HHS Public Access

Author manuscript

Nat Commun. Author manuscript; available in PMC 2013 July 01.

Published in final edited form as:

Nat Commun. 2013 ; 4: 1890. doi:10.1038/ncomms2883.

NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity

In-Ja L. Byeon[1,2], **Jinwoo Ahn**[1,2], **Mithun Mitra**[3], **Chang-Hyeock Byeon**[1,2], **Kamil Hercík**[3,4], **Jozef Hritz**[1,5], **Lisa M. Charlton**[1,2], **Judith G. Levin**[3,*], and **Angela M. Gronenborn**[1,2,*]

[1]Department of Structural Biology, University of Pittsburgh School of Medicine, 3501 Fifth Ave., Pittsburgh, PA 15261, USA

[2]Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA

[3]Section on Viral Gene Regulation, Program on Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892-2780, USA

Abstract

Human APOBEC3A (A3A) is a single-stranded DNA (ssDNA) cytidine deaminase that restricts viral pathogens and endogenous retrotransposons and plays a role in the innate immune response. Furthermore, its potential to act as a genomic DNA mutator has implications for a role in carcinogenesis. A deeper understanding of A3A's deaminase and nucleic acid binding properties, which is central to its biological activities, has been limited by the lack of structural information. Here, we report the NMR solution structure of A3A and show that the critical interface for interaction with ssDNA substrates includes residues extending beyond the catalytic center. Importantly, by monitoring deaminase activity in real time, we find that A3A displays similar catalytic activity on A3A-specific TTCA- or A3G-specific CCCA-containing substrates, involving key determinants immediately 5′ of the reactive C. Our results afford novel mechanistic insights into A3A-mediated deamination and provide the structural basis for further molecular studies.

Correspondence and requests for materials should be addressed to: A.M.G. (amg100@pitt.edu).

[4]Current Address: Fraunhofer IME – Division Molecular Biology, Biological Operating Systems: Infectious Diseases, Forckenbeckstraβe 6, 52074 Aachen, Germany

[5]Current Address: CEITEC, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

[*]These authors contributed equally to overseeing the project.

AUTHOR CONTRIBUTIONS

The study was conceived by A.M.G. and J.G.L. The NMR experiments were designed by A.M.G. and I.-J.L.B.; I.-J.L.B. and C.-H.B. performed experiments and data analysis; I.-J.L.B. and A.M.G. interpreted the NMR data; J.A., L.M.C., and C.-H.B. prepared purified A3A; J.H. built structural models of A3A, complexed with short oligonucleotides; the biochemical assays were designed by J.G.L., M.M., and K.H.; M.M. and K.H. performed the experiments and M.M., K.H., and J.G.L. analyzed the results; the manuscript was written by I.-J.L.B., M.M., J.G.L., and A.M.G.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

ACCESSION CODES

The A3A atomic coordinates and NMR constraints have been deposited in the RCSB Protein Data Bank under accession code 2m65, and the NMR chemical shift data have been deposited in the Biological Magnetic Resonance Bank under accession code 19108.

INTRODUCTION

The human APOBEC3 (A3) proteins are a family of deoxycytidine deaminases that convert dC residues in single-stranded DNA (ssDNA) to dU and act as DNA mutators. These proteins, which play an important role in the innate immune response, function as host restriction factors and display a broad range of activities against endogenous and exogenous retroelements[1–3]. There are seven members in the A3 family, each having one (A3A, A3C, A3H) or two (A3B, A3D, A3F, A3G) zinc (Zn)-binding domains with **H**X_1**E**$X_{23–24}$**C**$X_{2–4}$**C** motifs, where X is any amino acid[4]. The histidine and cysteine residues coordinate Zn^{2+}, while glutamic acid is thought to function as a proton shuttle during the deaminase reaction[5].

The single-domain A3A protein, the subject of this study, has multiple activities. A3A degrades foreign DNA introduced into human cells[6,7] and blocks replication of exogenous viruses such as human papilloma virus[8], Rous sarcoma virus[9], parvoviruses[10,11], and human T-lymphotropic virus type 1[12]. In addition, it strongly inhibits retrotransposition of LINE-1, *Alu*, and LTR retroelements[10,13–16] that cause insertional mutations. Unlike A3G[17], A3A is capable of deaminating 5-methylcytosine[18,19], an epigenetic marker in genomic DNA, as well as inducing cell cycle arrest[20] and somatic hypermutation of nuclear and mitochondrial DNAs in a dynamic interplay between A3 editing and DNA catabolism[20,21]. How all these activities are regulated, however, is not fully understood, although Tribbles 3, a human protein, was recently reported to protect nuclear DNA from A3A-mediated deamination[22].

A3A is highly expressed in cells of the myeloid lineage, such as monocytes and macrophages, and its expression is upregulated by treatment with interferon-alpha (IFN-α)[10,23–26]. Interestingly, silencing of A3A in monocytes is associated with increased susceptibility to HIV-1 infection, suggesting that the presence of A3A may be protective against HIV-1[23]. Recent studies indicate that endogenous A3A in macrophages restricts HIV-1 replication by reducing synthesis of viral DNA during reverse transcription[27]. This result is consistent with an independent observation that HIV-1 transcripts in IFN-α-treated infected macrophages appear to be edited predominantly by A3A[28].

Although abundant information on the biological activities of the A3 proteins has been reported, only limited structural data are available. For example, the structure of the C-terminal domain (CTD) of A3G, which contains the catalytic center for its deaminase activity[17,29,30], was solved by NMR[31–33] and X-ray crystallography[34,35], but the structure of full-length A3G has been more elusive. While this manuscript was in preparation, the X-ray structure of another single-domain A3 protein, A3C, was reported[36].

Given human A3A's function as an inhibitor of retroviruses and retroelements with significant effects on cellular activities, the availability of an atomic, three-dimensional structure clearly is of significant importance. Here we report the NMR solution structure of human A3A, define the interface that is critical for its interaction with single-stranded oligonucleotide substrates, and characterize its catalytic activity. Detailed analysis of A3A binding to nucleic acids and real-time monitoring of the deamination reaction by NMR allow us to propose a mechanism for substrate selection and specificity. These studies

provide the structural basis for a deeper understanding of A3A's biological activities and broaden our knowledge of the molecular properties of the A3 proteins.

RESULTS

Biochemical characterization of purified A3A

To ensure that the purified, recombinant wild-type A3A (199 aa), (containing a C-terminal His_6-tag (LEHHHHHH)), was enzymatically active, a uracil DNA glycosylase (UDG)-dependent gel-based deaminase assay was performed (Fig. 1a,b). The catalytic activity of this tagged protein was evaluated using a fluorescently labeled 40-nt ssDNA substrate, containing the TTCA deaminase recognition site, in the presence of increasing concentrations of protein, ranging from 20 to 1000 nM. Over 50% of the substrate was converted to the deaminated product with as little as 20 nM A3A and complete conversion was seen with 200 nM A3A. These results demonstrate that the recombinant A3A protein is highly active as a cytidine deaminase.

Binding of A3A to ssDNA and ssRNA was evaluated in electrophoretic mobility shift assay (EMSA) experiments with ^{32}P-labeled 40-nt oligonucleotides and varying concentrations of A3A (Fig. 1c,d). Complexes are seen at A3A concentrations of ≥50 μM with ssDNA (Fig. 1c, lane 4) and ≥20 μM with ssRNA (Fig. 1d, lane 15). To achieve ~50% complexation, 2-fold more A3A (80 μM) was required for ssDNA (Fig. 1c, lane 7), compared to ssRNA (40 μM) (Fig. 1d, lane 17), suggesting that A3A has a somewhat higher binding affinity for ssRNA than to ssDNA. Estimated K_d values of ~80 μM for the A3A-ssDNA complex agree well with values derived from NMR titration data (see below).

NMR sample preparation and behavior in solution

Full-length A3A was monomeric and soluble up to ~0.2 mM at pH 6.5, 200 mM NaCl, and the ^{1}H-^{15}N HSQC spectrum exhibited well dispersed, narrow resonances (Fig. 2). Nearly complete backbone and side chain assignments were obtained using unlabeled, uniformly ^{15}N-, ^{13}C/^{15}N-, and ^{2}H/^{13}C/^{15}N-labeled A3A with specific-protonations (see METHODS) for three different conditions: (1) ~0.2 mM A3A at pH 6.5, 200 mM NaCl (Fig. 2); (2) ~0.2 mM A3A at pH 6.9; and (3) ~0.3 mM A3A at pH 8.1, all in 25 mM sodium phosphate buffer. Since essentially identical NOE patterns and chemical shifts were observed for the different conditions (only the amide resonances exhibited small, pH-dependent chemical shift differences), the structure was assumed to be unaffected by these different conditions. A mutant A3A, L63N/C64S/C171Q, was also used to help resolve ambiguous assignments (see METHODS).

A3A structure

The structure of full-length A3A was calculated based on 3279 NMR-derived experimental constraints. The final model satisfies all experimental constraints, displays excellent covalent geometry, and the 30-conformer ensemble exhibits atomic r.m.s. deviations of 0.60±0.05 and 1.18±0.05 Å with respect to the mean coordinate positions for the backbone (N, C_α, C′) and all heavy atoms, respectively (Table 1). A stereo-view of the ensemble of

conformers as well as a ribbon representation of the lowest energy structure from the ensemble are depicted in Fig. 3a&b, respectively.

Overall, the A3A structure consists of six helices surrounding a central β sheet of 5 strands (Fig. 3a,b), common to all APOBEC proteins whose structures are known, i.e., A3G-CTD, A3C, and A2 (Supplementary Table S1). The structured region of A3A is limited to residues 10-194 (Fig. 3a,b) since very intense, sharp amide resonances, exhibiting random-coil chemical shifts (Fig. 2), are present for the N-terminal nine (1–9) and final five (195–199) residues in the A3A sequence as well as the His_6-tag residues, indicating a substantial degree of flexibility. In addition to the termini, the loop connecting β2′ and α2 (loop 3, residues 57–70) (Fig. 3a) is also highly plastic and undergoes motions on an intermediate (μ-msec) timescale, since many amide resonances exhibit severe line broadening and are of low intensity (Q58, N61, L62) or entirely missing (A59, Y67, and G68) in the ^{1}H-^{15}N HSQC spectrum (Fig. 2). This is suggestive of multiple conformations undergoing chemical exchange.

Comparison with other APOBEC protein structures

We carried out a detailed comparison of the present A3A solution structure with other available structures of APOBEC proteins (Supplementary Table S1). Among these structures, the present A3A NMR structure is most similar to the X-ray structure of the A3G-CTD quintuple mutant (residues 191-384, PDB: 3IR2)[35] (Fig. 3c,d), exhibiting an average pair-wise backbone atomic r.m.s. difference of 1.86 Å. The next closest structures are the X-ray structures of wild-type A3C, another single-domain human A3 protein (1-190, PDB: 3VOW)[36], and wild-type A3G-CTD (residues 197-380, PDB: 3E1U/3IQS)[34], both exhibiting average backbone r.m.s. differences of ~2.3–2.4 Å. The NMR solution structures of monomeric A2 proteins (murine A2, PDB: 2RPZ, and a CS-HM Rosetta model for human A2[37]) are also similar to A3A (atomic r.m.s. differences of ~3.4 Å). In the previously solved A3G-CTD structures, some local differences were reported[31–34]. Our current A3A structure clearly possesses an interrupted β2 strand (β2-bulge-β2′) (Fig. 3b and Supplementary Fig. S1), as well as an N-terminal α-helix (α1, residues 15–21).

A more detailed comparison was carried out between the current A3A and the most similar A3G-CTD structures (A3G191-384-2K3A; Fig. 3c,d and Supplementary Table S1). We focused on the regions around the active site where the A3A and A3G sequences differ. A two-amino acid (W104 and G105) insertion, located between two zinc-coordinating cysteines (C101 and C106) in the active site in A3A, is the most prominent sequence change. Despite this insertion, the positions of the catalytic site residues H70 (H257 in A3G), E72 (E259), C101 (C288), and C106 (C291) are very similar in both structures (Fig. 3c). The insertion, however, distorts the N-terminal end of helix α3, with W104 bulging out, causing positioning of the backbone and/or sidechain atoms of S103 and G105 in A3A equivalent to those of F289 and S290 in A3G. This structural adjustment places the hydrophobic F102 and W104 side chains at the protein surface, next to the active site. In contrast, A3G possesses only one hydrophobic residue (F289) in this region. Loop 7 is located near the active site (Fig. 3), and its C-terminal half in A3G is known to play an important role in substrate selection (TC or CC)[38]. In A3A, loop 7 (127–135) adds

additional hydrophobic residues (Y132, P134, and L135), while in A3G two of the equivalent residues (D317 and R320) are polar, and the proline is replaced by glycine, resulting in conformational differences between the A3A and A3G loop 7 (Fig. 3c). Note that among the human A3 proteins, the polar character of this region is unique for A3G (Fig. 3e).

A3A interacts with dCTP and dUTP

Although A3A is a well-known cytidine deaminase, the determinants of substrate specificity (i.e., preferential recognition of TC[7,10,39]) have not been characterized in detail. As an initial approach, we investigated binding of A3A to dCTP, several ssDNA oligonucleotide substrates and/or their dU containing products by NMR.

dUTP (Fig. 4a) and dCTP (Fig. 4e) binding was monitored by ^{1}H-^{15}N HSQC spectroscopy and affected regions were mapped onto the A3A structure (insets). Perturbations of amide resonances of the active site amino acids H70, E72, C101, and C106 as well as other surrounding ones (K30, L55, H56, N57, Q58, K60, N61, L63, C64, H70, A71, W98, S99, and Y132) were very similar, indicating that both the substrate and the product bind to the active site. Several of the amino acids whose resonances were affected reside in loop 3, with N61, L63, and C64 located distal to the active site, indicating that a remote conformational change was induced by dCTP/dUTP binding or that a particular preferred conformation was selected from the flexible ensemble in the substrate-free A3A state. Titration curves of selected perturbed resonances upon dCTP (Fig. 4e) and dUTP (Fig. 4a) binding were used to extract K_d values of 536±72 and 578±115 μM, respectively (Table 2). Note that since up to 30% of the dCTP was converted to dUTP during the titration, the value for dCTP should be regarded as an approximation.

A3A binds ssDNA substrates via an extended surface

The interaction of ATTTCATTT (~0.1–3 mM) with A3A (~0.15 mM) under our NMR conditions resulted in conversion to ATTTUATTT within <1 min, rendering it impossible to measure a true K_d value for ATTTCATTT. However, given the similarity between dUTP and dCTP binding to A3A (Fig. 4a vs. 4e), it seemed reasonable to assume that ATTTUATTT and ATTTCATTT would also have similar binding properties. The K_d value for the ATTTUATTT product was 58±8 μM (Fig. 4b, Table 2), and its interaction surface on the A3A structure (Fig. 4b, **inset**) covers a larger surface, extending beyond the catalytic site.

This extended interface involves an area to the left side of the active site, opposite of loop 3 (where residues are colored in hot pink (Δδ>0.050 p.p.m.) or cyan (Δδ between 0.028 and 0.050 p.p.m.) in the ribbon diagram). Interestingly, the two-residue insertion (W104/G105), unique to A3A, is part of the interface: the W104 backbone amide (Δδ=0.122 p.p.m.) and the W104 e1 side chain resonance (Δδ=0.079 p.p.m.) as well as the G105 (Δδ=0.066 p.p.m.) and S103 (Δδ=0.072 p.p.m.) amide resonances exhibit the largest perturbations. In addition, loop 7 and helix α4 are also part of the extended interface, since effects on the amide resonances of residues D133 (Δδ=0.112 p.p.m.), L135 (Δδ=0.102 p.p.m.), and E138 (Δδ=0.079 p.p.m.) were observed. Additional binding studies of A3A using a 15-nt oligonucleotide,

ATTATTTUATTTATT yielded essentially the same binding site (Fig. 4c, **inset**) and affinity (K_d=57±11 μM) (Fig. 4c, Table 2) as found with the 9-mer.

Although smaller perturbations were observed for some additional resonances, these changes were non-saturable and most likely reflect non-specific binding (e.g., the I17 resonance marked by a dashed line in Fig. 4c). The non-specific binding site of the 15-mer oligonucleotide mapped onto the A3A structure involves residues 12–19 (α1), 36–37, 40–42 (β1), 44–49 (β2), 53–56 (β2′) and 179–186 (α6) (Supplementary Fig. S2). To validate the binding data, the interaction of an A3A catalytic site mutant, E72Q, with a 15-nt substrate (ATTATTTCATTTATT) was evaluated. The mutant exhibited the same affinity (K_d=55±11 μM, data not shown) as wild-type A3A with the ATTATTTUATTTATT product, confirming that binding of substrate and product occurs with essentially identical affinities.

Further titration experiments using smaller oligonucleotides (ATTT(C/U)A, Fig. 4d; TCATTT, Fig. 4f) delineated a binding site similar to the one observed with the nonanucleotide. This implies that three nucleotides, (TCA), constitute the essential moiety for binding to A3A. The affinities of the hexanucleotides, however, were ~3-fold weaker (K_d~190 μM) than the one measured for ATTT(C/U)ATTT (K_d=58±8 μM) (Table 2), suggesting a stabilizing effect of the flanking nucleotides in the complex.

Interaction of A3A with a 9-nt substrate, AAACCCAAA, containing the A3G deaminase-specific recognition site[2], maps to the same surface (Fig. 4g, **inset**) as ATTT(C/U)ATTT (Fig. 4b, **inset**). This suggests that both dT and dC at the -1 and -2 positions, i.e., pyrimidine bases, can interact with the extended interface next to the catalytic site of A3A. However, slightly weaker binding (K_d=91±15 μM, Fig. 4g, Table 2) was observed for AAACCCAAA compared to ATTTCATTT. Replacement of the central CCC in AAACCCAAA by CCA (AAACCAAAA, Fig. 4h) or CAA (AAACAAAAA, Table 2) resulted in identical (K_d=94±11 μM) or slightly weaker (K_d=161±19 μM) affinities, respectively.

Structural model for A3A-ssDNA complexes

Model structures of the A3A-oligonucleotide complexes were created by flexible docking[40] and selecting binding poses compatible with chemical shift perturbations (see METHODS for details). Initial docking results for ATTTCATTT and AAACCCAAA indicated that only the central region of the oligonucleotides interacts specifically with A3A, leaving the two ends free. Therefore, we only generated final structural models of A3A with the pentanucleotides TTCAT (Fig. 5a) and CCCAA (Fig. 5b). In the A3A/TTCAT complex model, the central reactive C occupies the deep pocket delineated by the active site and surrounding residues T31, N57, Q58, H70, E72, W98, C101, C106, D131 and the zinc ion. The thymidine (T_{-1}) immediately preceding the C interacts with a surface formed by residues D133, P134, L135 (loop 7) and F102 (loop 5), while T_{-2} contacts residues F102, S103, W104, and G105 (loop 5), L135 (loop 7), and E138 (α4). The adenosine (A_{+1}) at the 3′ side of the C interacts with residues D131, Y132, and D133 (loop 7). The 5-methyl groups of T_{-1} and T_{-2} are in close contact with several hydrophobic residues, e.g., F102, W104, and L135, suggesting that this hydrophobic interaction may contribute to tighter binding of TTCA- compared to CCCA-containing ssDNA. The model for the A3A/CCCAA

complex (Fig. 5b) is very similar to that of A3A/TTCAT (Fig. 5a), including the localization of the individual nucleotides.

Deamination by real-time NMR

The A3A-catalyzed deamination reaction with several substrates was followed by 2D ^{1}H- ^{13}C HSQC (Fig. 6a) or 1D ^{1}H (Fig. 6b–d) NMR. Using dCTP as the substrate we determined that slow, but measurable deaminase activity is present (Fig. 6a): at 25 °C, 5.2 mM dCTP was completely converted to dUTP by 170 μM A3A in ~50 h. Longer substrates, such as the 9-nt ATTTCATTT, were deaminated much more rapidly (Fig. 6b), and in such cases, 1D ^{1}H NMR spectroscopy was used (Fig. 6b **inset**). The 9-nt substrate (0.98 mM) was completely deaminated by 0.196 μM A3A in ~1.7 h with an initial rate of ~0.8 mM/h. Other substrates, a 15-mer, ATTATTTCATTTATT, and two hexanucleotides, ATTTCA and TCATTT, (Supplementary Fig. S3a–c, respectively) exhibited essentially identical rates to that observed with the nonanucleotide.

A3A and A3G purportedly possess a difference in substrate specificity[2,6,7,10,39,41]. To quantify A3A's substrate specificity, we also assayed the deamination of an A3G-specific substrate, AAAC$_{-2}$C$_{-1}$CAAA, by A3A. Interestingly, all three cytidines were deaminated in a sequential 3′→5′ manner (Fig. 6c), with the third dC (C) converted with an initial rate of ~0.2 mM/h, only four-fold slower than the dC in the A3A-specific ATTTCATTT substrate. When comparing substrates with varying numbers of cytosines, such as AAACCCAAA (Fig. 6c), AAACCAAAA (Fig. 6d), and AAACAAAAA (Supplementary Fig. S3d), deamination of the 3′ dC (C) was about 2- to 4-fold faster in AAACCCAAA than in the other two substrates, indicating a preference for two pyrimidine bases at the 5′ side of the reactive C.

We also noted that the kinetics of the C→U conversions of C$_{-1}$ and C$_{-2}$ are much slower, quite distinct from that of C in the reaction with the AAAC-2C$_{-1}$CAAA substrate. Similar behavior was observed with AAAC$_{-2}$C$_{-1}$UAAA (Supplementary Fig. S3e), a substrate containing a uridine instead of the most reactive C. In contrast, when dC is followed by dA, and not by dU, e.g., the C in AAAC$_{-1}$CAAA (Fig. 6d) and AAACAAAAA (Supplementary Fig. S3d), the reaction proceeds readily, albeit somewhat slower than with AAAC$_{-2}$C$_{-1}$CAAA (Fig. 6c). Given our observation that dC and dU essentially possess the same A3A binding affinities and interfaces, it seems reasonable to assume that once dC is converted to dU, competitive binding and inhibition occurs, slowing the deamination of C$_{-1}$ and C$_{-2}$ in AAAC$_{-2}$C$_{-1}$UAAA. Therefore, optimal substrates for A3A contain a cytidine preceded by two pyrimidine bases and followed by an adenine, i.e., (T/C)(T/C)CA.

Further studies were performed to determine the catalytic constants for A3A deaminase activity on single-cytidine containing substrates, i.e., ATTTCATTT, ATTATTTCATTTATT, ATTTCA, TCATTT, and AAACAAAAA. The kinetics of deamination revealed that all TCA-containing oligonucleotide substrates, regardless of length, exhibited identical turn-over rates (k_{cat} ~60–70/min) (Table 2, Fig. 6e–g) *k*. The hexanucleotides, however, displayed a ~3-fold lower apparent second order rate constant ($_{cat}/K_M$ ~5 × 10^3/M•s) than the one measured for the 9-nt and 15-nt substrates (k_{cat}/K_M ~1.8 × 10^4/M•s), caused mainly by an increase in K_M. For all A3A substrates, K_M and K_d values

were very similar. A large reduction (~40-fold) of k_{cat}/K_M was seen only with a substrate that is lacking pyrimidine bases at the 5′ side of the reactive C, i.e., AAACAAAAA, caused by a large (13-fold) reduction in k_{cat} and a small (~3-fold) increase in K_M (less-favorable binding), highlighting the importance of a pyrimidine base at the 5′ side.

Collectively, the data in Fig. 6 demonstrate that the nucleotide context surrounding the reactive C is a major determinant of enzymatic activity.

DISCUSSION

Here, we present the NMR solution structure of human A3A as well as a detailed analysis of its nucleic acid interaction surface, providing new insights into substrate selection and binding. The A3A NMR structure very closely resembles the A3G-CTD and A3C X-ray structures[31–36] (Supplementary Table S1) and not surprisingly, there are also similarities between the activities of A3A and the A3G-CTD. Using NMR, we show that A3A binds TTCA- or CCCA-containing single-stranded oligonucleotides (≥9 nt) with K_d values ranging from 50–100 μM (Table 2), in excellent agreement with binding affinity data (~80 μM) estimated from EMSA (Fig. 1c) and with fluorescence depolarization results obtained by Love et al[39].

Intriguingly, A3A binds ssDNA much more weakly (~1000-fold) than A3G, for which K_d values ranging from 50 to 240 nM have been reported[17,39,42–44]. However, like A3A, K_d values in the μM range (200–450 μM) were obtained for the single-domain A3G-CTD protein[31,32], implying that the tighter binding of full-length A3G is associated with its double domain structure. Indeed, the A3G-NTD contains numerous positively charged amino acids that contribute to high efficiency binding to ssDNA[17,29], whereas both A3A and the A3G-CTD are slightly acidic and are unable to interact in this manner.

Importantly, titration of A3A with diverse ssDNA substrates by NMR made it possible to distinguish specific binding from non-specific binding (Fig. 4c). Our results showed that A3A can bind both TTCA- and CCCA-containing oligonucleotides, using the same five A3A contacts, namely the active site, loop 3, loop 5 including the exposed di-peptide W104-G105, loop 7, and helix α4. Surprisingly, although the W104NεH side chain resonance experiences the largest chemical shift changes upon substrate binding (Fig. 4), W104 mutations do not seem to influence or abrogate A3A's catalytic activity[16,18].

Note that A3A oligonucleotide-binding regions are highly localized and clustered around the active site, in stark contrast to results in the two studies that mapped ssDNA binding to A3G-CTD[31,32]. Importantly, our A3A-ssDNA complex model (Fig. 5) suggests that the DNA bends to insert the reactive cytidine into the active site, permitting only the immediate neighboring (−1, −2, and +1) nucleotides to interact locally near the catalytic site. Interestingly, there are reports of RNA bending in the crystal structure of a tRNA adenosine deaminase/RNA complex[45] and DNA contraction during A3G scanning of the ssDNA substrate[46].

NMR real-time kinetic data for the deamination reaction of A3A using ATTTCATTT and ATTATTTCATTT as substrates yielded values of k_{cat} ~70 min^{-1} and K_M ~60 μM, which

differ from the values (k_{cat} ~15 min^{-1} and K_M ~230 nM) obtained using a 43-nt TTCT-containing ssDNA[18]. These differences may be related to variations in the ssDNA sequences and experimental conditions. Interestingly, Carpenter et al. reported that A3A is a stronger (~200-fold) deaminase than full-length A3G[18]. Indeed, although similar amounts of deaminase product were observed in the NMR deamination assays described here for A3A (Fig. 6b) and for A3G-CTD in Furukawa et al.[32], the amount of A3A used was 1000-fold less than the amount of A3G-CTD.

The fact that A3A also deaminates the most 3′ dC (C) in AAACCCAAA (thought to be an A3G-specific substrate[2]) with only a ~5-fold reduction in k_{cat}/K_M, compared to TTCA-containing 9-nt and 15-nt oligonucleotides (Table 2) appears puzzling at first. However, these results are in agreement with the deaminase specificity observed for A3A in cell-based assays. For example, in an investigation of foreign DNA restriction in human primary cells, A3A preferentially deaminated TC and CC sequences in the green fluorescent protein gene of the transfected plasmid DNA[6]. In addition, a strong bias for deamination of TC (~50–70%) and CC (~15–25%) dinucleotide sequences was detected in nascent HIV-1 cDNAs isolated from infected macrophages[28] and in an *in vitro* HIV-1 model replication assay, performed using purified A3A protein[39]. Furthermore, a preference for TC and CC substrate recognition sites was also observed for A3A editing of HTLV-1[12]. Thus, A3A's active site possesses the flexibility to accommodate and deaminate dC residues (C) in both TC and CC dinucleotides.

Interestingly, in a study by Shinohara et al.[7], it was shown that A3A mediates genomic DNA editing in human cells, but no editing site preference was detected. In contrast, Suspène et al.[21] found that A3A preferentially deaminates cytidines in TC and CC dinucleotides in genomic DNA, when cells are exposed to uracil DNA glycosylase inhibitor. APOBEC-mediated genomic DNA mutations have been implicated in carcinogenesis[47] and for example, A3B was shown to be a source of DNA mutations in breast cancer[48]. These observations suggest that the strong mutagenic potential of A3A might be detrimental to the stability of the human genome.

Thus, the dual function of A3A as a host restriction factor and as a DNA mutator that can potentially act on genomic DNA, an activity that may be associated with malignancies, suggests that A3A can act as a "double-edged sword". The high resolution NMR structure of A3A presented here is a first step in aiding future structure-function studies for addressing these seemingly diverse A3A functions. Furthermore, the addition of the A3A structure to the still limited list of currently known APOBEC structures contributes to efforts towards elucidating the molecular mechanisms of the innate immune response.

METHODS

Protein expression and purification

Wild-type (Accession number NM_145699), E72Q and L63N/C64S/C171Q mutant synthetic A3A genes with a C-terminal His_6-tag (LEHHHHHH) were inserted into the NdeI-XhoI site of the pET21 plasmid (Novagen) for expression in *E. coli* Rosetta 2 (DE3). Uniform ^{15}N- and ^{13}C-labeling of the proteins was carried out by growth in modified

minimal medium at 18 °C using $^{15}NH_4Cl$ and $^{13}C_6$-glucose as the sole nitrogen and carbon sources, respectively. Uniform ^{2}H-, ^{15}N- and ^{13}C-labeling of the proteins was achieved using $^{2}H_2O$, $^{15}NH_4Cl$ and $^{13}C_6/^{2}H_7$-glucose as deuterium, nitrogen and carbon sources, with two different selective protonation of the side chains of (1) Tyr/Phe/Ile residues and (2) Tyr/Phe/Trp/Ile/Val/Leu residues, by adding 0.10–0.15 mg of $^{13}C/^{15}N$-tyrosine, -phenylalanine and -isoleucine (for sample 1) and $^{13}C/^{15}N$-tyrosine and -phenylalanine, unlabeled tryptophan, 2-keto-butyrate (1,2,3,4-^{13}C, 98%; 3,3′-^{2}H, 98%, CIL, Andover, MA) and 2-keto-3-methyl-butyrate (1,2,3,4-^{13}C, 99%; 3,4,4′,4″-^{2}H, 98%, CIL) (for sample 2), respectively. These chemicals were added to the culture 1 h prior to induction with 0.4 mM isopropyl-1-thio-β-D-galactopyranoside (total induction time=16 h). Proteins were purified over a 5-ml Hi-Trap His column (GE Healthcare) and Hi-Load Superdex 200 (1.6 cm × 60 cm) column, equilibrated in buffer containing 25 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5% glycerol, 2 mM DTT, and 0.02% NaN_3. Fractions containing A3A were further purified over an 8-ml MONO-Q column (GE Healthcare) in 25 mM Tris-HCl buffer (pH 8.5), 5% glycerol, 2 mM DTT, and 0.02% sodium azide, employing a linear gradient of 0–1 M NaCl. The final A3A preparations were >99% pure, as estimated by SDS-PAGE. The molecular mass of the A3A proteins were confirmed by LC-ESI-TOF mass spectrometry (Bruker Daltonics, Billerica, MA).

Multi-angle light scattering

Size-exclusion chromatography/multi-angle light scattering (SEC-MALS) data were obtained at room temperature using an analytical Superdex 200 (S200) column with inline multi-angle light scattering, refractive index (Wyatt Technology, Inc., Santa Barbara, CA) and UV (Agilent Technologies, Santa Clara, CA) detectors. 100-µl of 78.4 µM A3A were applied to the S200 column pre-equilibrated and eluted with 25 mM sodium phosphate buffer (pH 6.5), 200 mM NaCl, 0.02% sodium azide, and 1 mM DTT at a flow-rate of 0.5 ml/minute.

Deaminase assay using fluorescent tagged ssDNA substrates

Deaminase assay conditions were adapted from Iwatani et al.[17]. Forty-µl reactions, containing 180 nM of a 40-nt ssDNA (JL913, 5′-ATT ATT ATT ATT ATT ATT ATT TCA TTT ATT TAT TTA TTT A-3′), labeled at its 5′ end with Alexa Fluor® 488, (Integrated DNA Technologies (IDT) [Coralville, IA]) and varying amounts of A3A in 10 mM Tris-HCl buffer, pH 8.0, 50 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 8.0, and 10 units of E. coli UDG [New England BioLabs]) were incubated at 37 °C for 1 h. The reaction was stopped by incubation with Proteinase K (40 µg, Ambion) at 65 °C for 20 min, followed by sequential addition of 10 µl of 1 N NaOH for 15 min at 37 °C and 10 µl of 1N HCl. Ten-µl aliquots of the final mixture were subjected to electrophoresis in a 10% denaturing polyacrylamide gel. Gels were scanned in fluorescence mode on a Typhoon 9400 Imager and the data were quantified using ImageQuant software (GE Healthcare).

Electrophoretic mobility shift assay (EMSA)

Ten-µl reactions, containing varying amounts of A3A, 20 nM of a 5′ ^{32}P-labeled 40-nt ssDNA (JL895, identical sequence to JL913 but without the Alexa Fluor® 488 label) (Lofstrand) or 40-nt ssRNA (JL931, identical sequence to JL895, except that U was

substituted for T) (IDT), were incubated with 4 U SUPERaseIN (Ambion) at 37 °C for 10 min in 50 mM Tris-HCl buffer, pH 7.0, 100 mM NaCl, 1 mM DTT, 1% Ficoll-400, and 2.5 mM EDTA. Aliquots from each reaction were loaded onto an 8% native polyacrylamide gel in 40 mM Tris-acetate buffer, pH 8.4, 1 mM EDTA, and 5% glycerol. One μl of DNA Loading Gel Solution (Quality Biological, Inc.), containing bromphenol blue and xylene cyanole, was added to a control sample without A3A. Gels were run at 4 °C (5 mA) until the bromphenol blue dye had migrated ~2/3 through the gel. Radioactive products were detected with a Typhoon 9400 Imager and were quantified using ImageQuant software.

NMR Spectroscopy

All NMR spectra for the structure determination of A3A were recorded at 25 °C on Bruker AVANCE900, AVANCE800, AVANCE700, and AVANCE600 spectrometers, equipped with 5-mm triple-resonance, *z*-axis gradient cryoprobes. The NMR samples contained unlabeled, $^{13}C/^{15}N$- or $^{2}H/^{13}C/^{15}N$-labeled A3A with two types of selective protonations (see above). At pH 6.9 and 8.1, A3A showed higher solubility (~0.5 and ~1 mM, respectively) than at pH 6.5 (~0.2 mM). However, at these higher protein concentrations, aggregation occurred. We therefore performed all of our NMR experiments at concentrations of ~0.2–0.3 mM. The sample temperature in the spectrometer was calibrated with 100% methanol. Backbone and side chain resonance assignments were carried out using 2D ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC and NOESY and 3D HNCACB, HN(CO)CACB, HNCA, HN(CO)CA and HCCH-TOCSY experiments[49]. Distance constraints were derived from 3D simultaneous ^{13}C- and ^{15}N-edited NOESY[50] and 2D NOESY experiments. All NOESY spectra were acquired at 800 or 900 MHz, using a mixing time of 100 ms (non-perdeuterated samples) or 150 ms (perdeuterated samples). Spectra were processed with TOPSPIN 2.1 (Bruker) and NMRPipe[51] and analyzed using SPARKY3 (version 3.113) (T.D. Goddard and D.G. Kneller, University of California, San Francisco) and NMRView J (version 8.0.3)[52].

NMR structure calculation

All NOE cross peaks were assigned from the 3D and 2D NOESY spectra using the SPARKY3 assignment tool. Structure calculations were performed for A3A residues 1–199, using the anneal.py protocol in XPLOR-NIH[53]. An iterative approach with extensive, manual cross-checking of all distance constraints against the NOESY data and the generated structures was employed. The final number of the NMR-derived constraints were 3279, with 2827 NOE distances, 152 H-bond distances, identified from NOE patterns for helices and β-sheets, and 300 ϕ and ψ backbone torsion angles from TALOS calculations[54]. In addition, a Zn^{2+} ion was added at a late stage in the calculations to coordinate with H70, C101 and C106[31,32] using constraints based on the X-ray structures of A3G-CTD[34,35]. 512 structures were generated and the 30 lowest energy structures were selected and analyzed using PROCHECK-NMR[55] (Table 1). All structure figures were generated with MOLMOL[56].

A3A nucleotide binding site mapping and titration by NMR

To monitor nucleotide binding to A3A, aliquots of 14–100 mM mono or oligodeoxynucleotide stock solutions were added to 0.10–0.18 mM $^{13}C/^{15}N$-labeled A3A.

100 mM dCTP and dUTP stock solutions were purchased from Promega (Madison, WI) and HPLC-purified DNA oligonucleotides were obtained from IDT or MIDLAND Co. (Midland, TX). A series of 2D ^{1}H-^{15}N HSQC titration spectra were acquired and binding isotherms were obtained by plotting ^{1}HN proton chemical shift change vs. nucleotide concentrations for 5–8 unambiguously traceable amide resonances. Dissociation constants were calculated by nonlinear best fitting of the isotherms using KaleidaGraph (Synergy Software, Reading, PA) and values for 5–8 resonances were averaged.

Real-time studies of A3A-catalyzed deamination

A series of 2D ^{1}H-^{13}C HSQC and/or 1D ^{1}H NMR spectra were acquired as a function of time, after addition of A3A to solutions of mono or oligodeoxynucleotide. Concentrations of A3A and deoxynucleotides for the different samples are provided in the Figure captions. The intensities of well-resolved ^{13}C-5-^{1}H resonances (volumes in 2D ^{1}H-^{13}C HSQC spectra) and/or ^{1}H-5 resonances (integrals in 1D ^{1}H spectra) of cytosine and uracil were used for quantitation. Real-time monitoring of the A3A-catalyzed deamination reaction by NMR permitted the extraction of initial (<5%→dC dU conversion) rates for a series of substrate concentrations (50, 100, 200, 400, and 600 μM). k_{cat} (V_{max}/[A3A]) and K_M values were obtained using the Michaelis-Menten module in Kaleidagraph.

Molecular docking

Structures of TTCAT, CCCAA, ATTTCATTT, and AAACCCAAA ssDNAs were generated by MacroMoleculeBuilder (version 2.8)[57]. The oligonucleotide and A3A NMR structures were converted to a mol2 file format using the Hermes program (version 1.4) and standard Tripos atom and bond types. Docking was performed with the flexible docking program GOLD version 5.1[40], in combination with the Chemscore scoring function[58]. The zinc ion was chosen as the center point in the docking and the radius was set to 20 Å. For each nucleotide, five independent docking runs into the five lowest energy NMR structures of A3A, each producing fifty binding poses, were performed using a population of 100 ligands, with a maximal number of genetic algorithm operations dependent on ligand size: 50000 for TTCAT and CCCAA and 150000 for ATTTCATTT and AAACCCAAA. 1250 docking poses for the pentanucleotides were subjected to a two-step selection process utilizing the NMR data. In the first step, complexes were discarded if the oligonucleotide engaged in contacts with residues that did not exhibit experimental chemical shift changes. In the second step, structures were selected that exhibited good agreement with the NMR binding-site mapping data, taking the size of the chemical shift changes and the scoring value of the binding pose into account. The final structural models for A3A, complexed with TTCAT (Fig. 5a) and CCCAA (Fig. 5b), represent ~80% of all the structures that result from the selection process.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Maria DeLucia for help with A3A purification, Doug Bevan and Phil Greer for computer technical support, and Michael J. Delk for NMR instrumental support. J.H. acknowledges financial support by an International Outgoing Fellowship of the European Community program “Support for training and career development of researchers (Marie Curie)”, under Contract No. PIOF-GA-2009-235902. This work was supported in part by the National Institutes of Health NIGMS Grant P50GM082251 (A.M.G.) and the Intramural Research Program at the National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development (M.M., K.H., and J.G.L.).

References

1. Chiu YL, Greene WC. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol. 2008; 26:317–353. [PubMed: 18304004]
2. Malim MH. APOBEC proteins and intrinsic resistance to HIV-1 infection. Philos Trans R Soc Lond B Biol Sci. 2009; 364:675–687. [PubMed: 19038776]
3. Duggal NK, Emerman M. Evolutionary conflicts between viruses and restriction factors shape immunity. Nat Rev Immunol. 2012; 12:687–95. [PubMed: 22976433]
4. Jarmuz A, et al. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics. 2002; 79:285–296. [PubMed: 11863358]
5. Betts L, Xiang S, Short SA, Wolfenden R, Carter CW Jr. Cytidine deaminase. The 2.3 Å crystal structure of an enzyme: transition-state analog complex. J Mol Biol. 1994; 235:635–656. [PubMed: 8289286]
6. Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. Nat Struct Mol Biol. 2010; 17:222–229. [PubMed: 20062055]
7. Shinohara M, et al. APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells. Sci Rep. 2012; 2:806. [PubMed: 23150777]
8. Vartanian JP, Guetard D, Henry M, Wain-Hobson S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science. 2008; 320:230–3. [PubMed: 18403710]
9. Wiegand HL, Cullen BR. Inhibition of alpharetrovirus replication by a range of human APOBEC3 proteins. J Virol. 2007; 81:13694–13699. [PubMed: 17913830]
10. Chen H, et al. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Curr Biol. 2006; 16:480–485. [PubMed: 16527742]
11. Narvaiza I, et al. Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase. PLoS Pathog. 2009; 5:e1000439. [PubMed: 19461882]
12. Ooms M, Krikoni A, Kress AK, Simon V, Münk C. APOBEC3A, APOBEC3B, and APOBEC3H haplotype 2 restrict human T-lymphotropic virus type 1. J Virol. 2012; 86:6097–6108. [PubMed: 22457529]
13. Bogerd HP, et al. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A. 2006; 103:8780–8785. [PubMed: 16728505]
14. Muckenfuss H, et al. APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem. 2006; 281:22161–22172. [PubMed: 16735504]
15. Kinomoto M, et al. All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic Acids Res. 2007; 35:2955–2964. [PubMed: 17439959]
16. Bulliard Y, et al. Structure-function analyses point to a polynucleotide- accommodating groove essential for APOBEC3A restriction activities. J Virol. 2011; 85:1765–1776. [PubMed: 21123384]
17. 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]
18. Carpenter MA, et al. Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. J Biol Chem. 2012; 287:34801–8. [PubMed: 22896697]

19. Wijesinghe P, Bhagwat AS. Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. Nucleic Acids Res. 2012; 40:9206–17. [PubMed: 22798497]
20. Landry S, Narvaiza I, Linfesty DC, Weitzman MD. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. EMBO Rep. 2011; 12:444–450. [PubMed: 21460793]
21. Suspène R, et al. Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. Proc Natl Acad Sci U S A. 2011; 108:4858–4863. [PubMed: 21368204]
22. Aynaud MM, et al. Human Tribbles 3 Protects Nuclear DNA from Cytidine Deamination by APOBEC3A. J Biol Chem. 2012; 287:39182–92. [PubMed: 22977230]
23. Peng G, et al. Myeloid differentiation and susceptibility to HIV-1 are linked to APOBEC3 expression. Blood. 2007; 110:393–400. [PubMed: 17371941]
24. Koning FA, et al. Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. J Virol. 2009; 83:9474–9485. [PubMed: 19587057]
25. Refsland EW, et al. Quantitative profiling of the full *APOBEC3* mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. Nucleic Acids Res. 2010; 38:4274–4284. [PubMed: 20308164]
26. Thielen BK, et al. Innate immune signaling induces high levels of TC-specific deaminase activity in primary monocyte-derived cells through expression of APOBEC3A isoforms. J Biol Chem. 2010; 285:27753–27766. [PubMed: 20615867]
27. Berger G, et al. APOBEC3A is a specific inhibitor of the early phases of HIV-1 infection in myeloid cells. PLoS Pathog. 2011; 7:e1002221. [PubMed: 21966267]
28. Koning FA, Goujon C, Bauby H, Malim MH. Target cell-mediated editing of HIV-1 cDNA by APOBEC3 proteins in human macrophages. J Virol. 2011; 85:13448–13452. [PubMed: 21957290]
29. Navarro F, et al. Complementary function of the two catalytic domains of APOBEC3G. Virology. 2005; 333:374–386. [PubMed: 15721369]
30. Newman EN, et al. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. Curr Biol. 2005; 15:166–170. [PubMed: 15668174]
31. Chen KM, et al. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. Nature. 2008; 452:116–119. [PubMed: 18288108]
32. Furukawa A, et al. Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. EMBO J. 2009; 28:440–451. [PubMed: 19153609]
33. Harjes E, et al. An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. J Mol Biol. 2009; 389:819–832. [PubMed: 19389408]
34. Holden LG, et al. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. Nature. 2008; 456:121–124. [PubMed: 18849968]
35. Shandilya SMD, et al. Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. Structure. 2010; 18:28–38. [PubMed: 20152150]
36. Kitamura S, et al. The APOBEC3C crystal structure and the interface for HIV- 1 Vif binding. Nat Struct Mol Biol. 2012; 19:1005–10. [PubMed: 23001005]
37. Krzysiak TC, Jung J, Thompson J, Baker D, Gronenborn AM. APOBEC2 is a monomer in solution: implications for APOBEC3G models. Biochemistry. 2012; 51:2008–17. [PubMed: 22339232]
38. Bransteitter R, Prochnow C, Chen XS. The current structural and functional understanding of APOBEC deaminases. Cell Mol Life Sci. 2009; 66:3137–3147. [PubMed: 19547914]
39. Love RP, Xu H, Chelico L. Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A. J Biol Chem. 2012
40. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins. 2003; 52:609–23. [PubMed: 12910460]
41. Aguiar RS, Lovsin N, Tanuri A, Peterlin BM. Vpr.A3A chimera inhibits HIV replication. J Biol Chem. 2008; 283:2518–2525. [PubMed: 18057006]
42. 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]

43. Nowarski R, Britan-Rosich E, Shiloach T, Kotler M. Hypermutation by intersegmental transfer of APOBEC3G cytidine deaminase. Nat Struct Mol Biol. 2008; 15:1059–66. [PubMed: 18820687]
44. Iwatani Y, et al. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. Nucleic Acids Res. 2007; 35:7096–7108. [PubMed: 17942420]
45. Losey HC, Ruthenburg AJ, Verdine GL. Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. Nat Struct Mol Biol. 2006; 13:153–9. [PubMed: 16415880]
46. Senavirathne G, et al. Single-stranded DNA scanning and deamination by APOBEC3G cytidine deaminase at single molecule resolution. J Biol Chem. 2012; 287:15826–35. [PubMed: 22362763]
47. Roberts SA, et al. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol Cell. 2012; 46:424–35. [PubMed: 22607975]
48. Burns MB, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature. 2013; 494:366–70. [PubMed: 23389445]
49. Clore GM, Gronenborn AM. Determining the structures of large proteins and protein complexes by NMR. Trends Biotechnol. 1998; 16:22–34. [PubMed: 9470228]
50. Sattler M, Maurer M, Schleucher J, Griesinger C. A Simultaneous ^{15}N, ^{1}HHSQC and ^{13}C, ^{1}H-HSQC with sensitivity enhancement and a heteronuclear gradient-echo. J Biomol NMR. 1995; 5:97–102. [PubMed: 22911437]
51. Delaglio F, et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6:277–93. [PubMed: 8520220]
52. Johnson BA, Blevins RA. Nmr View - a Computer-Program for the Visualization and Analysis of Nmr Data. J Biomol NMR. 1994; 4:603–614. [PubMed: 22911360]
53. Brunger AT, et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr. 1998; 54 (Pt 5):905–21. [PubMed: 9757107]
54. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR. 1999; 13:289–302. [PubMed: 10212987]
55. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996; 8:477–86. [PubMed: 9008363]
56. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph. 1996; 14:51–5. 29–32. [PubMed: 8744573]
57. Flores SC, Sherman MA, Bruns CM, Eastman P, Altman RB. Fast flexible modeling of RNA structure using internal coordinates. IEEE/ACM Trans Comput Biol Bioinform. 2011; 8:1247–57. [PubMed: 21778523]
58. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des. 1997; 11:425–45. [PubMed: 9385547]

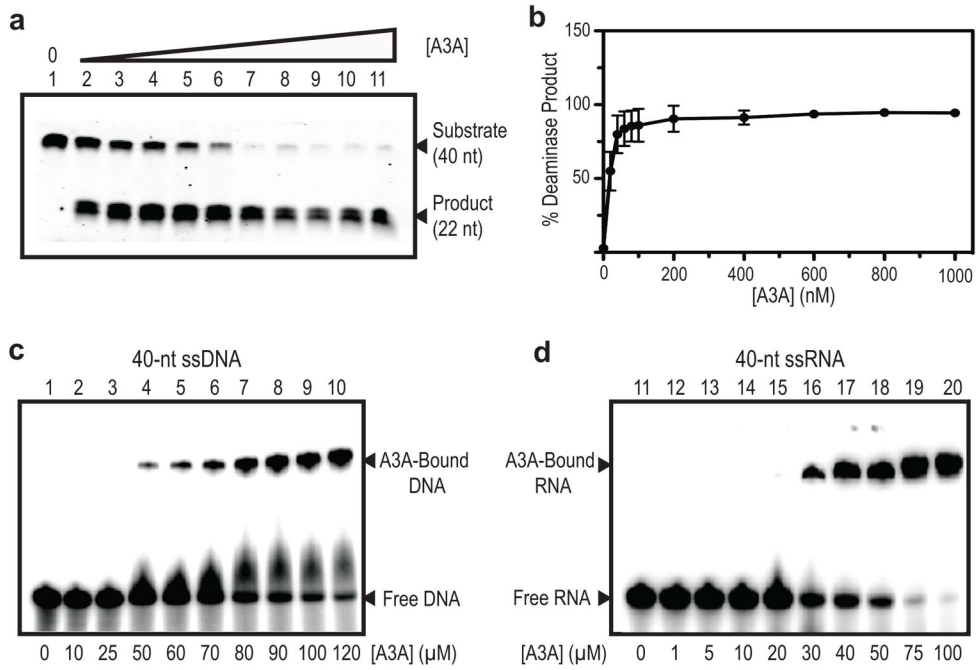

Figure 1. Deaminase activity and ssDNA and ssRNA binding of A3A

(a) Deamination of a 40-nt ssDNA as a function of A3A concentration measured in a uracil DNA glycosylase (UDG)-dependent assay. Reactions were performed as described in METHODS with increasing concentrations of A3A. Lanes: 1, no A3A; 2, 20 nM; 3, 40 nM; 4, 60 nM; 5, 80 nM; 6, 100 nM; 7, 200 nM; 8, 400 nM; 9, 600 nM; 10, 800 nM; 11, 1000 nM and relative amounts of substrate and product were assessed by gel electrophoresis. A representative gel (of three independent assays) was chosen for the figure. **(b)** Quantification of the relative amounts of deaminase product vs. A3A concentration from gel analysis as shown in **(a)**. The error bars represent the standard deviation for three independent measurements. **(c)** and **(d)** Binding of A3A to 40-nt ssDNA **(c)** or 40-nt ssRNA **(d)** evaluated by EMSA (details in METHODS). The positions of the A3A-bound as well as free DNA or RNA are indicated. A3A concentrations are listed under each lane. A representative gel (of five independent assays) was chosen for the figure.

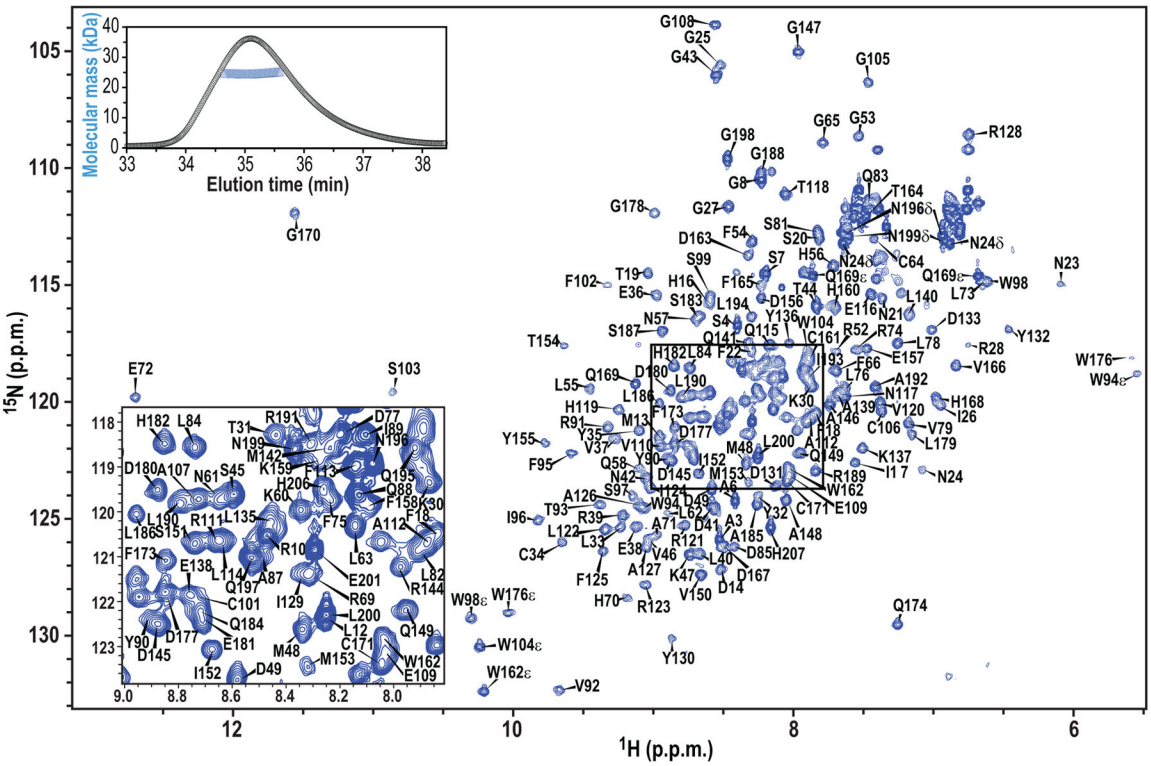

Figure 2. A3A NMR assignments

600 MHz ^{1}H-^{15}N HSQC NMR spectrum of 0.17 mM ^{13}C/^{15}N-labeled A3A in 25 mM sodium phosphate, 200 mM NaCl, pH 6.5, 25°C. Assignments are indicated by residue name and number. An expansion of the boxed region is provided in the lower left corner. SEC-MALS data are shown in the inset in the upper left corner, with the elution profile shown with black circles and the estimated molecular masses across the peak with blue triangles.

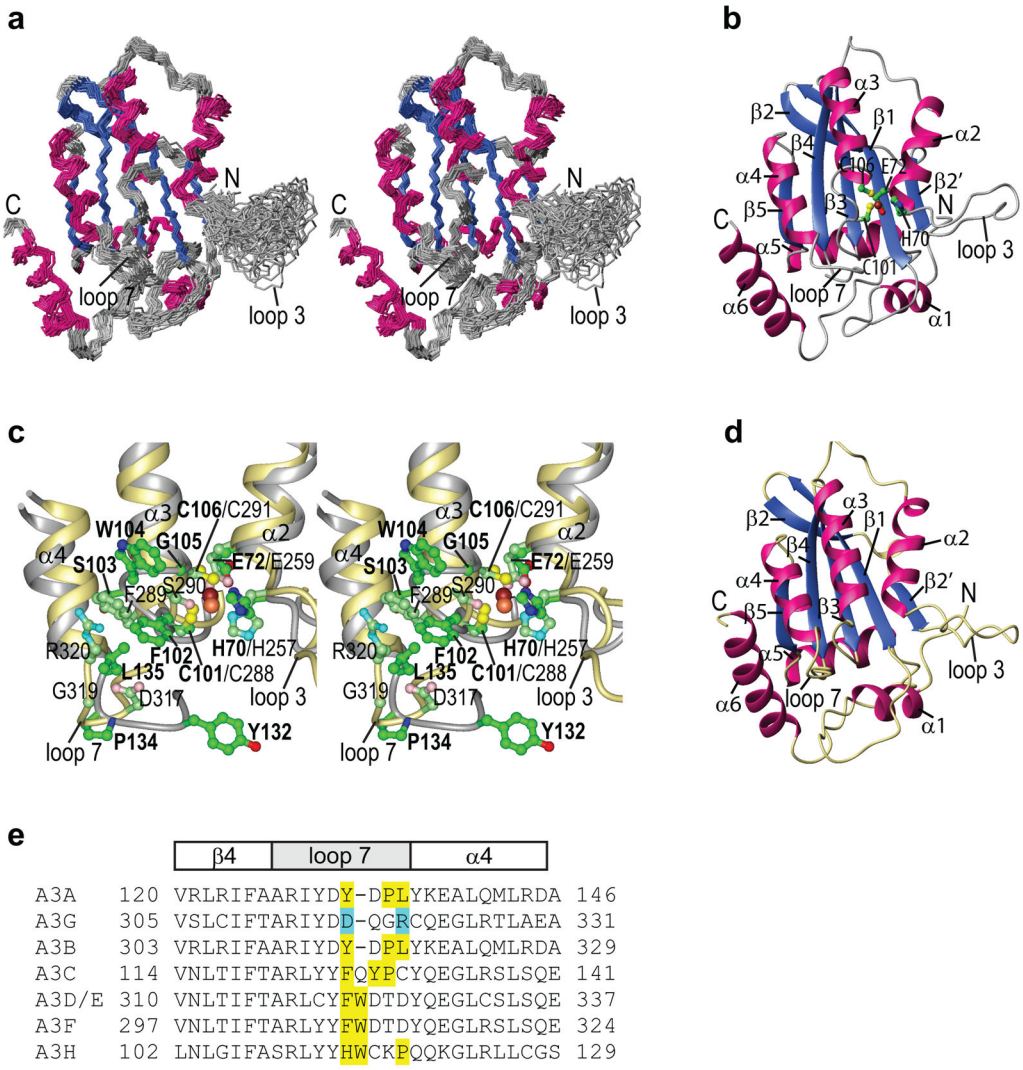

Figure 3. A3A NMR solution structure

(a) Stereo-view of the final 30 conformer ensemble (N, Cα, C′). Regions of helical and beta sheet structures are colored hot pink and royal blue, respectively, and the remainder of the structure in grey. **(b)** Ribbon representation of the lowest energy structure of the ensemble, using the same color scheme as in **(a)**. Secondary structure elements are labeled and the active site residues (H70, E72, C101, and C106) and the zinc ion are shown in ball and stick representation with carbon, nitrogen, oxygen, sulfur, and zinc atoms in green, blue, red, yellow, and brown, respectively. **(c)** Stereo-view of the superimposition of the active site regions of the current A3A NMR and the A3G-CTD (PDB: 3IR2) X-ray structures. The backbone traces of A3A and A3G-CTD are colored grey and khaki, respectively. Side chains are shown in ball-and-stick representation with carbon, nitrogen, oxygen, sulfur and zinc atoms of A3A and A3G in green, blue, red, yellow and brown, and pale green, cyan, pink, yellow and orange, respectively. A3A residues are labeled in bold. **(d)** Ribbon representation of the A3G-CTD (A3G191-384-2K3A, PDB: 3IR2) X-ray structure[35]. (**e**) Amino acid sequences of the loop 7 region in different A3 proteins. Large, hydrophobic

residues are highlighted in yellow and the polar residues D317 and R320 in A3G are highlighted in cyan.

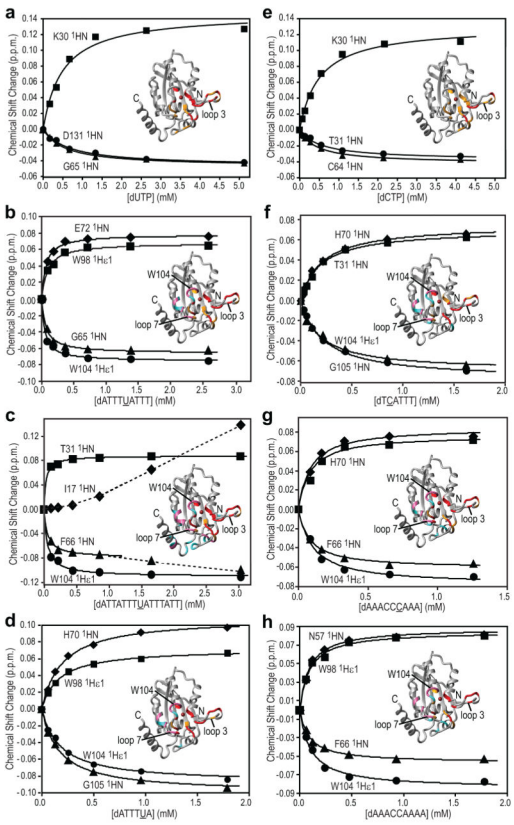

Figure 4. Binding of A3A to mononucleotides and ssDNAs

Titration curves for representative HN resonances and binding site mapping (inset) for binding of dUTP **(a)**, ATTTUATTT **(b)**, ATTATTTUATTTATT **(c)**, ATTTUA **(d)**, dCTP **(e)**, TCATTT **(f)**, AAACCCAAA **(g)** and AAACCAAAA **(h)**. A3A residues whose resonances exhibit large ^{1}H,^{15}N-combined chemical shift changes upon nucleotide addition are colored red (>0.05 p.p.m.) and orange (0.028–0.050 p.p.m.). Those only affected by ssDNA binding, but not dCTP/dUTP, are shown in dark pink (>0.050 p.p.m) and cyan (0.028–0.050 p.p.m.). All ^{1}H-^{15}N HSQC spectra were recorded at 25°C using 25 mM sodium phosphate buffer, pH 6.9. ^{1}H,^{15}N-combined chemical shift changes were calculated using $\sqrt{\Delta\delta_{HN}{}^{2}+(\Delta\delta_{N}/6)^{2}}$, $\Delta\delta HN$ and $\Delta\delta N$ *comprising* ^{1}HN ^{15}N chemical shift differences observed for A3A before and after adding ligands (~90% saturation).

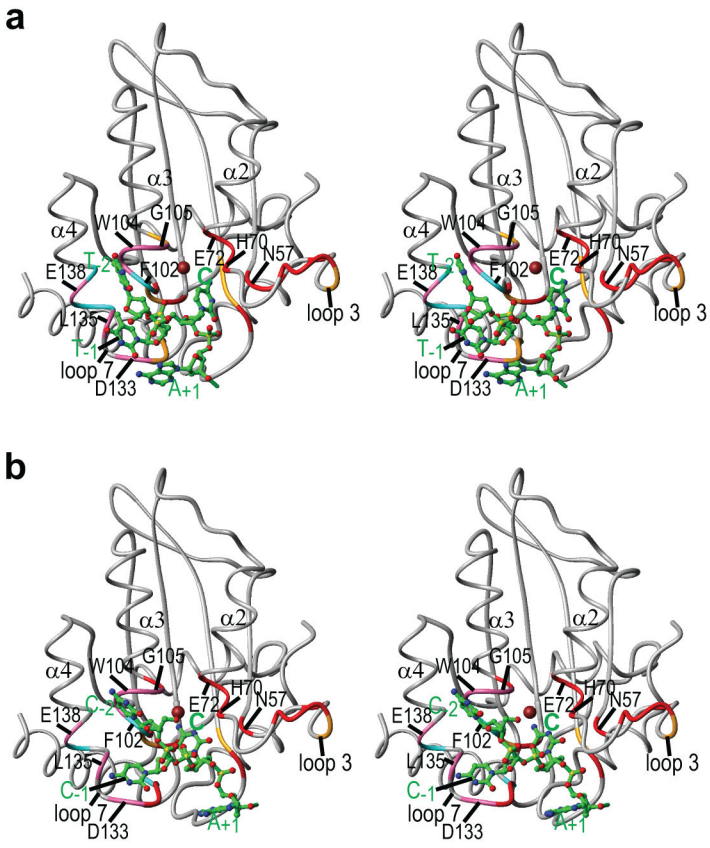

Figure 5. Model of A3A complexed with TTCAT or CCCAA

Stereo-views of the A3A backbone structure and the 5-nt ssDNAs (TTCAT **(a)** and CCCAA **(b)),** as well as interacting residues (from Fig. 4b or Fig. 4g) in ball-and-stick representation, with carbon, nitrogen, oxygen, phosphorus, and zinc atoms in green, blue, red, gold, and brown, respectively. The 3′-end thymidine (**a**) and adenosine (**b**), that appear to be random, are omitted for clarity.

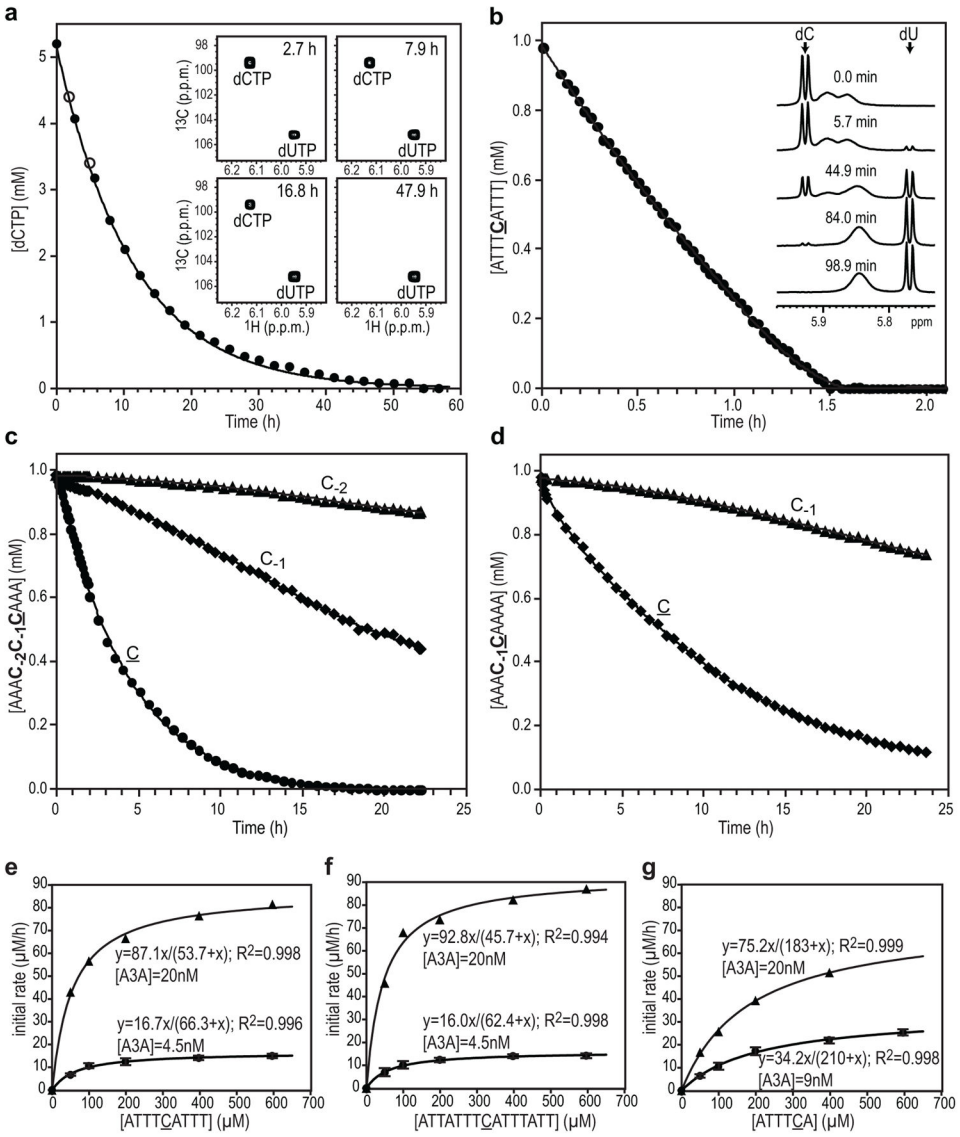

Figure 6. A3A-catalyzed deamination of dCTP and several ssDNA substrates monitored by real-time NMR

Concentrations of dCTP **(a)**, 9-nt ssDNA ATTTCATTT **(b)**, AAAC-2C-1CAAA **(c)** and AAAC-1CAAAA **(d)** vs. incubation time are provided. All concentrations of unreacted substrates (cytidine) and end products (uridine) were determined by measuring the intensities of the ^{13}C-5-^{1}H resonances of cytosine and uracil in the 2D ^{1}H-^{13}C HSQC spectra (**a**; 600MHz) or the ^{1}H-5 resonances in 1D ^{1}H spectra (**b–d**; 900 MHz), as a function of time The A3A concentration was 0.17 mM (**a**) or 0.197 μM **(b–d)**. Best fit curves are shown by a solid line. Representative 2D ^{1}H-^{13}C HSQC (**a**) or 1D ^{1}H NMR (**b**) spectra acquired at the indicated times are shown in the inset. (**e–g**), Kinetics of A3A-catalyzed deamination reactions for ATTTCATTT (**e**), ATTATTTCATTTATT (**f**) and ATTTCA (**g**): initial reaction rates (<5%) are plotted vs. substrate concentrations. Reactions were monitored by 1D ^{1}H real-time NMR (900MHz) at 25°C in 25 mM sodium phosphate buffer, pH 6.9. Two different A3A concentrations were used: **e** and **f**, 4.5 nM (circles) and 20 nM

(triangles); **g**, 9.0 nM (circles) and 20 nM (triangles). Two independent experiments were performed with 4.5 nM and 9 nM A3A and average values with standard deviations are shown. Reactions with the higher A3A concentration (20 nM) resulted in K_M and k_{cat} values very similar to those obtained with low A3A concentrations (4.5 or 9 nM).

Table 1

Statistics for the final 30 conformer ensemble of A3A

Number of NOE distance constraints	
Intra-residue (i−j=0)	1114
Sequential (|i−j|=1)	655
Medium range (2≤|i−j|≤4)	309
Long range (|i−j|≥5)	749
Total	2827
Number of hydrogen bond constraints	152
Number of dihedral angle constraints	
ϕ	151
ψ	149
Total	300
Structural Quality	
Violations[a]	
Distances constraints (Å)	0.016 ± 0.002
Dihedral angles constraints (°)	0.299 ± 0.037
Deviation from idealized covalent geometry	
Bond lengths (Å)	0.001 ± 0.000
Bond Angles (°)	0.398 ± 0.006
Improper torsions (°)	0.222 ± 0.005
Average r.m.s.d. of atomic coordinates (Å)[b]	
Backbone heavy atoms	0.60 ± 0.05
All heavy atoms	1.18 ± 0.05
Ramachandran plot analysis (%)[c]	
Most favorable region	73.7 ± 2.3
Additional allowed regions	22.5 ± 2.6
Generously allowed regions	2.9 ± 1.1
Disallowed regions	0.9 ± 0.6

[a] No individual member of the ensemble exhibited distance violations >0.5 Å or dihedral angle violations >5 °

[b] Average r.m.s. deviation of atomic coordinates for residues (11–57 and 70–194) with respect to the mean structure. A3A regions (1–9, 58–69 and 195–199) were excluded from the statistics because they exhibit a flexible, random-coil conformation.

[c] Statistics were calculated using PROCHECK for full-length A3A (residues 1–199).

Table 2

Dissociation and catalytic constants for A3A interaction with ss deoxymono and deoxyoligonucleotides[a]

Nucleotide sequences	K_d (μM)	K_M (μM)	k_{cat} (1/min)	k_{cat}/K_M (1/M•s)
dCTP	536±72	600[c]	0.033±0.003[c]	0.91
dUTP	578±115			
dATTT**C**ATTT	58±8[b]	66±7	71±6	1.8×10^4
dATTATTT**C**ATTTATT	57±11[b]	62±5	66±5	1.8×10^4
dATTT**C**A	184±20[b]	210±17	65±6	5.2×10^3
dT**C**ATTT	194±18[b]	219±29	55±5	4.2×10^3
dAAA**CC**CAAA	91±15[b]	100[c]	20±3[d]	3.3×10^3
dAAA**C**CAAAA	94±11[b]	100[c]	13±4[d]	2.2×10^3
dAAA**C**AAAAA	161±19[b]	192±32	5.5±1.1	4.7×10^2

[a] Dissociation constants (K_d) and catalytic constants (Michaelis-Menten) derived from NMR data at 25 °C (pH 6.9), as described in METHODS.

[b] he K_d values for these dC-containing oligonucleotides reflect values for their dU-containing products since these substrates are rapidly deaminated under the NMR experimental conditions (A3A ~ 0.1mM).

[c] Since accurate measurements of K_M are precluded when more than one dC is present, the K_M values of these substrates are assumed to be similar to the K_d values and are set to 100 (K_d values rounded up to the nearest hundred). This assumption is based on data obtained with substrates containing a single dC.

[d] k_{cat} values are derived from the initial rate in a sample with 1000 μM substrate and 0.197 μM A3A.