamics software V7.01. For each protein concentration, the measurements were performed in triplicate with ten 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 1.5mg/ml of protein in A2 buffer were scanned from 20°C-100°C with a scan rate of 80°C/hr. The data was analyzed using Origin 7 software.

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 carried out on 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 were collected, namely 2D ^1H - ^{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 > 1mM without immediate aggregation and yielded well-resolved spectra. A superposition of ^1H - ^{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 assignable

non-proline resonances, 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 8.34ppm (BMRB database). Such strong deshielding is indicative of hydrogen bonding and seems to be present for the equivalent amide group in all APOBEC spectra that are available to date. In murine A2, the E100 amide is located even slightly further downfield at 11.932ppm (BMRB), and the amide of the corresponding catalytic glutamic acid, E259, in the C-terminal catalytic domain (CTD) of A3G, though not as far downfield, is located at 10.955ppm (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 forty amino acids of A2 can be qualitatively evaluated from the ^1H - ^{15}N HSQC spectrum of full-length A2 (1-224) (Figure 2A, large inset on left). Almost all resonances that belong to residues in this region of the protein are very narrow and cluster around 8ppm (^1H), suggesting that this stretch of polypeptide is highly mobile. Secondary structure prediction by TALOS+ (51) also suggests that this region is very flexible and lacks any well structured elements. Although the absence or presence of first forty amino acids of A2 did not significantly alter the ^1H - ^{15}N HSQC spectra of the catalytic domain of A2 (see superposition of the spectra in Figure 2A), indicating that no gross structural changes are induced in the deaminase domain, closer examination of the combined ^1H - ^{15}N chemical shift differences between the two ^1H - ^{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 $\alpha 1$, strand $\beta 2$, and loops neighboring these elements. This is not too surprising since helix $\alpha 1$ and the short preceding loop are the structural elements in the A2 crystal structure closest to the 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 85-97, which comprise strand $\beta 2$ 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 forty 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-G67 is in two distinct conformations in the human A2 crystal structure, either forming 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 linewidths, considering that a dimer or tetramer would possess molecular masses of 44 or 88kDa and 52 or 104kDa, respectively, for A2 (41-224) and A2 (1-224). Proteins of this size generally exhibit very broad resonances and necessitate TROSY-type experiments. Given the unexpected high spectral quality, ^1H - ^{15}N heteronuclear NOE, R_1 and R_2 relaxation experiments for characterizing the backbone dynamics were carried out. The observed average R_2/R_1 ratios for A2 (41-224) yielded a rotational correlation time of 17ns, compatible with a 28kDa protein and not a 44 or 88kDa protein (Figure S1).

RDCs were measured for the A2 deaminase domain, A2 (41-224), using two different alignment media, Pf1 (43) and C12E5/hexanol (42) (Figure S2). To evaluate whether the structure of A2 (41-224) in solution is identical or very close to monomeric or dimeric units of the tetramer in the crystal, experimental RDCs measured in both media were compared to those calculated based on 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 (Pf1: $R = 0.933$, $Q_{\text{Saupe}} = 0.345$, C12E5/hexanol: $R = 0.937$, $Q_{\text{Saupe}} = 0.248$). In order to compare the measured RDCs with those predicted based on 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 , $Q_{\text{Saupe}} = 0.325$ and 0.353). The experimental RDCs measured in C12E5/hexanol, exhibited even more profound differences from those predicted based on the dimer, with $R = 0.744$ and 0.750 , and $Q_{\text{Saupe}} = 0.663$ 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 the truncated A2 (41-224) in solution, multi-angle light scattering (SEC-MALS) experiments were performed. The data in Figure 3 demonstrate unequivocally that at protein concentrations of 0.5-2mg/ml, only a single state is observed for both protein constructs, with molecular masses corresponding to the monomer state for A2 (41-224) of $22.4\text{kDa} \pm 0.055\text{kDa}$ ($MW_{\text{Theoretical}} = 21.7\text{kDa}$) and for A2 (1-224) of $26.5\text{kDa} \pm 0.026\text{kDa}$ ($MW_{\text{Theoretical}} = 26.1\text{kDa}$). We also tested the protein by dynamic light scattering (DLS) and both A2 proteins at concentrations of 1mg/ml yielded average particle sizes of $2.25\text{nm} \pm 0.04\text{nm}$, corresponding to a molecular mass of $24.5\text{kDa} \pm 1.08\text{kDa}$ for A2 (41-224) and $2.71\text{nm} \pm 0.03\text{nm}$, corresponding to a molecular mass of $34.7\text{kDa} \pm 0.9\text{kDa}$ for A2 (1-224). In both cases, the solutions were monodisperse with peak polydispersity $< 10\%$.

Full-length APOBEC2 Model Structure

A2 is a highly conserved protein among vertebrates (52) with the murine and human forms exhibiting 90% sequence identity. The murine A2 (46-224) NMR structure had been determined by the RIKEN Structural Genomics/Proteomics Initiative (PDB ID; 2RPZ) and since 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 25kDa size by combining traditional NMR data, such as chemical shifts, RDCs, and NOE constraints, with distance constraints from already 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 C12E5/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 extended 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). 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 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 $C\alpha$ chemical shifts generally serve as a good predictor for helicity of a polypeptide region (53). Comparing the $C\alpha$ 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/ψ angles. Based on the analysis of $C\alpha$ chemical shifts exhibited by the N-terminus of A2, we note that no stable helical segments are suggested by this data (Figure S2). Thus, the varying conformations for the N-terminal tail predicted by CS-HM Rosetta are also indicating 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/ψ angles. Although ^1H - ^{15}N HSQC spectra of A2 (1-224) strongly suggests that the N-terminus is predominately 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 ^1H - ^{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 $\beta 2$ face of the protein. Transient H-bonding as well as steric occlusion may play a role in this difference. Therefore, taken together, the above data suggest that the A2 N-terminus exhibits a highly fluctuating structure that may contain nascent helices as modeled by the CS-HM Rosetta.

As expected, the deaminase domain of the full-length A2 model 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 5A-C). In particular, the long loop connecting helix $\alpha 1$ with strand $\beta 1$ exhibits different conformations (Figure 5D). In the murine structure the loop points away from the putative active-site, while in the crystal structure this loop obscures the active-site. In the full-length A2 model, 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 $\alpha 1$. Like in the murine NMR solution structure, our model shows this helix at a 76° angle with respect to helix $\alpha 1$ of the crystal structure, nearly perpendicular to the rest of the fold. Since 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 NMR data collection, there was a distinct difference in the long-term behavior of full-length A2 as compared to the deletion construct. A2 (41-224) readily precipitated at 37°C, while no precipitate was seen in the A2 (1-224) samples, even after a week. In principle, precipitation can occur via non-specific aggregation of the folded or unfolded protein. We, therefore, assessed the thermodynamic stability of both proteins by DSC. Thermal unfolding of both proteins was not reversible and full-length A2 exhibited a T_m of $63.77^\circ\text{C} \pm 0.01^\circ\text{C}$. The deletion construct, on the other hand, exhibited a complex thermal melting pattern with an apparent early unfolding event (T_m of $\sim 52^\circ\text{C}$) 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 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 $\beta 2$ strands, forming a dimer with a contiguous β sheet, and two dimers engage in head-to-head interactions via helix $\alpha 6$, 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 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 RNA 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 acid 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 which 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 present 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 truncation used for crystallization, A2 (41-224), and full-length A2 are monomeric in solution, and even at a concentration of 1mM, soluble A2 was not seen to oligomerize. Full-length A2 did not exhibit any tendency to aggregate/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 a 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 a month and thermal unfolding results in quantitative precipitation.

Extensive intermolecular contacts between the $\beta 2$ 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 $\beta 2$ 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 $\beta 2$ 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 $\beta 2$. Ultimately, only solving the high resolution structure of A3G or another double domain APOBEC, preferably with a substrate mimic, will be able to reveal how such molecules orient the two domains to carry out 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 $\alpha 1$ to strand $\beta 1$. In the A2 crystal structure, this loop conceals the catalytic E100; however, in the murine A2 NMR structure this loop is located away from the active-site. In our full-length human A2 structure, this loop is flexible and can either cover the active site or not. 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 (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 be a way to 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 to engage 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 current 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 based on 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 ssDNA, 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 onto this question.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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

This work was supported by the National Institutes of Health (NIGMS) Grant P50GM082251-to A.M.G. and F32 GM087138 to T.C.K

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, and 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	Public Data Bank
RDC	Residual Dipolar Coupling
SEC-MALS	tandem Size Exclusion Multi-Angle Light Scattering
TEV	catalytic domain of NIa protein from Tobacco Etch Virus
TROSY	Transverse Relaxation-Optimized Spectroscopy

References

1. Liao W, Hong SH, Chan BH, Rudolph FB, Clark SC, Chan L. APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochem Biophys Res Commun.* 1999; 260:398–404. [PubMed: 10403781]
2. Neuberger MS, Harris RS, Di Noia J, Petersen-Mahrt SK. Immunity through DNA deamination. *Trends Biochem Sci.* 2003; 28:305–312. [PubMed: 12826402]
3. Teng B, Burant CF, Davidson NO. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science.* 1993; 260:1816–1819. [PubMed: 8511591]

4. Navaratnam N, Morrison JR, Bhattacharya S, Patel D, Funahashi T, Giannoni F, Teng BB, Davidson NO, Scott J. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J Biol Chem.* 1993; 268:20709–20712. [PubMed: 8407891]
5. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature.* 2002; 418:646–650. [PubMed: 12167863]
6. Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr Biol.* 2004; 14:1392–1396. [PubMed: 15296758]
7. Zheng YH, Irwin D, Kurosu T, Tokunaga K, Sata T, Peterlin BM. Human APOBEC3F is another host factor that blocks human immunodeficiency virus type 1 replication. *J Virol.* 2004; 78:6073–6076. [PubMed: 15141007]
8. Dang Y, Wang X, Esselman WJ, Zheng YH. Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. *J Virol.* 2006; 80:10522–10533. [PubMed: 16920826]
9. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, Flory E, Schumann GG, Munk C. APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem.* 2006; 281:22161–22172. [PubMed: 16735504]
10. Baumert TF, Rosler C, Malim MH, von Weizsacker F. Hepatitis B virus DNA is subject to extensive editing by the human deaminase APOBEC3C. *Hepatology.* 2007; 46:682–689. [PubMed: 17625792]
11. Wiegand HL, Cullen BR. Inhibition of alpharetrovirus replication by a range of human APOBEC3 proteins. *J Virol.* 2007; 81:13694–13699. [PubMed: 17913830]
12. OhAinle M, Kerns JA, Li MM, Malik HS, Emerman M. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe.* 2008; 4:249–259. [PubMed: 18779051]
13. Rogozin IB, Basu MK, Jordan IK, Pavlov YI, Koonin EV. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle.* 2005; 4:1281–1285. [PubMed: 16082223]
14. Mikl MC, Watt IN, Lu M, Reik W, Davies SL, Neuberger MS, Rada C. Mice deficient in APOBEC2 and APOBEC3. *Mol Cell Biol.* 2005; 25:7270–7277. [PubMed: 16055735]
15. Piec I, Lustrat A, Alliot J, Chambon C, Taylor RG, Bechet D. Differential proteome analysis of aging in rat skeletal muscle. *FASEB J.* 2005; 19:1143–1145. [PubMed: 15831715]
16. Sato Y, Probst HC, Tatsumi R, Ikeuchi Y, Neuberger MS, Rada C. Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *J Biol Chem.* 2010; 285:7111–7118. [PubMed: 20022958]
17. Pennings JL, van Dartel DA, Pronk TE, Hendriksen PJ, Piersma AH. Identification by gene coregulation mapping of novel genes involved in embryonic stem cell differentiation. *Stem Cells Dev.* 2011; 20:115–126. [PubMed: 20569193]
18. Vonica A, Rosa A, Arduini BL, Brivanlou AH. APOBEC2, a selective inhibitor of TGFbeta signaling, regulates left-right axis specification during early embryogenesis. *Dev Biol.* 2011; 350:13–23. [PubMed: 20880495]
19. Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell.* 2008; 135:1201–1212. [PubMed: 19109892]
20. Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, Abramowitz LK, Bartolomei MS, Rambow F, Bassi MR, Bruno T, Fanciulli M, Renner C, Klein-Szanto AJ, Matsumoto Y, Kobi D, Davidson I, Alberti C, Larue L, Bellacosa A. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell.* 2011; 146:67–79. [PubMed: 21722948]
21. Guo JU, Su Y, Zhong C, Ming GL, Song H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell.* 2011; 145:423–434. [PubMed: 21496894]
22. Frank D, Keshet I, Shani M, Levine A, Razin A, Cedar H. Demethylation of CpG islands in embryonic cells. *Nature.* 1991; 351:239–241. [PubMed: 2041571]

23. Jost JP, Oakeley EJ, Zhu B, Benjamin D, Thiry S, Siegmund M, Jost YC. 5-Methylcytosine DNA glycosylase participates in the genome-wide loss of DNA methylation occurring during mouse myoblast differentiation. *Nucleic Acids Research*. 2001; 29:4452–4461. [PubMed: 11691933]
24. Metivier R, Gallais R, Tiffocche C, Le Peron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F, Salbert G. Cyclical DNA methylation of a transcriptionally active promoter. *Nature*. 2008; 452:45–50. [PubMed: 18322525]
25. Song F, Mahmood S, Ghosh S, Liang P, Smiraglia DJ, Nagase H, Held WA. Tissue specific differentially methylated regions (TDMR): Changes in DNA methylation during development. *Genomics*. 2009; 93:130–139. [PubMed: 18952162]
26. Okuyama S, Marusawa H, Matsumoto T, Ueda Y, Matsumoto Y, Endo Y, Takai A, Chiba T. Excessive activity of apolipoprotein B mRNA editing enzyme catalytic polypeptide 2 (APOBEC2) contributes to liver and lung tumorigenesis. *Int J Cancer*. 2011
27. Harris RS, Petersen-Mahrt SK, Neuberger MS. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell*. 2002; 10:1247–1253. [PubMed: 12453430]
28. Lada AG, Krick CF, Kozmin SG, Mayorov VI, Karpova TS, Rogozin IB, Pavlov YI. Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. *Biochemistry (Mosc)*. 2011; 76:131–146. [PubMed: 21568845]
29. Prochnow C, Bransteitter R, Klein MG, Goodman MF, Chen XS. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature*. 2007; 445:447–451. [PubMed: 17187054]
30. Kumasaka T, Yamamoto M, Furuichi M, Nakasako M, Teh AH, Kimura M, Yamaguchi I, Ueki T. Crystal structures of blasticidin S deaminase (BSD): implications for dynamic properties of catalytic zinc. *J Biol Chem*. 2007; 282:37103–37111. [PubMed: 17959604]
31. Teh AH, Kimura M, Yamamoto M, Tanaka N, Yamaguchi I, Kumasaka T. The 1.48 Å resolution crystal structure of the homotetrameric cytidine deaminase from mouse. *Biochemistry*. 2006; 45:7825–7833. [PubMed: 16784234]
32. Randau L, Stanley BJ, Kohlway A, Mehta S, Xiong Y, Soll D. A cytidine deaminase edits C to U in transfer RNAs in Archaea. *Science*. 2009; 324:657–659. [PubMed: 19407206]
33. Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, Navaratnam N. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics*. 2002; 79:285–296. [PubMed: 11863358]
34. Lecossier D, Bouchonnet F, Clavel F, Hance AJ. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science*. 2003; 300:1112. [PubMed: 12750511]
35. Zhang KL, Mangeat B, Ortiz M, Zoete V, Trono D, Telenti A, Michielin O. Model structure of human APOBEC3G. *PLoS One*. 2007; 2:e378. [PubMed: 17440614]
36. Chelico L, Prochnow C, Erie DA, Chen XS, Goodman MF. Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. *J Biol Chem*. 2010; 285:16195–16205. [PubMed: 20212048]
37. Chelico L, Pham P, Calabrese P, Goodman MF. APOBEC3G DNA deaminase acts processively 3' → 5' on single-stranded DNA. *Nat Struct Mol Biol*. 2006; 13:392–399. [PubMed: 16622407]
38. Kornhaber GJ, Snyder D, Moseley HN, Montelione GT. Identification of zinc-ligated cysteine residues based on ¹³Cα and ¹³Cβ chemical shift data. *J Biomol NMR*. 2006; 34:259–269. [PubMed: 16645816]
39. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*. 1995; 6:277–293. [PubMed: 8520220]
40. Keller RLJ. The Computer Aided Resonance Assignment Tutorial. CANTNA Verlag. 2004
41. Mulder FA, Schipper D, Bott R, Boelens R. Altered flexibility in the substrate-binding site of related native and engineered high-alkaline *Bacillus subtilis*ins. *J Mol Biol*. 1999; 292:111–123. [PubMed: 10493861]
42. Ruckert M, Otting G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc*. 2000; 122:7793–7797.
43. Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol*. 1998; 5:1065–1074. [PubMed: 9846877]

44. Ottiger M, Delaglio F, Bax A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson.* 1998; 131:373–378. [PubMed: 9571116]
45. Goddard, TD.; Kneller, DG. SPARKY. 3.3.110 ed. University of California San Francisco; 2004.
46. Zweckstetter M. NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc.* 2008; 3:679–690. [PubMed: 18388951]
47. Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry.* 1989; 28:8972–8979. [PubMed: 2690953]
48. Chen KM, Harjes E, Gross PJ, Fahmy A, Lu Y, Shindo K, Harris RS, Matsuo H. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature.* 2008; 452:116–119. [PubMed: 18288108]
49. Harjes E, Gross PJ, Chen KM, Lu Y, Shindo K, Nowarski R, Gross JD, Kotler M, Harris RS, Matsuo H. An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *J Mol Biol.* 2009; 389:819–832. [PubMed: 19389408]
50. Furukawa A, Nagata T, Matsugami A, Habu Y, Sugiyama R, Hayashi F, Kobayashi N, Yokoyama S, Takaku H, Katahira M. Structure and real-time monitoring of the enzymatic reaction of APOBEC3G which is involved in anti-HIV activity. *Nucleic Acids Symp Ser (Oxf).* 2009:87–88.
51. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR.* 2009; 44:213–223. [PubMed: 19548092]
52. Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol.* 2005; 22:367–377. [PubMed: 15496550]
53. Wishart DS, Sykes BD. The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR.* 1994; 4:171–180. [PubMed: 8019132]
54. McDougall WM, Okany C, Smith HC. Deaminase activity on ssDNA occurred in vitro when APOBEC3G forms homotetramers and higher-order complexes. *J Biol Chem.* 2011
55. Hache G, Liddament MT, Harris RS. The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *J Biol Chem.* 2005; 280:10920–10924. [PubMed: 15647250]
56. Navarro F, Bollman B, Chen H, Konig R, Yu Q, Chiles K, Landau NR. Complementary function of the two catalytic domains of APOBEC3G. *Virology.* 2005; 333:374–386. [PubMed: 15721369]
57. Iwatani Y, Takeuchi H, Strebel K, Levin JG. Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. *J Virol.* 2006; 80:5992–6002. [PubMed: 16731938]
58. Chen KM, Martemyanova N, Lu Y, Shindo K, Matsuo H, Harris RS. Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G. *FEBS Lett.* 2007; 581:4761–4766. [PubMed: 17869248]
59. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, Stevens RC, Goodman MF, Chen XS. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature.* 2008; 456:121–124. [PubMed: 18849968]
60. Furukawa A, Nagata T, Matsugami A, Habu Y, Sugiyama R, Hayashi F, Kobayashi N, Yokoyama S, Takaku H, Katahira M. Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO J.* 2009; 28:440–451. [PubMed: 19153609]
61. Shandilya SM, Nalam MN, Nalivaika EA, Gross PJ, Valesano JC, Shindo K, Li M, Munson M, Royer WE, Harjes E, Kono T, Matsuo H, Harris RS, Somasundaran M, Schiffer CA. Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure.* 2010; 18:28–38. [PubMed: 20152150]
62. Ma DK, Guo JU, Ming GL, Song H. DNA excision repair proteins and Gadd45 as molecular players for active DNA demethylation. *Cell Cycle.* 2009; 8:1526–1531. [PubMed: 19377292]
63. Hajkova P, Jeffries SJ, Lee C, Miller N, Jackson SP, Surani MA. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science.* 2010; 329:78–82. [PubMed: 20595612]

64. Zhang W, Liu Z, Crombet L, Amaya MF, Liu Y, Zhang X, Kuang W, Ma P, Niu L, Qi C. Crystal structure of the mismatch-specific thymine glycosylase domain of human methyl-CpG-binding protein MBD4. *Biochem Biophys Res Commun*. 2011

```

APOBEC2 : MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIPLPFIVTGERIPANFEEKFQRNVEYSSGKNT
APOBEC3H : -----MNPQIRN-----PMKAMYEGTFYFQKKNLWEANDNET
APOBEC3C : -----MTSEKGPSTG-----DPTLRRRIEPWEFDVFDYDF--RELKKEA
APOBEC3A : -----MEASPASG-----PRHIMDEHIEITSNENN--GIGRHKT

APOBEC2 : FICYVVEAQQGGQVQASR--CYLEDEHA-----AAHAEAEFFNTILP--AFDEALRNINVTWYVSS
APOBEC3H : RKALLCYQLTPQNGSTPTR--CYFENKKK-----CHAEICFINEIKSM-GLDETQCNQVTCYLTW
APOBEC3C : WICETVEGIKRRSVVSWKT--CVFRNOVDSETH---CHAEKCFISWFCDD-ILSENTKMQVTWYITW
APOBEC3A : YICYEVEERLDNGTSVKMDQHRSEFLHNQAKNLLCGFYGRHAELRFLDLVPSL-QLDEAQITRVTWFTSW

APOBEC2 : SPCA--ACADRRIKTIISKTKNIRLLILVGRLEFMWEEPEIQAAIRKKKEA-CKIRIMKPQDFEYVWQNF
APOBEC3H : SPCS--SCAWEDVDRIKAHDHINIGTFASRLYHWCCKPQKGLRLLCGSOVPPVEVMGFPKFDGWNF
APOBEC3C : SPCP--DCAGEVAERLARHSNVLTIETARLYYFQYPCYQEGRLRSLSQECVAEIMDYEDFKYCWENF
APOBEC3A : SPCW--ECSQAIREILSRHPGVTLVTVYVARLEFWMDOQNROGLRDVNSVTVTQIMRASEYYHCWRNF

APOBEC2 : VEQEE-----GESKAFQPEDIQENFLYYEEIDADILK-----
APOBEC3H : VDHEKP-----LSFNPKMLEEDKNSRAIKRRLERIKIPGVRAQGGRY
APOBEC3C : VYND-----NEPKPKWKGKTNFRLLKRLRESLQ-----
APOBEC3A : VNYPPGDEAHWPQYPPLWMMLYALELHCIIISLPPCLKISRRWQHLLTFFRLHLQCHYQTIIPPHILL
APOBEC3A : VDHO-----GCFQFPWDGIDEHSQALSGIRLRAILQNGN-----

```

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 ▼ denotes the conserved catalytic glutamic acid while ■ denotes the conserved
 histidine and two cysteines that coordinate the zinc ion at the active site.

Figure 2. Comparison between Full-length and Truncated A2

(A) The ^1H - ^{15}N HSQC spectra of A2 (1-224) is superimposed with that of A2 (41-224). 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 8-8.5ppm, indicating that they reside in a flexible, rapidly exchanging region (large left inset). (B) The combined ^1H , ^{15}N chemical shift differences between A2 (1-224) and A2 (41-224) are plotted. Residues only assigned in A2 (1-224) are labeled with a red asterisk and those only assigned in A2 (41-224) with a blue asterisk. (C) The A2 (41-224) crystal structure monomer, residing at the tetramer interface was used as the template to display the location

of residues that were only seen in the A2 (1-224) spectra, red, A2 (41-224) spectra, blue, and those residues that exhibited a combined chemical shift difference greater than the average + 1 standard deviation, green.

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 (dotted line), 1 (dashed line), and 2mg/ml (solid line) predicts a MW of $22.4\text{kDa} \pm 0.055\text{kDa}$ and $26.5\text{kDa} \pm 0.026\text{kDa}$ respectively. (C) The average particle size, predicted molecular mass of the two constructs, and polydispersity of the solution from dynamic light scattering is listed.

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) is displayed. (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 shown in red, along with the deaminase domain in black. The A2 (41-224) dimer of the crystal structure (2NYT) is provided for comparison. The monomer residing at the tetramer interface in the crystal structure is shown in grey and that at the dimer interface in olive.

Figure 5. Comparison of the Full-length A2 Model with other A2 Structures

The CS-HM Rosetta model of A2 (1-224) in red is aligned with: (A) the RIKEN murine A2 (46-224) NMR structure (2RPZ) in cyan, (B) the A2 monomer comprising the tetramer interface of the human A2 (41-224) crystal structure (2NYT) in grey, and (C) the A2 (41-224) monomer not residing at the tetramer interface in olive. (D) For each pair the similarities and differences in helix $\alpha 1$ and loop $\alpha 1\beta 1$ are highlighted. For clarity, the $\alpha 1\beta 1$ loops are colored differently (black = A2 (1-224); purple = tetramer interface A2 (41-224) monomer; orange = dimer interface A2 (41-224) monomer). In all structures, the catalytic E100 and the Zn coordinating residues H98, C128, and C131 are shown in stick representation.

Figure 6. DSC Analysis of Full-length and Truncated A2

1.5mg/ml of A2 (1-224) in A2 buffer was heated from 20°C to 100°C. A2 (1-224) displays a single melting transition with $T_m = 63.77^\circ\text{C} \pm 0.01^\circ\text{C}$.