Published in final edited form as:

Nature. 2008 November 6; 456(7218): 121–124. doi:10.1038/nature07357.

Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications

Lauren G. Holden^{1,*}, Courtney Prochnow^{1,*}, Y. Paul Chang^{1,*}, Ronda Bransteitter¹, Linda Chelico¹, Udayaditya Sen¹, Raymond C. Stevens², Myron F. Goodman¹, and Xiaojiang S. Chen¹

¹ Molecular and Computational Biology, University of Southern California, Los Angeles, California 90089, USA

² Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

Abstract

The APOBEC family members are involved in diverse biological functions. APOBEC3G restricts the replication of human immunodeficiency virus (HIV), hepatitis B virus and retroelements by cytidine deamination on single-stranded DNA or by RNA binding^{1–4}. Here we report the high-resolution crystal structure of the carboxy-terminal deaminase domain of APOBEC3G (APOBEC3G-CD2) purified from *Escherichia coli*. The APOBEC3G-CD2 structure has a five-stranded β -sheet core that is common to all known deaminase structures and closely resembles the structure of another APOBEC protein, APOBEC2 (ref. ⁵). A comparison of APOBEC3G-CD2 with other deaminase structures shows a structural conservation of the active-site loops that are directly involved in substrate binding. In the X-ray structure, these APOBEC3G active-site loops form a continuous ‘substrate groove’ around the active centre. The orientation of this putative substrate groove differs markedly (by 90 degrees) from the groove predicted by the NMR structure⁶. We have introduced mutations around the groove, and have identified residues involved in substrate specificity, single-stranded DNA binding and deaminase activity. These results provide a basis for understanding the underlying mechanisms of substrate specificity for the APOBEC family.

We have purified the human wild-type C-terminal cytidine deaminase domain of APOBEC3G (APOBEC3G-CD2, residues 197–380) expressed in *E. coli*. APOBEC3G-CD2 (with and without a glutathione *S*-transferase (GST) tag) is highly soluble, and deaminates cytidine to uridine on single-stranded DNA (ssDNA) with a specific activity of 5 fmol $\mu\text{g}^{-1} \text{min}^{-1}$, which is about 25-fold lower than that of the full-length APOBEC3G (GST–APOBEC3G; 126 fmol $\mu\text{g}^{-1} \text{min}^{-1}$) expressed in *E. coli* (Fig. 1a). We analysed the processive and polar properties of APOBEC3G-CD2 and full-length APOBEC3G (Fig. 1b). Similar to the insect-cell-derived full-length APOBEC3G^{7,8}, the full-length APOBEC3G expressed in *E. coli* processively deaminates cytidine in two 5'CCC3' motifs located on a ssDNA substrate, during one binding event (Fig. 1b). The full-length APOBEC3G also exerts a 3' to 5' deamination bias by preferentially deaminating the cytidine in the CCC motif near the 5' end of the ssDNA substrate (Fig. 1b). In contrast, the APOBEC3G-CD2 exhibits an approximate twofold decrease in

*These authors contributed to the work equally.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Author Information The atomic coordinates and the structure of APOBEC3G-CD2 have been deposited in the Protein Data Bank under accession number 3E1U. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to X.S.C. (xiaojiang.chen@usc.edu).

processivity and virtually no 3' to 5' deamination bias (Fig. 1b). These results indicate that APOBEC3G-CD2 partially retains the catalytic properties of full-length APOBEC3G, but that the CD1 domain in the context of the full-length APOBEC3G is probably required for the strong processive property and the 3' to 5' deamination bias on ssDNA.

We crystallized the wild-type APOBEC3G-CD2 and solved the structure by using the multi-wavelength anomalous dispersion (MAD) phasing method with selenium-substituted methionine (Se-Met) diffraction data. The 2.3 Å resolution X-ray structure shows a core β -sheet that is composed of five β -strands surrounded by six α -helices (Fig. 1c, d). Helices 2–4 (h2–h4) are packed alongside one face of the core β -sheet (Fig. 1c), whereas helix 1 (h1) and helix 5 (h5) are packed against the opposite face of the β -sheet (Fig. 1c, d). Helix 6 (h6) is located at the edge of the β -sheet core and is perpendicular to the β 5 strand (Fig. 1c).

A recently reported NMR structure of an APOBEC3G-CD2 mutant (APOBEC3G-2K3A, Protein Data Bank (PDB) accession 2JYW)⁶ resembles the X-ray structure of the wild-type APOBEC3G-CD2. However, the superposition of the two structures shows notable differences (Supplementary Fig. 1a). The β 2 strand and the amino-terminal helix (h1) are absent from the NMR structure (Supplementary Fig. 1b, c, in grey). The β 2 strand in the X-ray structure does not make crystal contact with neighbouring monomers. Thus, the formation of the β 2 strand is unlikely to be the result of crystal contact. Furthermore, a similar β 2 strand within a five-stranded β -sheet core is the common structural feature that is observed in all wild-type cytidine deaminase structures available so far (Supplementary Fig. 2a, b, d, e). Therefore, an intact full-length β 2 strand and the five-stranded β -sheet core is probably the feature of wild-type APOBEC3G-CD2 and all other APOBEC proteins. The structural differences observed in the NMR structure could have resulted from the five mutations on the APOBEC3G protein used for NMR study (Supplementary Fig. 1b, c), or from the different methodology used for determining the structure, or from both.

A superposition of the core structures of APOBEC3G-CD2 and APOBEC2 monomers shows substantial overlap for all five β -strands and all six helices (Fig. 2a), suggesting that these core structures of APOBEC family members are highly conserved. Yet, the structural overlap shows notable differences in the active centre (AC) loops, referred to as AC loops 1 and 3, which potentially mark the differences in substrate use and activity of the two proteins (Fig. 2b, c). The AC loop 1, which connects h1 with the β 1 strand, is located further away from the active site in APOBEC3G than in APOBEC2 (Fig. 2b, c). The APOBEC3G AC loop 3 is longer than that of APOBEC2 and is positioned further away from the active site (Fig. 2b, c).

The structure shows elaborate bonding interactions that can stabilize the open conformation of APOBEC3G AC loops 1 and 3. For example, AC loop 1 forms an extensive bonding network through R215, which anchors this loop to other parts of the structure. R215 interactions include the direct contact with R313 and W285 located near the core structure (Fig. 2d). We demonstrate later that the R215E mutation in APOBEC3G abolishes deamination activity, which is consistent with a previous study⁹. Similarly, the APOBEC3G AC loop 3 is stabilized by multiple hydrogen bonds between the main-chain atoms of residues R256, F252, L253, H248 and Q245 within the loop (Fig. 2e). The loop residue R256 interacts with D264 on a core helix by a strong salt bridge, and R256 hydrophobically packs with the loop residue F252 by the long aliphatic chain (Fig. 2e). All of these interactions should help stabilize the conformation of AC loop 3. An APOBEC3G R256E mutant, which probably disrupts the AC loop 3 conformation, greatly impairs deamination activity.

In the active site of APOBEC3G-CD2, a zinc atom is coordinated by the three residues H257, C288 and C291, and a water molecule (Fig. 3a). The closely positioned water molecule can be activated to become a Zn-hydroxide for nucleophilic attack in the deamination reaction¹⁰. Two

residues (N244 and H257) on the APOBEC3G AC loop 3 show a structural conservation with many distantly related Zn-deaminases, specifically TadA and human CDA^{10,11} (Fig. 3b, c). The two equivalent TadA residues (N42 and H53) on a TadA loop (similar to the AC loop 3) directly contact the target base of the RNA substrate (Fig. 3b). These residues overlap well with the APOBEC3G residues N244 and H257 on the AC loop 3 in the superposition of the two structures (Fig. 3b)¹¹. Similarly, two equivalent residues (N54 and C65) on a human CDA loop contact the substrate/inhibitor¹⁰, and also overlap with N244 and H257 on the AC loop 3 of APOBEC3G (Fig. 3c). This structural conservation suggests that the APOBEC3G-CD2 residues, N244 and H257, are also involved in substrate contact. In an *in vitro* assay, the APOBEC3G N244A mutant had no detectable deamination activity (Fig. 3f). The structural conservation of the position of these residues suggests that the open conformation of the APOBEC3G AC loop 3 is in a position ready to bind nucleic acid.

A surface representation of the APOBEC3G-CD2 X-ray structure shows that the AC loops 1 and 3 and the regions near the active site form a deep, spacious groove that runs horizontally across the active centre pocket (Fig. 3d). This groove is not present in the APOBEC3G-2K3A NMR structure because of the structural differences⁶ (Supplementary Fig. 3d–f). The structural features in this groove strongly suggest a role for binding ssDNA substrates. The groove starts between the AC loops 1 and 3 on the right side of the displayed structure (Fig. 3d), leads into a deep pocket where the Zn atom is located and slightly exposed, and continues towards the left side over helix 6. The target base must be positioned into the active site so that the attachment of the Zn hydroxyl group can occur on the cytidine base during the deamination reaction (Fig. 3c). The ssDNA lying across this horizontal groove can present a cytidine base so that it is directed towards the active site Zn in the correct orientation and angle to permit deamination, as shown in the case of TadA and human CDA (Fig. 3a–c)^{10,11}.

Within this horizontal groove are a group of charged residues (R213, R215, N244, R256, R313, D316, D317, R320, R374 and R376) and hydrophobic residues (W285, Y315 and F289; Fig. 3e). In our mutagenesis study, we show that all of these residues are important for the deamination activity on ssDNA (Fig. 3f). However, they affect the deamination activity in different ways. The R374 and R376 residues are located on one end of the groove and are positioned to interact with a negatively charged ssDNA phosphate backbone. The ssDNA binding of the R374E/R376D double mutant is impaired by 46% in comparison to that of the wild-type APOBEC3G, and the deamination activity is even more disrupted (Fig. 3f). On the edge of the groove, the AC loop 1 R213 residue can make contact with ssDNA. Consistent with a previous report⁶, the R213E mutant has only weak deamination activity (Fig. 3f).

Three of the charged residues (R256, R215 and R313) are involved in elaborate bonding networks for the AC loops (Figs 2d, e and 3c), and should be important for maintaining the groove conformation. The mutants R215E, R256E and R313E/R320D show only minimal or no deamination activity (Fig. 3f). The primary functional role of these residues may be to maintain the conformation of the substrate groove rather than to directly contact ssDNA. Mutation of the R313 residue can disrupt its interaction with W285, which is located on the floor of the groove near the active site Zn (Fig. 3e). Y315 next to W285 is also on the floor of the groove. Both residues could stack with bases of ssDNA and position the DNA into the active site (Fig. 3d, e). Mutants W285A and Y315A show no detectable deamination activity (Fig. 3f), consistent with a previous report⁹. Another hydrophobic residue on the edge of the groove is F289, and the F289A mutant has greatly reduced deaminase activity (Fig. 3f).

Notably, next to Y315 and W285 are two negatively charged residues (D316 and D317) on the floor of the groove (Fig. 3d, e). The mutant D316R/D317R has higher deamination activity (1.6-fold), as well as higher ssDNA binding (twofold) compared to the wild-type APOBEC3G (Fig. 3f). These enhanced activities could be caused by the increased total positive charge in

the groove. Furthermore, this mutant showed altered substrate specificity (Fig. 3f, inset). Unlike wild-type APOBEC3G that strongly favours deamination at the 3'C of a 5'CCC3'hot-spot motif, the D316R/D317R mutant deaminates the middle C and the 3'C at about the same rate (Fig. 3f, inset). This result indicates that these negative residues, D316 and D317, are important for positioning the substrate so that the 3'C is most likely to be deaminated by wild-type APOBEC3G.

The mutagenesis study supports our model of the horizontal groove, and verifies that the residues located within and around the groove are important for deamination activity, ssDNA binding and substrate orientation. These results provide a basis to pursue further studies of APOBEC3G and other important APOBEC proteins (including activation-induced cytidine deaminase, AID), which will facilitate our understanding of how they act within our innate and adaptive immune responses to restrict HIV and other infectious pathogens.

METHODS SUMMARY

APOBEC3G-CD2 was expressed and purified as a recombinant GST-fusion protein in *E. coli*. Purified GST-fusion protein was digested by PreScission protease. Further purification of the APOBEC3G-CD2 protein was completed using Superdex-75 gel filtration chromatography in 50 mM HEPES, pH 7.0, 250 mM NaCl and 1 mM dithiothreitol. Native and Se-Met-labelled proteins were concentrated to 25 mg ml⁻¹. Crystals were grown at 18 °C by hanging-drop vapour diffusion from a reservoir solution of 100 mM MES, pH 6.5, 40% PEG 200. In an assay for deamination activity, APOBEC3G (0.024–10 µM) was allowed to react with 500 nM fluorescein-dT-incorporated ssDNA for 10 or 15 min and subsequently treated with uracil-DNA glycosylase and resolved on 16% urea-PAGE for analysis as described previously⁷. Specific activity, measured as fmoles of substrate deaminated per µg of enzyme per minute, was calculated from the per cent deamination of a ssDNA substrate over a range of enzyme concentrations. To analyse processivity and directionality, substrate use (%) was less than 15% to maintain single-hit kinetics. The 'processivity factor' is defined as the ratio of the observed fraction of double deaminations (occurring at both 5'C and 3'C on the same molecule) to the predicted fraction of independent double deaminations⁷. A processivity factor of greater than one indicates that most of the double deaminations are caused by the same APOBEC3G molecule acting processively on both C targets. The deamination bias is measured by the ratio of 5'C/3'C deaminations⁷. For the experiments measuring processivity and directionality, the ssDNA substrate sequence are found in the Methods. Crystallography statistics are found in Supplementary Information.

METHODS

Structure determination and refinement

Selenium-substituted methionine protein crystals were used for collecting Se-MAD data using the ALS synchrotron beam source. Data were processed with HKL3000 (ref. ¹²). A total of three selenium and one zinc sites were located by the SHELXD¹³ program using MAD data between 50 and 3.0 Å resolution range. The SHARP program was used to calculate the experimental and model-combined phases using the MAD data in the resolution range of 50 to 2.3 Å as well as for density modification. The model was built with O using the high quality electron density map obtained, and was refined with CNS to 2.3 Å resolution with excellent statistics. The final refinement statistics and geometry as defined by Procheck were in good agreement and are summarized in Supplementary Table 1. Structure figures were designed using PyMOL¹⁴.

Construction of APOBEC3G mutants

Mutant APOBEC3G proteins (D316R/D317R, R313E/R320D and R374E/R376D) were constructed by site-directed mutagenesis using the pAcG2T-APOBEC3G vector as the template. The following primers and their complementary strands were used: 5'-CTTCACTGCCC-GCATCTATAGAAGACAAGGAAGATGTCAGGAG-3'(D316R/D317R), 5'-CTGTGCATCTTCACTGCCGAGATCTATGATGATCAAGGAGATTGTCAG-GAGGGGCTGCGC-3'(R313E/R320D), and 5'-GAGCACAGCCAAGACCTG-AGTGGGGAGCTGGACGCCATTCTCCAGAATCAGG-3'(R374E/R376D). The entire coding region of the APOBEC3G mutant constructs was verified by DNA sequencing. The mutant plasmids were then co-transfected, according to the manufacturer's protocol, with linearized baculovirus DNA (BD Biosciences) to generate recombinant mutant APOBEC3G baculovirus. Wild-type and mutant APOBEC3G expression in Sf9 insect cells and purification was carried out as described previously⁸. Mutant *E. coli* GST-APOBEC3G proteins (R213E, R215E, K249E, R256E, W285A, F289A and Y315A) were constructed by site-directed mutagenesis using the pGEX-6P1-GST-APOBEC3G vector as the template. The following primers and their complementary strands were used: 5'-AATGAACCTTGGGTTGAAGGTCGTCACGAGACTTAC-3'(R213E), 5'-GAA-CCTTGGGTTTCGTGGTGAACACGAGACTTACCTG-3'(R215E), 5'-TGTAAC-CAGGCCCCGACGAGCACGGTTTCTGGAA-3'(K249E), 5'-GCACGGTT-TTCTGGAAGGTGAACACGCCGAACCTGTG-3'(R256E), 5'-GTTACCTGCTT-TACCTCTGCGTCCCCGTGCTTTTCC-3'(W285A), 5'-ACCTCTTGGTCCCC-CGTGCGCTTCCTGCGCACAAGAA-3'(F289A), 5'-ATCTTCACTGCACGTA-TTGCCGACGACCAGGGCCGT-3'(Y315A), 5'-CGTCGTGGTTTCTGTGTG-TGCCAGGCCCCGACAAGCAC-3'(N244A), 5'-CGTCGTGGTTTCTGTG-TAGACAGGCCCCGACAAGCAC-3'(N244A). The entire coding region of the APOBEC3G mutant constructs was verified by DNA sequencing. Plasmids were expressed in XA90 *E. coli* cells and were lysed by French press. Further purification was carried out as described previously⁸.

DNA binding

APOBEC3G DNA binding was monitored by changes in steady state fluorescence depolarization (rotational anisotropy). Reaction mixtures (70 µl), containing fluorescein-labelled DNA (50 nM) in buffer (50 mM HEPES, pH 7.3, 1 mM dithiothreitol and 5 mM MgCl₂) and varying concentration of 0 to 500 nM APOBEC3G, were incubated at 37 °C. The sequence of the ssDNA was TTAGATGAGTGTA(fluorescein-dT)GTGATATATGTGTAT. Rotational anisotropy was measured as described previously⁷. The fraction of DNA bound to protein was determined as described previously¹⁵.

Processivity and directionality substrates

The substrate used to determine processivity and directionality is 5'-AAAGAGAAAGTGATACCCAAAGAGT-AAAGT (fluorescein-dT)AGATAGAGAGTGATACCCAAAGAGTAAAGTTA-GTAAGATGTGTAAGTATGTAA-3'. For specific activity measurements, the ssDNA substrate sequence was GG(fluorescein-dT)AGTTTAGTGTTTGTAT-AGAATTAATACCCAAAGAAGTGATGTAATTGTTATGATAAGATTGAAA.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the staff at the Berkeley Laboratory's Advanced Light Source (ALS) BL8.2.1 and Advanced Photon Source 19ID in Argonne National Laboratory for assistance in data collection, and M. Klein and other members of the X.S.C. laboratory for help and discussion. This work is supported in part by CBM graduate training grants to L.H. and Y.P.C., and by National Institutes of Health grants to M.F.G. and X.S.C.

References

1. Peled JU, et al. The biochemistry of somatic hypermutation. *Annu Rev Immunol* 2008;26:481–511. [PubMed: 18304001]
2. Goodman MF, Scharff MD, Romesberg FE. AID-initiated purposeful mutations in immunoglobulin genes. *Adv Immunol* 2007;94:127–155. [PubMed: 17560274]
3. Conticello SG, Langlois MA, Yang Z, Neuberger MS. DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv Immunol* 2007;94:37–73. [PubMed: 17560271]
4. 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]
5. Prochnow C, Bransteitter R, Klein MG, Goodman MF, Chen XS. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature* 2007;445:447–451. [PubMed: 17187054]
6. Chen KM, et al. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* 2008;452:116–119. [PubMed: 18288108]
7. Chelico L, Pham P, Calabrese P, Goodman MF. APOBEC3G DNA deaminase acts processively 3' → 5' on single-stranded DNA. *Nature Struct Mol Biol* 2006;13:392–399. [PubMed: 16622407]
8. Chelico L, Sacho EJ, Erie DA, Goodman MF. A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV. *J Biol Chem* 2008;283:13780–13791. [PubMed: 18362149]
9. Chen KM, et al. Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G. *FEBS Lett* 2007;581:4761–4766. [PubMed: 17869248]
10. Chung SJ, Fromme JC, Verdine GL. Structure of human cytidine deaminase bound to a potent inhibitor. *J Med Chem* 2005;48:658–660. [PubMed: 15689149]
11. Losey HC, Ruthenburg AJ, Verdine GL. Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nature Struct Mol Biol* 2006;13:153–159. [PubMed: 16415880]
12. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997;276:307–326.
13. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. *Acta Crystallogr D* 2002;58:1772–1779. [PubMed: 12351820]
14. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; 2002.
15. Bertram JG, Bloom LB, O'Donnell M, Goodman MF. Increased dNTP binding affinity reveals a nonprocessive role for *Escherichia coli* β clamp with DNA polymerase IV. *J Biol Chem* 2004;279:33047–33050. [PubMed: 15210708]

Figure 1. The X-ray structure of enzymatically active APOBEC3G-CD2

a, Analysis of the deamination activity for full-length GST-APOBEC3G (GST-A3G) and GST-APOBEC3G-CD2 (GST-CD2) and APOBEC3G-CD2 (CD2). The 32-nucleotide (nt) band indicates deamination activity. **b**, APOBEC3G processivity and the 3'to 5'deamination bias was characterized on ssDNA with two CCC motifs. Single deaminations of the 5'C and the 3'C appear as 67- and 48-nucleotide fragments, respectively; deamination of both the 5'C and the 3'C results in a 30-nucleotide fragment (see Methods) **c**, **d**, Two views of the APOBEC3G-CD2 domain rotated 90° showing the five-stranded β -sheet core surrounded by six helices. The zinc is represented as a red sphere.

Figure 2. Structural comparison of APOBEC3G-CD2 with APOBEC2

a, Core structures of APOBEC3G-CD2 (yellow) and APOBEC2 (cyan) superimposed. The red sphere represents zinc. **b**, **c**, The superposition of APOBEC3G-CD2 and an APOBEC2 monomer, with the AC loop 1 collapsed over the active site (conformation 1, **b**) or forming an α -hairpin (conformation 2, **c**). **d**, In APOBEC3G-CD2, the AC loop 1 R215 residue forms hydrogen bonds (green dashes) with F204, E211, N207, E209 and W285 (pink). The R215 aliphatic chain hydrophobically packs with F204, R313 and W285. **e**, The APOBEC3G-CD2 AC loop 3 residues, R256, F252, G251, H248 and G244 (pink), form main-chain hydrogen bonds (green dashes). The conserved N244 is shown in cyan. The active site residues are H257, C288 and C291 (wheat).

Figure 3. Predicted substrate groove and deamination activity of APOBEC3G mutants

a, The active site residues of APOBEC3G-CD2. The water (W) and zinc molecules are cyan and red spheres, respectively. **b**, Superposition of APOBEC3G-CD2 (A3G-CD2, yellow) and TadA (light blue, PDB accession 2B3J). **c**, Superposition of APOBEC3G-CD2 (yellow) and human CDA (pink). **d**, Surface representation of APOBEC3G-CD2, showing a horizontal groove with residues (magenta) predicted to interact with ssDNA. ssDNA is represented by a green line. **f**, Mutational data of APOBEC3G purified from Sf9 (left) or from *E. coli* (right) are shown. The right inset shows the relative deamination of the 3'C (5'CCC) or the middle C (5'CCC) on a ssDNA substrate. Error bars represent the s.d.