

ARTICLE

Received 29 Jan 2013 | Accepted 12 Apr 2013 | Published 21 May 2013

DOI: 10.1038/ncomms2883

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

In-Ja L. Byeon<sup>1,2</sup>, Jinwoo Ahn<sup>1,2</sup>, Mithun Mitra<sup>3</sup>, Chang-Hyeock Byeon<sup>1,2</sup>, Kamil Hercík<sup>3,†</sup>, Jozef Hritz<sup>1,†</sup>, Lisa M. Charlton<sup>1,2</sup>, Judith G. Levin<sup>3,\*</sup> & Angela M. Gronenborn<sup>1,2,\*</sup>

Human APOBEC3A is a single-stranded DNA cytidine deaminase that restricts viral pathogens and endogenous retrotransposons, and has 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 APOBEC3A'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 nuclear magnetic resonance solution structure of APOBEC3A and show that the critical interface for interaction with single-stranded DNA substrates includes residues extending beyond the catalytic centre. Importantly, by monitoring deaminase activity in real time, we find that A3A displays similar catalytic activity on APOBEC3A-specific TTCA- or A3G-specific CCCA-containing substrates, involving key determinants immediately 5' of the reactive C. Our results afford novel mechanistic insights into APOBEC3A-mediated deamination and provide the structural basis for further molecular studies.

<sup>1</sup>Department of Structural Biology, University of Pittsburgh School of Medicine, 3501 Fifth Avenue, Pittsburgh, PA 15261, USA. <sup>2</sup>Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA. <sup>3</sup>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. † Present addresses: Fraunhofer IME – Division Molecular Biology, Biological Operating Systems: Infectious Diseases, Forckenbeckstraße 6, 52074 Aachen, Germany (K.H.); CEITEC, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic (J.H.). \* These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.M.G. (email: amg100@pitt.edu).

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 have 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<sup>1–3</sup>. 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 HX<sub>1</sub>EX<sub>23–24</sub>CX<sub>2–4</sub>C motifs, where X is any amino acid<sup>4</sup>. The histidine and cysteine residues coordinate Zn<sup>2+</sup>, while glutamic acid is thought to function as a proton shuttle during the deaminase reaction<sup>5</sup>.

The single-domain A3A protein, the subject of this study, has multiple activities. A3A degrades foreign DNA introduced into human cells<sup>6,7</sup> and blocks replication of exogenous viruses such as human papilloma virus<sup>8</sup>, Rous sarcoma virus<sup>9</sup>, parvoviruses<sup>10,11</sup> and human T-lymphotropic virus type 1 (ref. 12). In addition, it strongly inhibits retrotransposition of LINE-1, *Alu* and LTR retroelements<sup>10,13–16</sup>, which cause insertional mutations. Unlike A3G (ref. 17), A3A is capable of deaminating 5-methylcytosine<sup>18,19</sup>, an epigenetic marker in genomic DNA, as well as inducing cell cycle arrest<sup>20</sup> and somatic hypermutation of nuclear and mitochondrial DNAs in a dynamic interplay between A3 editing and DNA catabolism<sup>20,21</sup>. 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<sup>22</sup>.

A3A is highly expressed in cells of the myeloid lineage, such as monocytes and macrophages, and its expression is upregulated by treatment with interferon- $\alpha$ <sup>10,23–26</sup>. Interestingly, silencing of A3A in monocytes is associated with increased susceptibility to human immunodeficiency virus (HIV)-1 infection, suggesting that the presence of A3A may be protective against HIV-1 (ref. 23). Recent studies indicate that endogenous A3A in

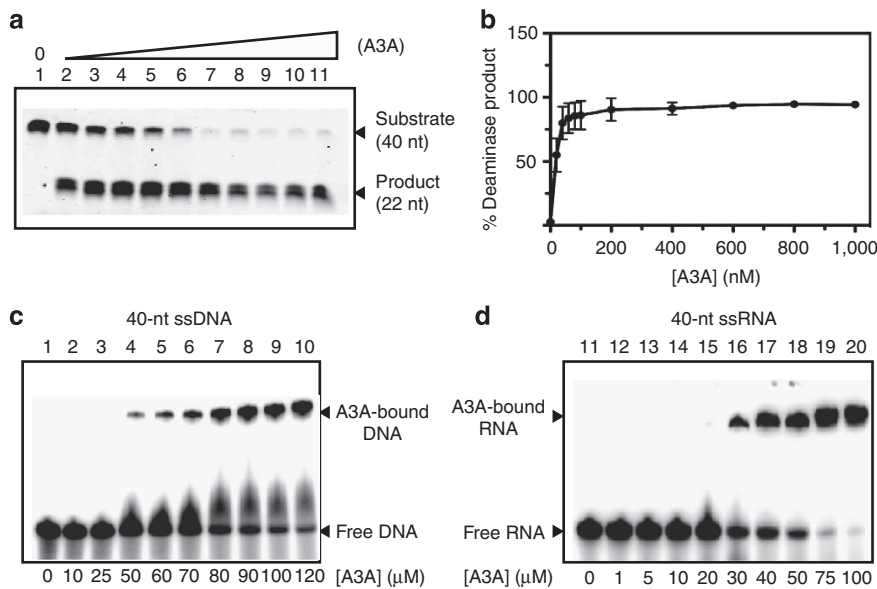
macrophages restricts HIV-1 replication by reducing synthesis of viral DNA during reverse transcription<sup>27</sup>. This result is consistent with an independent observation that HIV-1 transcripts in interferon- $\alpha$ -treated infected macrophages seem to be edited predominantly by A3A (ref. 28).

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 centre for its deaminase activity<sup>17,29,30</sup>, was solved by nuclear magnetic resonance (NMR)<sup>31–33</sup> and X-ray crystallography<sup>34,35</sup>, 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<sup>36</sup>.

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<sub>6</sub>-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 labelled 40-nt



**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 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 versus A3A concentration from gel analysis as shown in (a). The error bars represent the s.d. for three independent measurements. (c,d) Binding of A3A to 40-nt ssDNA (c) or 40-nt ssRNA (d) evaluated by electrophoretic mobility shift assay (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. In each case, a representative gel (of five independent assays) was chosen for the figure.

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.

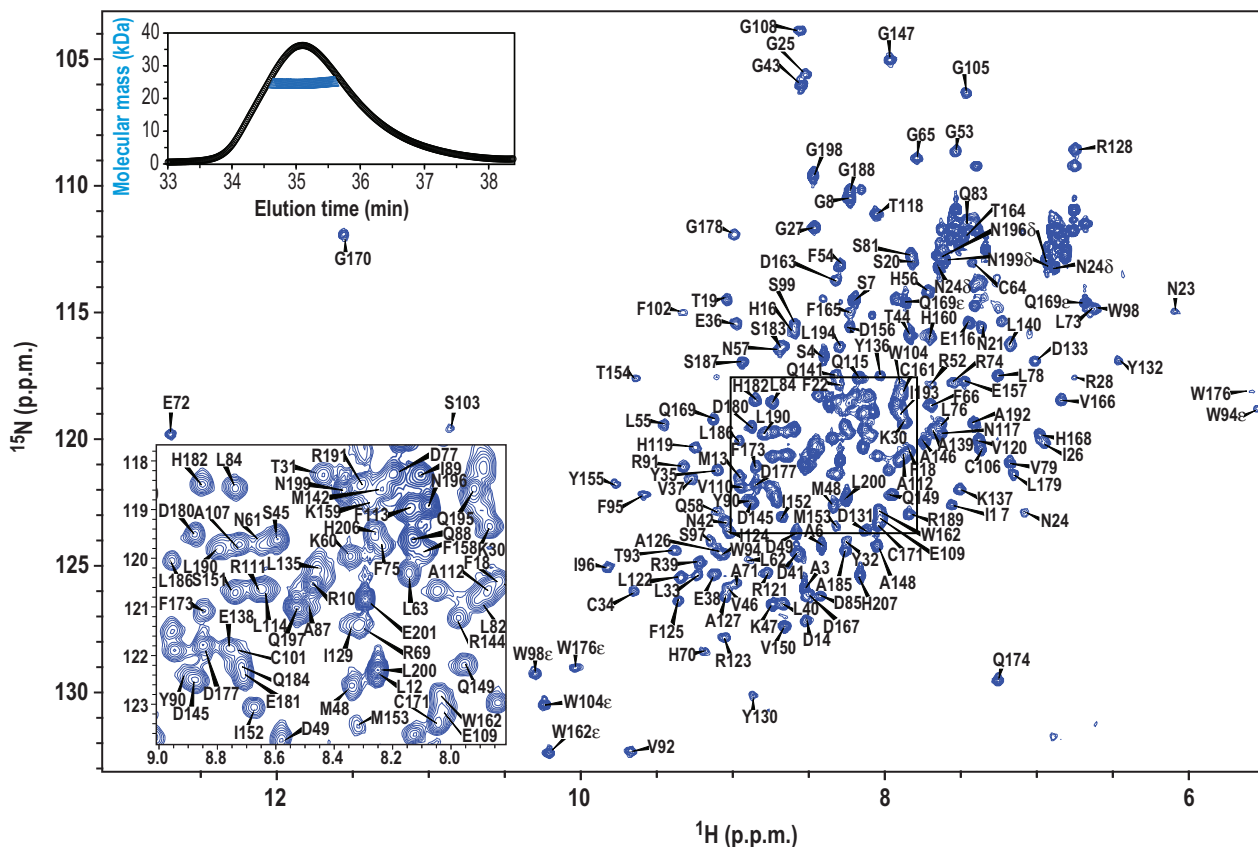
Binding of A3A to ssDNA and ssRNA was evaluated in electrophoretic mobility shift assay experiments with  $^{32}\text{P}$ -labelled 40-nt oligonucleotides and varying concentrations of A3A (Fig. 1c,d). Complexes are seen at A3A concentrations of  $\geq 50 \mu\text{M}$  with ssDNA (Fig. 1c, lane 4) and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

**NMR sample preparation and behaviour in solution.** Full-length A3A was monomeric and soluble up to  $\sim 0.2 \text{ mM}$  at pH 6.5, 200 mM NaCl, and the  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) spectrum exhibited well dispersed, narrow resonances (Fig. 2). Nearly complete backbone and side-chain assignments were obtained using unlabelled, uniformly  $^{15}\text{N}$ -,  $^{13}\text{C}/^{15}\text{N}$ - and  $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labelled A3A with specific-protonations (see Methods) for three different conditions: (1)  $\sim 0.2 \text{ mM}$  A3A at pH 6.5, 200 mM NaCl (Fig. 2); (2)  $\sim 0.2 \text{ mM}$  A3A at pH 6.9; and (3)  $\sim 0.3 \text{ mM}$  A3A at pH 8.1, all in 25 mM sodium phosphate

buffer. As essentially identical nuclear Overhauser effect (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 on the basis of 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  $0.60 \pm 0.05$  and  $1.18 \pm 0.05 \text{ \AA}$  with respect to the mean coordinate positions for the backbone (N,  $\text{C}_\alpha$ ,  $\text{C}'$ ) and all heavy atoms, respectively (Table 1). A stereoview 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  $\beta$  sheet of five strands (Fig. 3a,b), common to all APOBEC proteins whose structures are known; that is, A3G-CTD, A3C and A2 (Supplementary Table S1). The structured region of A3A is limited to residues 10–194 (Fig. 3a,b), as 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<sub>6</sub>-tag residues, indicating a substantial degree of flexibility. In addition to the termini, the loop connecting  $\beta 2'$  and  $\alpha 2$  (loop 3, residues 57–70; Fig. 3a) is also highly plastic and undergoes motions on an



**Figure 2 | A3A NMR assignments.** 600 MHz  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectrum of 0.17 mM  $^{13}\text{C}/^{15}\text{N}$ -labelled 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. Size-exclusion chromatography/multi-angle light scattering 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.

**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 ( $2 \leq  i-j  \leq 4$ )	309
Long range ( $ i-j  \geq 5$ )	749
Total	2827
Number of hydrogen bond constraints	152
Number of dihedral angle constraints	
$\phi$	151
$\psi$	149
Total	300
Structural quality	
Violations*	
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) <sup>†</sup>	
Backbone heavy atoms	0.60 ± 0.05
All heavy atoms	1.18 ± 0.05
Ramachandran plot analysis (%) <sup>‡</sup>	
Most favourable 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

\*No individual member of the ensemble exhibited distance violations >0.5 Å or dihedral angle violations >5°.

<sup>†</sup>Average r.m.s.d. 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.

<sup>‡</sup>Statistics were calculated using PROCHECK for full-length A3A (residues 1–199).

intermediate ( $\mu$ -ms) timescale, as many amide resonances exhibit severe line broadening, and are of low intensity (Q58, N61, L62) or entirely missing (A59, Y67 and G68) in the  $^1\text{H}$ - $^{15}\text{N}$  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)<sup>35</sup> (Fig. 3c,d), exhibiting an average pairwise 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)<sup>36</sup>, and wild-type A3G-CTD (residues 197–380, PDB: 3E1U/3IQS)<sup>34</sup>, 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 (ref. 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<sup>31–34</sup>. Our current A3A structure clearly possesses an interrupted  $\beta$ 2 strand ( $\beta$ 2-bulge- $\beta$ 2'; Fig. 3b and Supplementary Fig. S1), as well as an N-terminal  $\alpha$ -helix ( $\alpha$ 1, 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 Zn-coordinating cysteines (C101 and C106) in the active site of 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  $\alpha$ 3, with W104 bulging out, causing positioning of the backbone and/or side chain 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 have an important role in substrate selection (TC or CC)<sup>38</sup>. In A3A, loop 7 (127–135) adds additional hydrophobic residues (Y132, P134 and L135), whereas 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 characteristic of this region is unique for A3G (Fig. 3e).

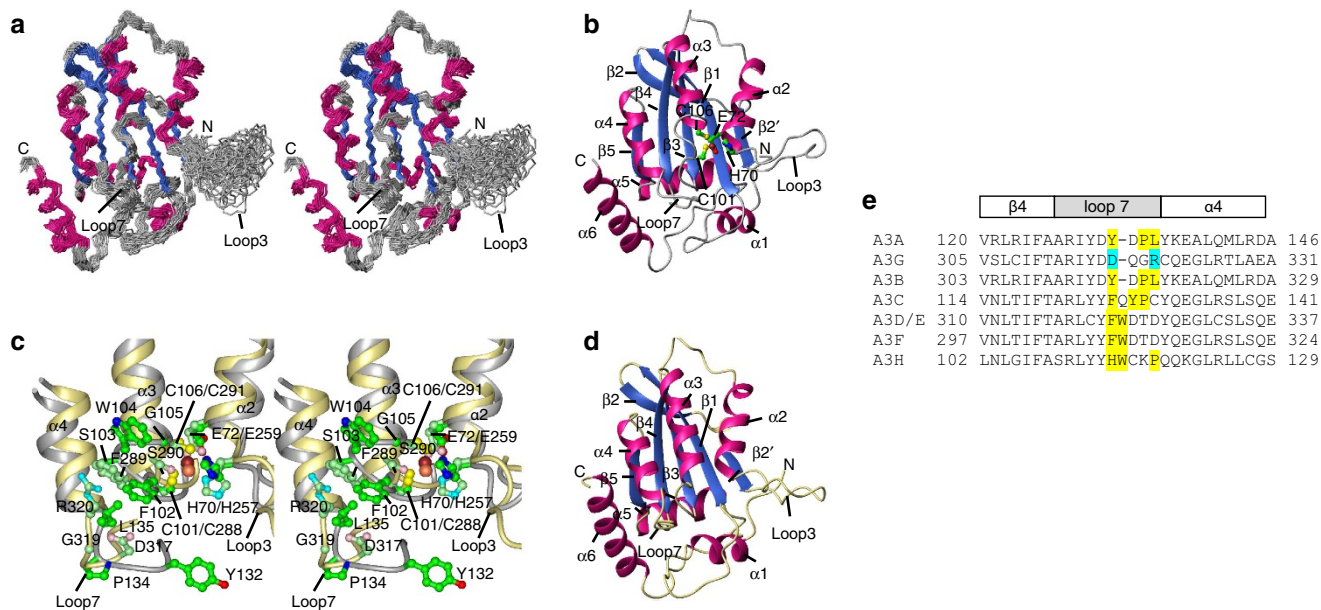
**A3A interacts with dCTP and dUTP.** Although A3A is a well-known cytidine deaminase, the determinants of substrate specificity (that is, preferential recognition of TC (refs 7,10,39)) have not been characterized in detail. As an initial approach, we investigated binding of A3A to dCTP, several ssDNA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

dUTP (Fig. 4a) and dCTP (Fig. 4e) binding was monitored by  $^1\text{H}$ - $^{15}\text{N}$  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  $536 \pm 72 \mu\text{M}$  and  $578 \pm 115 \mu\text{M}$ , respectively (Table 2). Note that as 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 5'-ATTTCATTT-3' (~0.1–3 mM) with A3A (~0.15 mM) under our NMR conditions resulted in conversion to 5'-ATTTUATTT-3' within <1 min, rendering it impossible to measure a true  $K_d$  value for 5'-ATTTCATTT-3'. However, given the similarity between dUTP and dCTP binding to A3A (Fig. 4a versus Fig. 4e), it seemed reasonable to assume that 5'-ATTTUATTT-3' and 5'-ATTTCATTT-3' would also have similar binding properties. The  $K_d$  value for the ATTTUATTT product was  $58 \pm 8 \mu\text{M}$  (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 coloured in hot pink ( $\Delta\delta > 0.050$  p.p.m.) or cyan ( $\Delta\delta$  between 0.028 and 0.050 p.p.m.) in the ribbon diagram). Interestingly, the two-





**Figure 3 | A3A NMR solution structure.** (a) Stereoview of the final 30 conformer ensemble (N, C $\alpha$ , C'). Regions of helical and beta sheet structures are coloured 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 colour scheme as in (a). Secondary structure elements are labelled and the active site residues (H70, E72, C101 and C106) and the Zn<sup>2+</sup> ion are shown in ball-and-stick representation with carbon, nitrogen, oxygen, sulphur and Zn atoms in green, blue, red, yellow and brown, respectively. (c) Stereoview 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 coloured grey and khaki, respectively. Side chains are shown in ball-and-stick representation with carbon, nitrogen, oxygen, sulphur and Zn atoms of A3A and A3G in green, blue, red, yellow and brown, and pale green, cyan, pink, yellow and orange, respectively. A3A residues are labelled in bold. (d) Ribbon representation of the A3G-CTD (A3G191-384-2K3A, PDB: 3IR2) X-ray structure<sup>35</sup>. (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.

residue insertion (W104/G105), unique to A3A, is part of the interface: the W104 backbone amide ( $\Delta\delta = 0.122$  p.p.m.) and the W104  $\epsilon 1$  side chain resonance ( $\Delta\delta = 0.079$  p.p.m.), as well as the G105 ( $\Delta\delta = 0.066$  p.p.m.) and S103 ( $\Delta\delta = 0.072$  p.p.m.) amide resonances exhibit the largest perturbations. In addition, loop 7 and helix  $\alpha 4$  are also part of the extended interface, as effects on the amide resonances of residues D133 ( $\Delta\delta = 0.112$  p.p.m.), L135 ( $\Delta\delta = 0.102$  p.p.m.) and E138 ( $\Delta\delta = 0.079$  p.p.m.) were observed. Additional binding studies of A3A using a 15-nt oligonucleotide, 5'-ATTATTTUATTTATT-3' yielded essentially the same binding site (Fig. 4c, inset) and affinity ( $K_d = 57 \pm 11 \mu\text{M}$ ; Fig. 4c, Table 2), as found with the 9-mer.

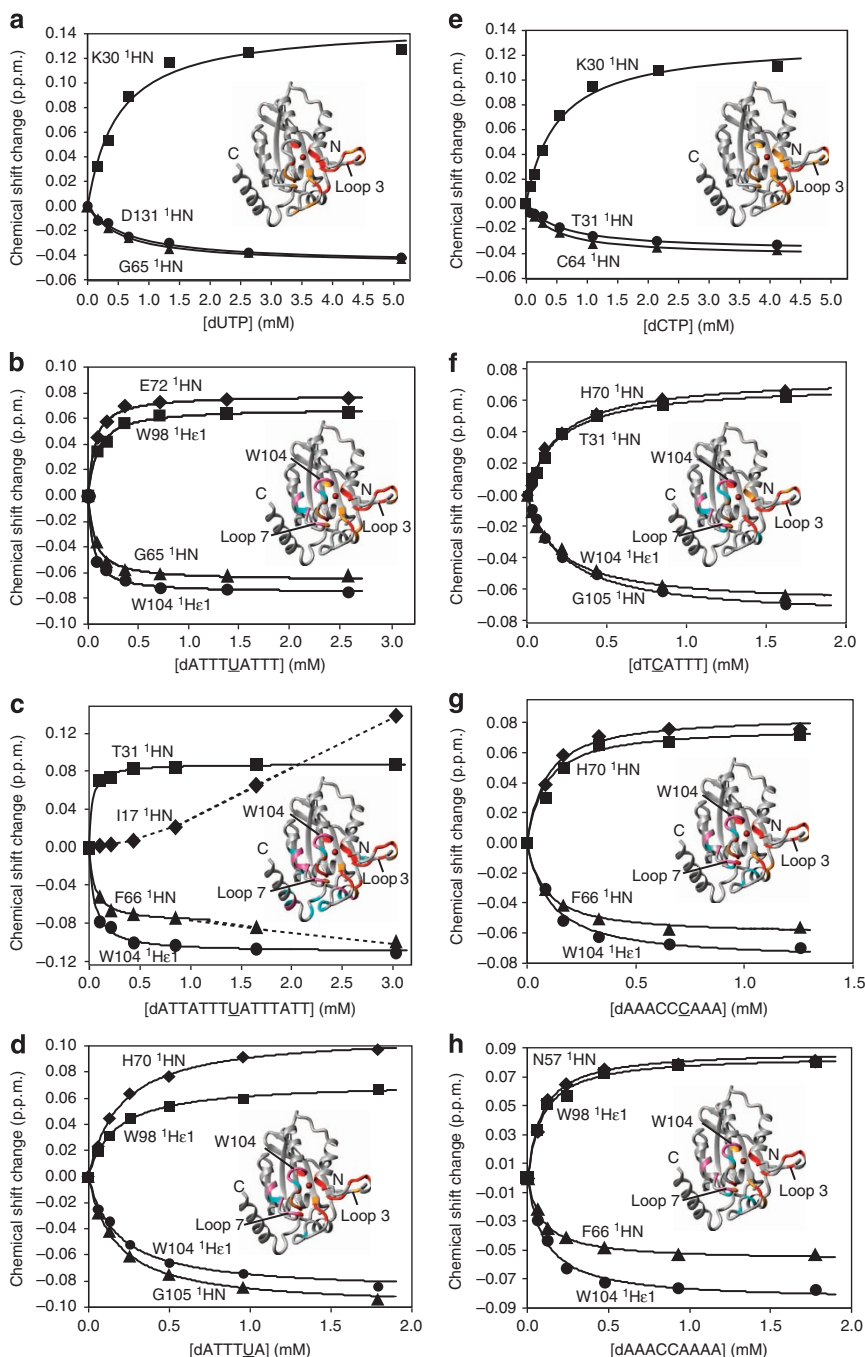
Although smaller perturbations were observed for some additional resonances, these changes were non-saturable and most likely reflect non-specific binding (for example, the I17 resonance marked by a dashed line in Fig. 4c). The non-specific binding site of the 15-mer oligonucleotide mapped onto the A3A structure involves residues 12–19 ( $\alpha 1$ ), 36–37, 40–42 ( $\beta 1$ ), 44–49 ( $\beta 2$ ), 53–56 ( $\beta 2'$ ) and 179–186 ( $\alpha 6$ ) (Supplementary Fig. S2). To validate the binding data, the interaction of an A3A catalytic site mutant, E72Q, with a 15-nt substrate 5'-(ATTATTTCATTTATT)-3' was evaluated. The mutant exhibited the same affinity ( $K_d = 55 \pm 11 \mu\text{M}$ , data not shown) as wild-type A3A with the 5'-ATTATTTUATTTATT-3' product, confirming that binding of substrate and product occurs with essentially identical affinities.

Further titration experiments using smaller oligonucleotides (5'-ATTT(C/U)A-3', Fig. 4d; 5'-TCATTT-3', Fig. 4f) delineated a binding site similar to the one observed with the nona-nucleotide. This implies that three nucleotides, (TCA), constitute the essential moiety for binding to A3A. The affinities of the hexanucleotides, however, were approximately threefold

weaker ( $K_d \sim 190 \mu\text{M}$ ) than the one measured for 5'-ATTT(C/U)ATTT-3' ( $K_d = 58 \pm 8 \mu\text{M}$ ; Table 2), suggesting a stabilizing effect of the flanking nucleotides in the complex.

Interaction of A3A with a 9-nt substrate, 5'-AAACCCAA-3', containing the A3G deaminase-specific recognition site<sup>2</sup>, maps to the same surface (Fig. 4g, inset) as 5'-ATTT(C/U)ATTT-3' (Fig. 4b, inset). This suggests that both dT and dC at the -1 and -2 positions, that is, pyrimidine bases, can interact with the extended interface next to the catalytic site of A3A. However, slightly weaker binding ( $K_d = 91 \pm 15 \mu\text{M}$ , Fig. 4g, Table 2) was observed for 5'-AAACCCAAA-3' compared with 5'-ATTTTCATTT-3'. Replacement of the central CCC in 5'-AAACCCAAA-3' by CCA (5'-AAACCAAAA-3', Fig. 4h) or CAA (5'-AAACA AAAA-3', Table 2) resulted in identical ( $K_d = 94 \pm 11 \mu\text{M}$ ) or slightly weaker ( $K_d = 161 \pm 19 \mu\text{M}$ ) affinities, respectively.

**Structural model for A3A-ssDNA complexes.** Model structures of the A3A-oligonucleotide complexes were created by flexible docking<sup>40</sup> and selecting binding poses compatible with chemical shift perturbations (see Methods for details). Initial docking results for 5'-ATTTTCATTT-3' and 5'-AAACCCAAA-3' 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 5'-TTCAT-3' (Fig. 5a) and 5'-CCC AA-3' (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 Zn<sup>2+</sup> ion. The thymidine (T<sub>-1</sub>) immediately



**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**), T<sub>C</sub>ATT (**f**), AAACC<sub>C</sub>AAA (**g**) and AAACC<sub>A</sub>AAA (**h**). A3A residues whose resonances exhibit large <sup>1</sup>H,<sup>15</sup>N-combined chemical shift changes upon nucleotide addition are coloured red (>0.050 p.p.m.) and orange (0.028–0.050 p.p.m.). Those only affected by ssDNA binding, but not by dCTP/dUTP, are shown in dark pink (>0.050 p.p.m.) and cyan (0.028–0.050 p.p.m.). All <sup>1</sup>H-<sup>15</sup>N HSQC spectra were recorded at 25°C using 25 mM sodium phosphate buffer, pH 6.9. <sup>1</sup>H,<sup>15</sup>N-combined chemical shift changes were calculated using  $\sqrt{\Delta\delta_{\text{HN}}^2 + (\Delta\delta_{\text{N}}/6)^2}$ , with  $\Delta\delta_{\text{HN}}$  and  $\Delta\delta_{\text{N}}$  the <sup>1</sup>HN and <sup>15</sup>N chemical shift differences observed for A3A before and after adding ligands (~90% saturation).

preceding the **C** interacts with a surface formed by residues D133, P134, L135 (loop 7) and F102 (loop 5), whereas T<sub>-2</sub> contacts residues F102, S103, W104, and G105 (loop 5), L135 (loop 7) and E138 (α4). The adenosine (A<sub>+1</sub>) at the 3' side of the **C** interacts with residues D131, Y132 and D133 (loop 7). The 5-methyl groups of T<sub>-1</sub> and T<sub>-2</sub> are in close contact with several

hydrophobic residues, for example, F102, W104 and L135, suggesting that this hydrophobic interaction may contribute to tighter binding of TT<sub>C</sub>A- compared with CC<sub>C</sub>A-containing ssDNA. The model for the A3A/CC<sub>C</sub>AA complex (Fig. 5b) is very similar to that of A3A/TT<sub>C</sub>AT (Fig. 5a), including the localization of the individual nucleotides.

**Table 2 | Dissociation and catalytic constants for A3A interaction with ss deoxymono- and deoxyoligonucleotides.\***

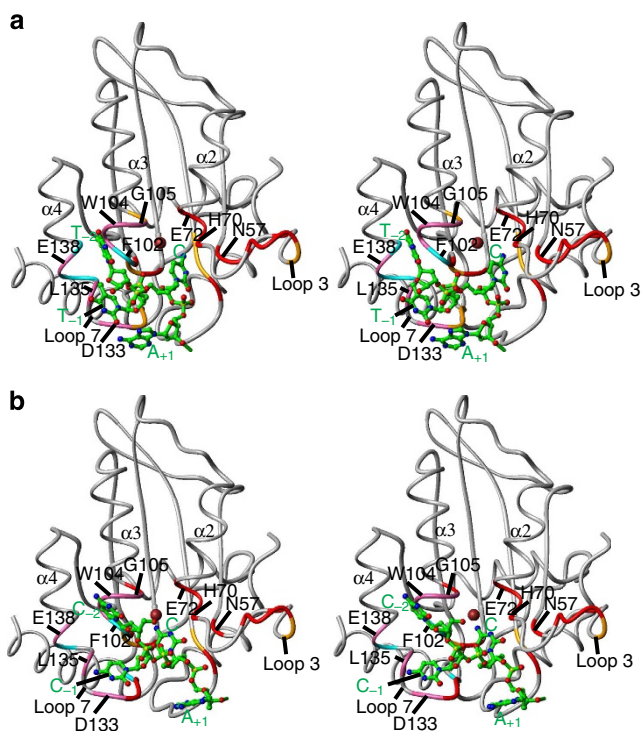
Nucleotide sequences	$K_d$ ( $\mu\text{M}$ )	$K_M$ ( $\mu\text{M}$ )	$k_{\text{cat}}$ ( $\text{min}^{-1}$ )	$k_{\text{cat}}/K_M$ ( $\text{M}^{-1}\text{s}^{-1}$ )
dCTP	536 ± 72	600 <sup>†</sup>	0.033 ± 0.003 <sup>‡</sup>	0.91
dUTP	578 ± 115			
5'-dATTTCATT-3'	58 ± 8 <sup>‡</sup>	66 ± 7	71 ± 6	$1.8 \times 10^4$
5'-dATTATTCATT-3'	57 ± 11 <sup>‡</sup>	62 ± 5	66 ± 5	$1.8 \times 10^4$
5'-dATTTCAT-3'	184 ± 20 <sup>‡</sup>	210 ± 17	65 ± 6	$5.2 \times 10^3$
5'-dTTCATT-3'	194 ± 18 <sup>‡</sup>	219 ± 29	55 ± 5	$4.2 \times 10^3$
5'-dAAACCCAAA-3'	91 ± 15 <sup>‡</sup>	100 <sup>‡</sup>	20 ± 3 <sup>§</sup>	$3.3 \times 10^3$
5'-dAAACCAAAA-3'	94 ± 11 <sup>‡</sup>	100 <sup>‡</sup>	13 ± 4 <sup>§</sup>	$2.2 \times 10^3$
5'-dAAACAAAAA-3'	161 ± 19 <sup>‡</sup>	192 ± 32	5.5 ± 1.1	$4.7 \times 10^2$

\*Dissociation constants ( $K_d$ ) and catalytic constants (Michaelis-Menten) derived from NMR data at 25 °C (pH 6.9), as described in Methods.

<sup>†</sup>As 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.

<sup>‡</sup>The  $K_d$  values for these dC-containing oligonucleotides reflect values for their dU-containing products, as these substrates are rapidly deaminated under the NMR experimental conditions (A3A ~ 0.1 mM).

<sup>§</sup> $k_{\text{cat}}$  values are derived from the initial rate in a sample with 1000  $\mu\text{M}$  substrate and 0.197  $\mu\text{M}$  A3A.



**Figure 5 | Model of A3A complexed with TTCAT or CCCAA.** Stereoviews 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 Zn atoms in green, blue, red, gold and brown, respectively. The 3'-end thymidine (a) and adenosine (b), which appear to be random, are omitted for clarity.

**Deamination by real-time NMR.** The A3A-catalyzed deamination reaction with several substrates was followed by two-dimensional (2D)  $^1\text{H}$ - $^{13}\text{C}$  HSQC (Fig. 6a) or 1D  $^1\text{H}$  (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  $\mu\text{M}$  A3A in ~50 h. Longer substrates, such as the 9-nt 5'-ATTTCATT-3', were deaminated much more rapidly (Fig. 6b), and in such cases 1D  $^1\text{H}$  NMR

spectroscopy was used (Fig. 6b inset). The 9-nt substrate (0.98 mM) was completely deaminated by 0.196  $\mu\text{M}$  A3A in ~1.7 h, with an initial rate of ~0.8  $\text{mM h}^{-1}$ . Other substrates, a 15-mer, 5'-ATTA TTTTCATT-3', and two hexanucleotides, 5'-ATTTCA-3' and 5'-TCATT-3', (Supplementary Fig. S3a-c, respectively) exhibited essentially identical rates to that observed with the nona-nucleotide.

A3A and A3G purportedly possess a difference in substrate specificity<sup>2,6,7,10,39,41</sup>. To quantify A3A's substrate specificity, we also assayed the deamination of an A3G-specific substrate, AAAC<sub>-2</sub>C<sub>-1</sub>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  $\text{mM h}^{-1}$ , only fourfold slower than the dC in the A3A-specific 5'-ATTTCATT-3' substrate. When comparing substrates with varying numbers of cytosines, such as 5'-AAACCAAAA-3' (Fig. 6c), 5'-AAACCAAAA-3' (Fig. 6d) and 5'-AAACAAAAA-3' (Supplementary Fig. S3d), deamination of the 3' dC (C) was about two- to four-fold faster in 5'-AAACCAAAA-3' 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<sub>-1</sub> and C<sub>-2</sub> are much slower, quite distinct from that of C in the reaction with the AAAC<sub>-2</sub>C<sub>-1</sub>CAAA substrate. Similar behaviour was observed with AAAC<sub>-2</sub>C<sub>-1</sub>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, for example, the C in 5'-AAAC<sub>-1</sub>CAAAA-3' (Fig. 6d) and 5'-AAACAAAAA-3' (Supplementary Fig. S3d), the reaction proceeds readily, albeit somewhat slower than with AAAC<sub>-2</sub>C<sub>-1</sub>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<sub>-1</sub> and C<sub>-2</sub> in AAAC<sub>-2</sub>C<sub>-1</sub>UAAA. Therefore, optimal substrates for A3A contain a cytidine preceded by two pyrimidine bases and followed by an adenine; that is, (T/C)(T/C)CA.

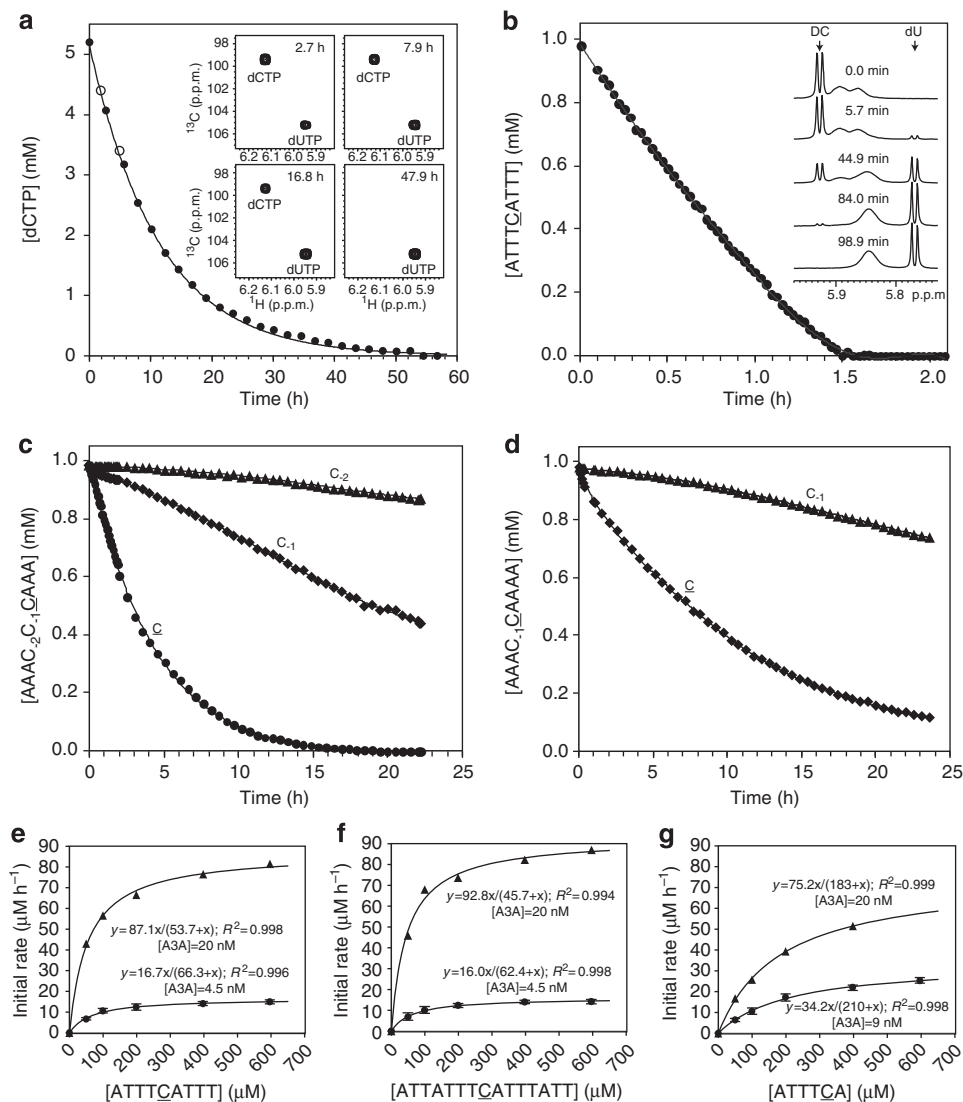
Further studies were performed to determine the catalytic constants for A3A deaminase activity on single cytidine-containing substrates; that is, 5'-ATTTCATT-3', 5'-ATTATTTTCATT-3', 5'-ATTTCAT-3', 5'-TCATT-3' and 5'-AAACAAAAA-3'. The kinetics of deamination revealed that all TCA-containing oligonucleotide substrates, regardless of length, exhibited identical turnover rates ( $k_{\text{cat}} \sim 60\text{--}70 \text{ min}^{-1}$ ; Table 2, Fig. 6e-g). The hexanucleotides, however, displayed an approximately threefold lower apparent second-order rate constant ( $k_{\text{cat}}/K_M \sim 5 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$ ) than the one measured for the 9-nt and 15-nt substrates ( $k_{\text{cat}}/K_M \sim 1.8 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ ), 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  $k_{\text{cat}}/K_M$  was seen only with a substrate that is lacking pyrimidine bases at the 5' side of the reactive C; that is, 5'-AAACAAAAA-3', caused by a large (13-fold) reduction in  $k_{\text{cat}}$  and a small (approximately threefold) increase in  $K_M$  (less-favourable 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<sup>31-36</sup> (Supplementary Table S1) and not surprisingly, there are also similarities between the activities of





**Figure 6 | A3A-catalyzed deamination of dCTP and several ssDNA substrates monitored by real-time NMR.** Concentrations of dCTP (a), 9-nt ssDNA ATTTCTATT (b), AAAC<sub>2</sub>C<sub>1</sub>CAAAA (c) and AAAC<sub>1</sub>CAAAA (d) versus incubation time are provided. All concentrations of unreacted substrates (cytosine) and end products (uridine) were determined by measuring the intensities of the <sup>13</sup>C-5-<sup>1</sup>H resonances of cytosine and uracil in the 2D <sup>1</sup>H-<sup>13</sup>C HSQC spectra (a; 600 MHz) or the <sup>1</sup>H-5 resonances in 1D <sup>1</sup>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 <sup>1</sup>H-<sup>13</sup>C HSQC (a) or 1D <sup>1</sup>H NMR (b) spectra acquired at the indicated times are shown in the inset. (e-g) Kinetics of A3A-catalyzed deamination reactions for ATTTCTATT (e), ATTATTTCTATTATT (f) and ATTTCA (g); initial reaction rates (<5%) are plotted versus substrate concentrations. Reactions were monitored by 1D <sup>1</sup>H real-time NMR (900 MHz) 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 and 9 nM A3A and average values with s.d. 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).

A3A and the A3G-CTD. Using NMR, we show that A3A binds TTCA- or CCCA-containing single-stranded oligonucleotides ( $\geq 9$  nt) with  $K_d$  values ranging from 50–100 μM (Table 2), in excellent agreement with binding affinity data ( $\sim 80$  μM) estimated from electrophoretic mobility shift assay (Fig. 1c) and with fluorescence depolarization results obtained by Love *et al.*<sup>39</sup>

Intriguingly, A3A binds ssDNA much more weakly ( $\sim 1000$ -fold) than A3G, for which  $K_d$  values ranging from 50 to 240 nM have been reported<sup>17,39,42–44</sup>. However, like A3A,  $K_d$  values (200–450 μM) were obtained for the single-domain A3G-CTD protein<sup>31,32</sup>, implying that the tighter binding of full-length A3G is associated with its double domain structure. Indeed, the A3G-N-terminal domain contains numerous positively charged amino acids that contribute to high efficiency binding to ssDNA<sup>17,29</sup>,

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 dipeptide W104-G105, loop 7 and helix  $\alpha 4$ . Surprisingly, although the W104N<sub>EH</sub> 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<sup>16,18</sup>.

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 (refs 31,32). Interestingly, 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 neighbouring ( $-1$ ,  $-2$  and  $+1$ ) nucleotides to interact locally near the catalytic site. Interestingly, there are reports of RNA bending in the crystal structure of a transfer RNA adenosine deaminase/RNA complex<sup>45</sup> and DNA contraction during A3G scanning of the ssDNA substrate<sup>46</sup>.

NMR real-time kinetic data for the deamination reaction of A3A using 5'-ATTTCATTT-3' and 5'-ATTATTTCAATT-3' as substrates yielded values of  $k_{\text{cat}} \sim 70 \text{ min}^{-1}$  and  $K_M \sim 60 \mu\text{M}$ , which differ from the values ( $k_{\text{cat}} \sim 15 \text{ min}^{-1}$  and  $K_M \sim 230 \text{ nM}$ ) obtained using a 43-nt TTCT-containing ssDNA<sup>18</sup>. These differences may be related to variations in the ssDNA sequences and experimental conditions. Interestingly, Carpenter *et al.* reported that A3A is a stronger ( $\sim 200$ -fold) deaminase than full-length A3G (ref. 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.*<sup>32</sup>, 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 5'-AAACC<sub>3</sub>AAA-3' (thought to be an A3G-specific substrate<sup>2</sup>) with only an approximate fivefold reduction in  $k_{\text{cat}}/K_M$ , compared with 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 (ref. 6). In addition, a strong bias for deamination of TC ( $\sim 50$ –70%) and CC ( $\sim 15$ –25%) dinucleotide sequences was detected in nascent HIV-1 complementary DNAs isolated from infected macrophages<sup>28</sup> and in an *in vitro* HIV-1 model replication assay, performed using purified A3A protein<sup>39</sup>. Furthermore, a preference for TC and CC substrate recognition sites was also observed for A3A editing of human T-lymphotropic virus type 1 (ref. 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.*<sup>7</sup>, it was shown that A3A mediates genomic DNA editing in human cells, but no editing site preference was detected. In contrast, Suspène *et al.*<sup>21</sup> found that A3A preferentially deaminates cytidines in TC and CC dinucleotides in genomic DNA, when cells are exposed to UDG inhibitor. APOBEC-mediated genomic DNA mutations have been implicated in carcinogenesis<sup>47</sup> and, for example, A3B was shown to be a source of DNA mutations in breast cancer<sup>48</sup>. 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<sub>6</sub>-tag (LEHHHHHH) were inserted into the *NdeI*-*XhoI* site of the

pET21 plasmid (Novagen) for expression in *Escherichia coli* Rosetta 2 (DE3). Uniform <sup>15</sup>N- and <sup>13</sup>C-labelling of the proteins was carried out by growth in modified minimal medium at 18 °C, using <sup>15</sup>NH<sub>4</sub>Cl and <sup>13</sup>C<sub>6</sub>-glucose as the sole nitrogen and carbon sources, respectively. Uniform <sup>2</sup>H-, <sup>15</sup>N- and <sup>13</sup>C-labelling of the proteins was achieved using <sup>2</sup>H<sub>2</sub>O, <sup>15</sup>NH<sub>4</sub>Cl and <sup>13</sup>C<sub>6</sub>/D<sub>7</sub>-glucose as deuterium, nitrogen and carbon sources, respectively, 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 <sup>13</sup>C/<sup>15</sup>N-tyrosine, -phenylalanine and -isoleucine (for sample 1), and <sup>13</sup>C/<sup>15</sup>N-tyrosine and -phenylalanine, unlabelled tryptophan, 2-keto-butyrates (1,2,3,4-<sup>13</sup>C, 98%; 3,3'-<sup>2</sup>H, 98%, CIL, Andover, MA, USA) and 2-keto-3-methyl-butyrates (1,2,3,4-<sup>13</sup>C, 99%; 3,4,4',4''-<sup>2</sup>H, 98%, CIL; for sample 2), respectively. These chemicals were added to the culture 1 h before 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 dithiothreitol (DTT) and 0.02% NaN<sub>3</sub>. 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-polyacrylamide gel electrophoresis. The molecular mass of the A3A proteins were confirmed by LC-ESI-TOF mass spectrometry (Bruker Daltonics, Billerica, MA, USA).

**Multi-angle light scattering.** Size-exclusion chromatography/multi-angle light scattering data were obtained at room temperature using an analytical Superdex 200 (S200) column with in-line multi-angle light-scattering refractive index (Wyatt Technology, Inc., Santa Barbara, CA, USA) and ultraviolet (Agilent Technologies, Santa Clara, CA) detectors. One hundred microlitres 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 min<sup>-1</sup>.

**Deaminase assay using fluorescent-tagged ssDNA substrates.** Deaminase assay conditions were adapted from Iwatani *et al.*<sup>17</sup> Forty microlitres 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'), labelled at its 5' end with Alexa Fluor 488, (Integrated DNA Technologies (IDT), Coralville, IA, USA) 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 1 N HCl. Ten microlitres 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.** Ten microlitres reactions, containing varying amounts of A3A, 20 nM of a 5' <sup>32</sup>P-labelled 40-nt ssDNA (JL895, identical sequence to JL913 but without the Alexa Fluor 488 label; Lofstrand Labs Ltd, Gaithersburg, MD, USA) or 40-nt ssRNA (JL931, identical sequence to JL895, except that U was substituted for T; IDT), were incubated with 4 U SUPERase IN (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 microlitre DNA Loading Gel Solution (Quality Biological, Inc.), containing bromphenol blue and xylene cyanol, was added to a control sample without A3A. Gels were run at 4 °C (5 mA) until the bromphenol blue dye had migrated  $\sim 2/3$  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 unlabelled, <sup>13</sup>C/<sup>15</sup>N- or <sup>2</sup>H/<sup>13</sup>C/<sup>15</sup>N-labelled A3A with two types of selective protonations (see above). At pH 6.9 and 8.1, A3A showed higher solubility ( $\sim 0.5$  and  $\sim 1$  mM, respectively) than that at pH 6.5 ( $\sim 0.2$  mM). However, at these higher protein concentrations, soluble aggregation occurred, as evidenced by severe line broadening. We therefore performed all of our NMR experiments at concentrations of  $\sim 0.2$ –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 <sup>1</sup>H-<sup>15</sup>N HSQC, <sup>1</sup>H-<sup>13</sup>C HSQC and nuclear Overhauser enhancement spectroscopy (NOESY) and three-dimensional (3D) HNCACB, HN(CO)CACB, HNCA, HN(CO)CA and HCCH-total correlation spectroscopy experiments<sup>49</sup>. Distance constraints were derived from 3D simultaneous <sup>13</sup>C- and <sup>15</sup>N-edited NOESY (ref. 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<sup>51</sup>, 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)<sup>52</sup>.

**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 (ref. 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  $\beta$ -sheets, and 300  $\phi$  and  $\psi$  backbone torsion angles from TALOS calculations<sup>54</sup>. In addition, a  $Zn^{2+}$  ion was added at a late stage in the calculations to coordinate with H70, C101 and C106 (refs 31,32) using constraints based on the X-ray structures of A3G-CTD (refs 34,35). Five hundred and twelve structures were generated and the 30 lowest energy structures were selected and analyzed using PROCHECK-NMR (ref. 55) (Table 1). All structure figures were generated with MOLMOL (ref. 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 <sup>13</sup>C/<sup>15</sup>N-labelled A3A. 100 mM dCTP and dUTP stock solutions were purchased from Promega (Madison, WI, USA) and HPLC-purified DNA oligonucleotides were obtained from IDT or Midland Co. (Midland, TX, USA). A series of 2D <sup>1</sup>H-<sup>15</sup>N HSQC titration spectra were acquired and binding isotherms were obtained by plotting <sup>1</sup>HN proton chemical shift change versus nucleotide concentrations for 5–8 unambiguously traceable amide resonances. Dissociation constants were calculated by non-linear best fitting of the isotherms using KaleidaGraph (Synergy Software, Reading, PA, USA), and values for 5–8 resonances were averaged.

**Real-time studies of A3A-catalyzed deamination.** A series of 2D <sup>1</sup>H-<sup>13</sup>C HSQC and/or 1D <sup>1</sup>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 <sup>13</sup>C-5-<sup>1</sup>H resonances (volumes in 2D <sup>1</sup>H-<sup>13</sup>C HSQC spectra) and/or <sup>1</sup>H-5 resonances (integrals in 1D <sup>1</sup>H spectra) of cytosine and uracil were used for quantification. 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  $\mu$ M).  $k_{cat}$  ( $V_{max}/[A3A]$ ) and  $K_M$  values were obtained using the Michaelis-Menten module in Kaleidagraph.

**Molecular docking.** Structures of 5'-TT $\overline{C}$ AT-3', 5'-CC $\overline{C}$ AA-3', 5'-ATT $\overline{C}$ ATTT-3' and 5'-AAACCCAAA-3' ssDNAs were generated by MacroMoleculeBuilder (version 2.8)<sup>57</sup>. The oligonucleotide and A3A NMR structures were converted to a mol2 file format using the Hermes programme (version 1.4) and standard Tripos atom and bond types. Docking was performed with the flexible docking programme GOLD version 5.1 (ref. 40), in combination with the Chemscore scoring function<sup>58</sup>. The  $Zn^{2+}$  ion was chosen as the centre 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: 50,000 for 5'-TT $\overline{C}$ AT-3' and 5'-CC $\overline{C}$ AA-3', and 150,000 for 5'-ATT $\overline{C}$ ATTT-3' and 5'-AAACCCAAA-3'. Docking poses (1250) 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 5'-TT $\overline{C}$ AT-3' (Fig. 5a) and 5'-CC $\overline{C}$ AA-3' (Fig. 5b), represent ~80% of all the structures that result from the selection process.

## References

- Chiu, Y. L. & Greene, W. C. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu. Rev. Immunol.* **26**, 317–353 (2008).
- Malim, M. H. APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 675–687 (2009).
- Duggal, N. K. & Emerman, M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat. Rev. Immunol.* **12**, 687–695 (2012).
- Jarmuz, A. *et al.* An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285–296 (2002).
- Betts, L., Xiang, S., Short, S. A., Wolfenden, R. & Carter, Jr C. W. Cytidine deaminase. The 2.3 Å crystal structure of an enzyme: transition-state analog complex. *J. Mol. Biol.* **235**, 635–656 (1994).
- Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J. & Harris, R. S. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat. Struct. Mol. Biol.* **17**, 222–229 (2010).
- Shinohara, M. *et al.* APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells. *Sci. Rep.* **2**, 806 (2012).
- Vartanian, J. P., Guetard, D., Henry, M. & Wain-Hobson, S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* **320**, 230–233 (2008).
- Wiegand, H. L. & Cullen, B. R. Inhibition of alpharetrovirus replication by a range of human APOBEC3 proteins. *J. Virol.* **81**, 13694–13699 (2007).
- Chen, H. *et al.* APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.* **16**, 480–485 (2006).
- Narvaiza, I. *et al.* Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase. *PLoS Pathog.* **5**, e1000439 (2009).
- Ooms, M., Krikoni, A., Kress, A. K., Simon, V. & Münk, C. APOBEC3A, APOBEC3B, and APOBEC3H haplotype 2 restrict human T-lymphotropic virus type 1. *J. Virol.* **86**, 6097–6108 (2012).
- Bogerd, H. P. *et al.* Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl Acad. Sci. USA* **103**, 8780–8785 (2006).
- Muckenfuss, H. *et al.* APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J. Biol. Chem.* **281**, 22161–22172 (2006).
- Kinomoto, M. *et al.* All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res.* **35**, 2955–2964 (2007).
- Bulliard, Y. *et al.* Structure-function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities. *J. Virol.* **85**, 1765–1776 (2011).
- Iwatani, Y., Takeuchi, H., Strebel, K. & Levin, J. G. Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. *J. Virol.* **80**, 5992–6002 (2006).
- Carpenter, M. A. *et al.* Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *J. Biol. Chem.* **287**, 34801–34808 (2012).
- Wijesinghe, P. & Bhagwat, A. S. Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res.* **40**, 9206–9217 (2012).
- Landry, S., Narvaiza, I., Linfesty, D. C. & Weitzman, M. D. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep.* **12**, 444–450 (2011).
- 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. USA* **108**, 4858–4863 (2011).
- Aynaud, M. M. *et al.* Human Tribbles 3 protects nuclear DNA from cytidine deamination by APOBEC3A. *J. Biol. Chem.* **287**, 39182–39192 (2012).
- Peng, G. *et al.* Myeloid differentiation and susceptibility to HIV-1 are linked to APOBEC3 expression. *Blood* **110**, 393–400 (2007).
- Koning, F. A. *et al.* Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J. Virol.* **83**, 9474–9485 (2009).
- Refsland, E. W. *et al.* Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res.* **38**, 4274–4284 (2010).
- Thielen, B. K. *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.* **285**, 27753–27766 (2010).
- Berger, G. *et al.* APOBEC3A is a specific inhibitor of the early phases of HIV-1 infection in myeloid cells. *PLoS Pathog.* **7**, e1002221 (2011).
- Koning, F. A., Goujon, C., Bauby, H. & Malim, M. H. Target cell-mediated editing of HIV-1 cDNA by APOBEC3 proteins in human macrophages. *J. Virol.* **85**, 13448–13452 (2011).
- Navarro, F. *et al.* Complementary function of the two catalytic domains of APOBEC3G. *Virology* **333**, 374–386 (2005).
- Newman, E. N. *et al.* Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr. Biol.* **15**, 166–170 (2005).
- Chen, K.-M. *et al.* Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116–119 (2008).
- Furukawa, A. *et al.* Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO J.* **28**, 440–451 (2009).
- Harjes, E. *et al.* An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *J. Mol. Biol.* **389**, 819–832 (2009).
- Holden, L. G. *et al.* Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **456**, 121–124 (2008).
- Shandilya, S. M. D. *et al.* Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure* **18**, 28–38 (2010).
- Kitamura, S. *et al.* The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat. Struct. Mol. Biol.* **19**, 1005–1010 (2012).
- Krzysiak, T. C., Jung, J., Thompson, J., Baker, D. & Gronenborn, A. M. APOBEC2 is a monomer in solution: implications for APOBEC3G models. *Biochemistry* **51**, 2008–2017 (2012).

38. Bransteitter, R., Prochnow, C. & Chen, X. S. The current structural and functional understanding of APOBEC deaminases. *Cell. Mol. Life Sci.* **66**, 3137–3147 (2009).
39. Love, R. P., Xu, H. & Chelico, L. Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A. *J. Biol. Chem.* **287**, 30812–30822 (2012).
40. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52**, 609–623 (2003).
41. Aguiar, R. S., Lovsin, N., Tanuri, A. & Peterlin, B. M. Vpr.A3A chimera inhibits HIV replication. *J. Biol. Chem.* **283**, 2518–2525 (2008).
42. Chelico, L., Pham, P., Calabrese, P. & Goodman, M. F. APOBEC3G DNA deaminase acts processively 3' → 5' on single-stranded DNA. *Nat. Struct. Mol. Biol.* **13**, 392–399 (2006).
43. Nowarski, R., Britan-Rosich, E., Shiloach, T. & Kotler, M. Hypermutation by intersegmental transfer of APOBEC3G cytidine deaminase. *Nat. Struct. Mol. Biol.* **15**, 1059–1066 (2008).
44. Iwatani, Y. *et al.* Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic Acids Res.* **35**, 7096–7108 (2007).
45. Losey, H. C., Ruthenburg, A. J. & Verdine, G. L. Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nat. Struct. Mol. Biol.* **13**, 153–159 (2006).
46. Senavirathne, G. *et al.* Single-stranded DNA scanning and deamination by APOBEC3G cytidine deaminase at single molecule resolution. *J. Biol. Chem.* **287**, 15826–15835 (2012).
47. Roberts, S. A. *et al.* Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* **46**, 424–435 (2012).
48. Burns, M. B. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366–370 (2013).
49. Clore, G. M. & Gronenborn, A. M. Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol.* **16**, 22–34 (1998).
50. Sattler, M., Maurer, M., Schleucher, J. & Griesinger, C. A Simultaneous <sup>15</sup>N, <sup>1</sup>H-HSQC and <sup>13</sup>C, <sup>1</sup>H-HSQC with sensitivity enhancement and a heteronuclear gradient-echo. *J. Biomol. NMR* **5**, 97–102 (1995).
51. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
52. Johnson, B. A. & Blevins, R. A. NMR View - a computer-program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
53. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta. Crystallogr. D Biol. Crystallogr.* **54**(Pt 5): 905–921 (1998).
54. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
55. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
56. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 29–32 (1996).
57. Flores, S. C., Sherman, M. A., Bruns, C. M., Eastman, P. & Altman, R. B. Fast flexible modeling of RNA structure using internal coordinates. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **8**, 1247–1257 (2011).
58. Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V. & Mee, R. P. 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.* **11**, 425–445 (1997).

## Acknowledgements

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. PEOF-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.).

## 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.

## Additional information

**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.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors claim no competing financial interests.

**Reprints and permission** information is available online at <http://npublishing.nature.com/reprintsandpermissions/>

**How to cite this article:** Byeon, I.-J. L. *et al.* NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nat. Commun.* **4**:1890 doi: 10.1038/ncomms2883 (2013).