Published in final edited form as:

DNA Repair (Amst). 2017 June ; 54: 8–12. doi:10.1016/j.dnarep.2017.03.007.

Activation-induced deoxycytidine deaminase: structural basis favoring WRC hot motif specificities unique among APOBEC family members

Phuong Pham¹, Samir A. Afif¹, Mayuko Shimoda^{2,3,4}, Kazuhiko Maeda^{3,4}, Nobuo Sakaguchi^{4,5}, Lars C. Pedersen⁶, and Myron F. Goodman^{1,7,*}

¹Department of Biological Sciences, University of Southern California, Los Angeles CA 90089

²Department of Immunology, Graduate School of Medical Sciences, Kumamoto University, 1-1-1 Honjo, Chuo-ku, Kumamoto 860-8556, Japan.

³Laboratory of Host Defence, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamada-oka, Suita 565-0871, Japan

⁴World Premier International Research Center Initiative, Immunology Frontier Research Center, Osaka University, 3-1 Yamada-oka, Suita 565-0871, Japan

⁵Tokyo Metropolitan Institute of Medical Science, 2-1-6, Kamikitazawa, Setagaya-ku, Tokyo, 156-8506, Japan

⁶Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709

⁷Department of Chemistry, University of Southern California, Los Angeles CA 90089

Keywords

AID X-ray crystal structure; Antibody diversity; IgV somatic hypermutation; scanning processivity

Activation-induced deoxycytidine deaminase (AID) is a member of the APOBEC-family DNA- and RNA-dependent deaminases. The generation of high-affinity antibodies requires that AID initiate somatic hypermutation (SHM) in immunoglobulin variable regions (IgV) and class-switch recombination in Ig switch regions (IgS). AID exhibits distinct motif specificities on single-stranded ssDNA, favoring deaminations in WRC (W = A/T, R = A/G) motifs while disfavoring SYC (S = G/C, Y = T/C) motifs [1, 2]. In contrast, three widely studied APOBEC dC deaminases, Apo3A, Apo3B, and Apo3G, exhibit strong YCC deamination signatures on ssDNA [3–5]. Recently, we obtained a high-resolution X-ray crystal structure for a variant of AID, AIDv(15) at 2.8 Å [6]. Native AID (AID) is poorly soluble and difficult to purify. We succeeded in obtaining highly soluble (> 15 mg/ml) chromatographically pure protein by introducing 12 mutations and three amino acid deletions at the N-terminus in helix 1 and loop 1 and three surface mutations at the end of

*To whom correspondence should be addressed: mgoodman@usc.edu.

strand $\beta 1$, which mirror residues found in Apo3A and reduce the charge of AID from +11.9 to +3.4 at neutral pH [6]. To enhance crystallization, a 15 residue C-terminal truncation of AIDv, AIDv(15), was also generated [6]. The C-terminal residues are known to be required for CSR, but they are not required for SHM [7, 8], and do not affect AID's preference for WRC motifs [9].

A comparison of the structure for AID with those of Apo3A and Apo3G, showed that the APOBEC substrate specificity loop (loop 7) for AID is larger and extends further away from the catalytic active cleft suggesting a structural basis of AID's ability to accommodate two neighboring purine residues [6]. Since AIDv contains an N-terminus similar to Apo3A, key questions that needed to be addressed were: 1) does AIDv(15) exhibit motif specificities consistent with AID *in vitro*? 2) Does the variant exhibit high deamination processivity *in vitro*, which is a hallmark property of AID [1, 10, 11]? 3) Is AIDv(15) able to carry out somatic hypermutation (SHM) of an IgV gene *in vivo*? T. Honjo and co-workers had earlier shown that maintaining the integrity of loop 1 in AID was necessary to observe class-switch recombination (CSR) and SHM in primary B cells [12]. We realized, therefore, that responding to "question 3" was a long shot at best, but still worth trying.

We are writing this commentary to discuss two recent publications. First, in a letter to DNA Repair, Professor Honjo and colleagues, showed that our AIDv construct was unable to generate SHM or CSR in mouse primary B cells or in a mouse B lymphoma cell line [13], consistent with his previous data with point mutants in the AID N-terminus [12]. Second, an Apo3A-ssDNA co-crystal structure is now available [14] that offers additional perspective on the role of loop 7 in AID vs. Apo3A in determining motif specificity.

AIDv(15) and AID exhibit similar spatial deamination spectra *in vitro*

Using an *in vitro* assay, deamination spectra for AID, AIDv(15), and Apo3A were obtained in a *lacZa* target sequence, and scored as C→T mutations, as described in Ref. [6]. Spectra for AID and AIDv(15) contain numerous deaminations in hot WRC motifs sharing most of the same locations (Fig. 1, "red" lines). AID and AIDv(15) exhibit a similar dense spatial distribution of mutation in all motifs, hot, cold and neutral (non-WRC or -SYC, "green" lines). Apo3A deaminations exhibit a diametrically opposite preference, with dC deaminations occurring predominantly in SYC motifs, with just a small fraction of clones having mutations in WRC (Fig. 1). Unlike the mutational pattern for the two AID enzymes, the spatial mutational pattern for Apo3A is relatively sparse.

A single molecule of AID remains bound to a ssDNA substrate for a median time ~ 4 to 5 min [15], during which it scans randomly in bidirectional short slides/hops [10, 11, 15]. During this lengthy binding period, large regions of DNA are scanned with numerous deaminations occurring processively. AIDv(15) retains the ability to catalyze multiple processive deaminations [6]. In contrast, Apo3A shows weak processivity, behaving in a much more distributive manner [4, 16].

AIDv(15) and AID exhibit similar spatial IgV SHM patterns in Ramos B-cells *in vivo*

AIDv(15) overexpressed in Ramos B-cells promotes IgV_H SHM. The mutation frequency is similar for the overexpressed variant compared to overexpressed native AID, both of which are 3-fold higher than control cells with AID expressed at normal levels (Fig. 2A). The spatial distribution of base substitution mutations in IgV_H is roughly similar for AIDv(15) and AID, with about 20% (14 out of 68 mutated sites) occurring at identical locations (Fig. 2B). Two mutational hot spots occur in a complementarity-determining region (CDR) at adjacent sites in the canonical AGCT hot motif for AIDv(15) (positions 84, 85), whereas a single hot spot for AID occurs at AGCT (position 85). An identical hot spot occurs for both enzymes at AGCT (position 213), while AID generates a unique AGCT hot spot at position 230. Closely spaced mutational hot spots are observed for AIDv(15) (positions 181, 284) and AID (positions 183, 286). Several indels are also observed in IgV_H, which occur only with the AID variant (Fig. 2B). These include 2 deletions and 4 exact sequence duplications. The deletions occur at the locations of the two closely spaced mutational hot spots. The ability of AIDv(15) to generate indels suggests the possibility that closely spaced deaminations may be occurring during transcription on both non-transcribed and transcribed strands, perhaps within a stalled transcription bubble, which could serve as foci for double-strand breaks and subsequent recombination.

AIDv(15) and our full-length variant AIDv are “silent” in primary mouse B cells and a mouse B lymphoma cell line, as reported in a recent letter to DNA Repair by Professor Honjo and colleagues [13]. However, human Ramos B cells are apparently far more “forgiving” than the primary mouse B cells and B lymphoma cell line, since SHM occurs at an increased rate when AIDv(15) is overexpressed in Ramos (Fig. 2A). The Ramos B-cell data along with biochemical data documenting similar motif deamination spectra suggest that our crystal structure for AIDv(15) can serve as a surrogate for native AID.

Structural Insights into AID motif specificity in relation to an Apo3A-ssDNA co-crystal

A comparison of the crystal structure of AIDv(15) with those of Apo3A and the catalytic subunit of Apo3G-CD2 suggested a structural basis for the distinctive WRC dC deamination bias exhibited by AID compared to YCC for other APOBEC family members [6]. The explanation lies in differences in loop 7. Loop 7 of AID consists of unique sequences from residues Leu113 to Pro123 that have been shown to dictate the substrate specificity of individual APOBEC proteins [17–19]. It has previously been demonstrated that swapping loops between APOBEC proteins switches the specificity of these enzymes [17–19].

An overlay of AIDv(15) [6] with the newly obtained Apo3A/DNA complex [14] (Figs. 3A, B) shows a larger loop 7 for AID, which extends away from the active site creating a larger cavity at the –1 position that could potentially accommodate a larger purine base without stringently excluding pyrimidines. AIDv(15) lacks the equivalent D131 and Y132 in

Apo3A that are important for the selectivity for T by Apo3A through hydrogen bonding and hydrophobic interactions, respectively (Fig. 3B) [14].

Notably, AIDv(15) exhibits a two-fold reduction in processivity when compared to AID (AID's processivity is ~20 deaminations/clone measured under the same single-hit and residence time conditions) [1, 6]. One explanation for the reduction in processivity may result from an augmentation in the 5' binding pocket for the DNA. Loop1 is located on the backside of the -1 and -2 bases in the Apo3A complex structure. The larger loop 1 for native AID may inhibit the DNA "U" conformation present in the Apo3A structure resulting in a different path of the 5' end of the DNA in AID.

Previously, we've suggested that the 5' end may extend along a positively charged cleft across helix 6 [6] (Figs. 3A, 3C). This was supported by observations that amino acid substitutions in helix 6 (R171A, Q175A, R178A) each reduced AIDv(15) processivity by an additional 2-fold [6]. Since these amino acids are remote from the active site (Fig. 3C), their influence in reducing processivity may reflect direct non-specific interactions with ssDNA outside the WRC motif, which presumably regulate AID's slide/hop scanning motion [1, 10, 11]. In keeping with its role as a haphazard mutator in IgV required to generate mutational diversity, AID acts with remarkable inefficiency (~ 1 to 10%) even when sampling its most favored AAC and AGC hot motifs, with more than 9 out of 10 encounters being non-productive [10, 11]. Thus, by substituting the loop 1 sequence of AID with that of Apo3A the 5' binding behavior of AIDv(15) may be affected resulting in decreased processivity during slide/hop scanning. Whether or not AID binds DNA in a similar "U" conformation as seen in Apo3A or perhaps extending across the positively charged cleft remains to be determined. AID's mysterious ssDNA binding, scanning, processivity and catalytic efficiency behavior should be better revealed when an AID-ssDNA co-crystal becomes available.

Acknowledgements

This research was supported, in whole or in part, by the National Institutes of Health [ES13192 and GM21422 to M.F.G], the Division of Intramural Research of the National Institute of Environmental Health Sciences [Project number: ZIA ES102645 to L.C.P.] and the JSPS Grant-in-Aid for Scientific Research (C) [26460580 to K.M.].

Abbreviations:

AID	activation induced deoxycytidine deaminase
SHM	somatic hypermutation
CSR	class switch recombination
Apo3A	APOBEC3A

References

- [1]. Pham P, Bransteitter R, Petruska J, Goodman MF, Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation, *Nature*, 424 (2003) 103–107. [PubMed: 12819663]

- [2]. MacCarthy T, Kalis SL, Roa S, Pham P, Goodman MF, Scharff MD, Bergman A, V-region mutation in vitro, in vivo, and in silico reveal the importance of the enzymatic properties of AID and the sequence environment, *Proc Natl Acad Sci U S A*, 106 (2009) 8629–8634. [PubMed: 19443686]
- [3]. Beale RC, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS, Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo, *J Mol Biol*, 337 (2004) 585–596. [PubMed: 15019779]
- [4]. Pham P, Landolph A, Mendez C, Li N, Goodman MF, A Biochemical Analysis Linking APOBEC3A to Disparate HIV-1 Restriction and Skin Cancer, *J Biol Chem*, 288 (2013) 29294–29304. [PubMed: 23979356]
- [5]. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, Yee D, Temiz NA, Donohue DE, McDougle RM, Brown WL, Law EK, Harris RS, APOBEC3B is an enzymatic source of mutation in breast cancer, *Nature*, 494 (2013) 366–370. [PubMed: 23389445]
- [6]. Pham P, Afif SA, Shimoda M, Maeda K, Sakaguchi N, Pedersen LC, Goodman MF, Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification, *DNA Repair (Amst)*, 43 (2016) 48–56. [PubMed: 27258794]
- [7]. Ta VT, Nagaoka H, Catalan N, Durandy A, Fischer A, Imai K, Nonoyama S, Tashiro J, Ikegawa M, Ito S, Kinoshita K, Muramatsu M, Honjo T, AID mutant analyses indicate requirement for class-switch-specific cofactors, *Nat Immunol*, 4 (2003) 843–848. [PubMed: 12910268]
- [8]. Barreto V, Reina-San-Martin B, Ramiro AR, McBride KM, Nussenzweig MC, C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion, *Mol Cell*, 12 (2003) 501–508. [PubMed: 14536088]
- [9]. Mu Y, Prochnow C, Pham P, Chen XS, Goodman MF, A Structural Basis for the Biochemical Behavior of Activation-induced Deoxycytidine Deaminase Class-switch Recombination-defective Hyper-IgM-2 Mutants, *J Biol Chem*, 287 (2012) 28007–28016. [PubMed: 22715099]
- [10]. Pham P, Calabrese P, Park SJ, Goodman MF, Analysis of a single-stranded DNA-scanning process in which activation-induced deoxycytidine deaminase (AID) deaminates C to U haphazardly and inefficiently to ensure mutational diversity, *J Biol Chem*, 286 (2011) 24931–24942. [PubMed: 21572036]
- [11]. Mak CH, Pham P, Afif SA, Goodman MF, A Mathematical Model for Scanning and Catalysis on Single-stranded DNA, Illustrated with Activation-induced Deoxycytidine Deaminase, *J Biol Chem*, 288 (2013) 29786–29795. [PubMed: 23979486]
- [12]. Shinkura R, Ito S, Begum NA, Nagaoka H, Muramatsu M, Kinoshita K, Sakakibara Y, Hijikata H, Honjo T, Separate domains of AID are required for somatic hypermutation and class-switch recombination, *Nat Immunol*, 13 (2004) 13.
- [13]. Kobayashi M, Takemoto M, Honjo T, The novel activation-induced deoxycytidine deaminase (AID) mutants, AID Δ v and AID Δ 15 are defective in SHM and CSR, *DNA Repair (Amst)*, (2017).
- [14]. Shi K, Carpenter MA, Banerjee S, Shaban NM, Kurahashi K, Salamango DJ, McCann JL, Starrett GJ, Duffy JV, Demir O, Amaro RE, Harki DA, Harris RS, Aihara H, Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B, *Nat Struct Mol Biol*, 24 (2017) 131–139. [PubMed: 27991903]
- [15]. Senavirathne G, Bertram JG, Jaszczur M, Chaurasiya KR, Pham P, Mak CH, Goodman MF, Rueda D, Activation-induced deoxycytidine deaminase (AID) co-transcriptional scanning at single-molecule resolution, *Nat Commun*, 6 (2015) 10209. [PubMed: 26681117]
- [16]. Love RP, Xu H, Chelico L, Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A, *J Biol Chem*, 287 (2012) 30812–30822. [PubMed: 22822074]
- [17]. Kohli RM, Abrams SR, Gajula KS, Maul RW, Gearhart PJ, Stivers JT, A portable hotspot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase, *J Biol Chem*, 284 (2009) 22898–22904. [PubMed: 19561087]
- [18]. Carpenter MA, Rajagurubandara E, Wijesinghe P, Bhagwat AS, Determinants of sequence-specificity within human AID and APOBEC3G, *DNA Repair (Amst)*, 9 (2010) 579–587. [PubMed: 20338830]

- [19]. Wang M, Rada C, Neuberger MS, Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID, *J Exp Med*, 207 (2010) 141–153. [PubMed: 20048284]
- [20]. DeLano WL, The PyMOL Molecular Graphics System, in, DeLano Scientific LLC, San Carlos, CA, 2002.

Figure 1. Deamination profile of AID, AIDv($\Delta 15$) and Apo3A.

Each *colored bar* represents a fraction of mutated clones with a C $\rightarrow$ U deamination at the indicated position on the *lacZ α* sequence (−217 to +149). *Red bars* identify C deaminations occurring in 5'-WRC hot spot motifs, *blue bars* represent 5'-SYC cold spot motifs, and *green bars* represent neutral motifs (neither WRC nor SYC).

A

	Transfected Ramos B-cells		
	Mock control	AID-WT	AIDv(Δ C15)
Number of sequenced IgV _H clones	50	58	67
Number of substitution mutations	16	64	75
Mutation frequency (x 10 ⁻³ per bp)	0.96	3	3.3

B

Figure 2. Comparison of AID-WT and AIDv(Δ 15)-induced IgV_H SHM in Ramos B-cells.

A) Retroviruses, carrying wild-type AID or AIDv(Δ 15) along with the mock control were used to infect Ramos B-cells. After culturing for 2 weeks, IgV_H region DNA from infected B-cells was subjected to DNA sequencing analysis. **B)** Mutation profiles of AID-WT (*top*) and AIDv(Δ 15) (*bottom*). Horizontal numbered red lines indicate a duplication of the sequences, dotted lines indicate deletion events, solid lines indicate AID WRC hot motifs. Green font indicates IgV CDR regions.

Figure 3. Comparison of AID and Apo3A structures.

A) Superposition of crystal structures of Apo3A (green) with DNA bound (yellow) to AIDv(-15) (pink). The catalytic zinc ion for AIDv(-15) is shown in green and the substrate specificity loop 7 is shown in raspberry. Loop 3 and Loop 7 represent the largest differences in these two structures. **B)** Closeup of the loop 7 binding pocket for the base at the -1 position to the substrate dC. D131 of Apo3A forms a hydrogen bond to the dT at the -1 position. **C)** Surface rendering of a full atom model of AIDv(-15) crystal structure with the DNA from Apo3A superimposed into the active site. Zinc ion is shown in green. The partial cyan atoms are the 18-GIG-23 of the modified loop1 from AIDv(-15) with the wild-type residues WAK between I19 and G23 absent. With these residues present the 5' end of the DNA could differ in position with respect to that of Apo3A. A magenta arrow represents a possible alternative binding position for the 5' end of the DNA. Mutations in residues

labeled in green reduce the processivity of AIDv(15). The surface was calculated using the Adaptive Poisson-Boltzmann Solver tool in PyMol [20].

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript