Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

DNA Repair (Amst). Author manuscript; available in PMC 2017 July 01.

Published in final edited form as:

DNA Repair (Amst). 2016 July ; 43: 48–56. doi:10.1016/j.dnarep.2016.05.029.

Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification

Phuong Pham[1,#], **Samir A. Afif**[1], **Mayuko Shimoda**[2,3,4], **Kazuhiko Maeda**[3,4], **Nobuo Sakaguchi**[4,5], **Lars C. Pedersen**[6,#], and **Myron F. Goodman**[1,7,*]

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

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

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

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

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

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

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

Abstract

Activation-induced deoxycytidine deaminase (AID) initiates somatic hypermutation (SHM) and class-switch recombination (CSR) by deaminating C→U during transcription of Ig-variable (V) and Ig-switch (S) region DNA, which is essential to produce high-affinity antibodies. Here we report the crystal structure of a soluble human AID variant at 2.8 Å resolution that favors targeting WRC motifs (W = A/T, R = A/G) *in vitro*, and executes Ig V SHM in Ramos B-cells. A specificity loop extending away from the active site to accommodate two purine bases next to C, differs significantly in sequence, length, and conformation from APOBEC proteins Apo3A and Apo3G, which strongly favor pyrimidines at -1 and -2 positions. Individual amino acid contributions to specificity and processivity were measured in relation to a proposed ssDNA binding cleft. This

*To whom correspondence should be addressed: mgoodman@usc.edu.
#Equally contributing authors

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Author contributions

P.P., L.C.P., N.S. and M.F.G. conceived and designed the study; P.P., L.C.P., S.A.A, M.S. and K.M. performed experiments and analyses; P.P., L.C.P., S.A.A, N.S. and M.F.G. drafted and edited the manuscript.

Competing financial interests

The authors declare no competing financial interests

Structural coordinates for AIDv(Δ15) was deposited to PDB (ID: 5JJ4)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

study provides a structural basis for residue contributions to DNA scanning properties unique to AID, and for disease mutations in human HIGM-2 syndrome.

Keywords

AID X-ray crystal structure; Antibody diversity; IgV somatic hypermutation; scanning processivity; human HIGM-2 syndrome

1. INTRODUCTION

Activation-induced deoxycytidine deaminase (AID) plays an indispensible role in immunoglobulin (Ig) diversification by initiating somatic hypermutation (SHM) and class-switch recombination (CSR) in B-cells [1]. When acting on-target, AID initiates SHM and CSR by deaminating C→U during the transcription of Ig variable (V) and switch (S) regions [2–4]. Replication of U generates C→T mutations at the deamination site; mutations at distal A and T sites occur when G•U mispairs are subjected to mismatch or base excision repair, involving error-prone DNA polymerases to fill in the repair gaps [5, 6]. Acting in concert, CSR and SHM give rise to diversified isotype-switched and antigen specific high affinity antibodies [2–4]. However, when acting off-target, AID can generate non-Ig genomic mutations causing B-cell lymphoma [7, 8].

AID deaminates C→U on ssDNA; it does not function on dsDNA or RNA [9]. AID scans ssDNA processively, catalyzing multiple deaminations per binding event [10–12]. AID has a distinctive catalytic specificity, favoring C targets in WRC (W = A/T, R = A/G) “hot” motifs, while acting much less frequently at SYC (S = G/C, Y = C/T) “cold” motifs [10, 13]. Despite exhibiting a clear mutational preference for hot motifs, AID is nonetheless highly inefficient typically deaminating its most favored hot spots, AAC and AGC, at only 1 – 7% per motif encounter [11, 12]. Owing to the intrinsic inefficiency of AID, one observes a broad clonal mutational heterogeneity within V-regions [14–16], which is key to generating a diverse Ab repertoire. The processive and stochastic behavior of AID scanning ssDNA [10–12] is also a major contributor to deleterious off-target consequences. Repetitive back-and-forth scanning of persisting regions of genomic ssDNA results in a characteristic mutational signature, namely the presence of random clusters of mutations in WRC hot motifs, termed *kataegis*, found in B-cell lymphomas [17] or yeast [18, 19].

AID is a member of the APOBEC protein family of polynucleotide C or dC deaminases that contribute to diverse cellular functions, such as innate immunity against retroviruses and endogenous retroelements [20, 21]. In contrast to AID, which acts at 5'-WRC motifs preferentially, all other APOBECs favor pyrimidines at adjacent -2 and -1 bases on the 5'-side of C. For example, APOBEC3A (Apo3A) favors YYC motifs [22, 23], whereas Apo3G strongly favors only a single CCC motif [24, 25]. Despite a high degree of homology among APOBEC protein members and the availability of structural information for deaminase domains of Apo3A, Apo3B, Apo3C, Apo3G, Apo3F and Apo2 [26–35], the structural basis underlying the mechanism for the unique WRC deamination signature of AID is not understood.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Here, we report the crystal structure of a soluble human AID variant, $AID_v(\Delta 15)$, at 2.8 Å resolution. Native AID, expressed either in insect cells or *E. coli*, is poorly soluble, tends to aggregate, and is difficult to obtain in a highly purified form at sufficiently high concentrations suitable for X-ray structural analysis. We show that $AID_v(\Delta 15)$ generates IgV SHM in Ramos B-cells, exhibiting mutational spectra similar to native AID *in vivo*. AID differs distinctively from the other APOBECs, in the size and orientation of its substrate specificity loop. The larger AID loop extends away from the active site, allowing AID to accommodate two purines next to a target C. A biochemical analysis carried out using AID mutants reveals significant contributions to catalytic specificity and processivity made by amino acids within and outside the specificity loop. In addition, the structure provides fresh insight into how point mutations that cause human hyper IgM type 2 syndrome (HIGM-2) may alter the structure and function of AID.

2. MATERIAL AND METHODS

2.1. AID/APOBEC enzymes

*Sf*9 expressed GST-tagged AID, Apo3G and untagged Apo3A were purified as described [10, 23, 25]. A plasmid for expression of AIDv in baculovirus infected *Sf*9 insect cells was constructed by cloning the coding portion of AIDv into the pAcG2T vector (BD Biosciences). Recombinant virus was obtained by co-transfection of pAcG2T-AIDv and linearized baculovirus DNA (BD Biosciences) into *Sf*9 insect cells (Invitrogen). AIDv was expressed in *Sf*9 insect cells following infection with the recombinant baculovirus at an MOI of 3 and harvested after 72 h. Cells harvested from 1 L cultures were suspended in 50 ml insect cell lysis buffer (10 mM HEPES pH 7.5, 250 mM NaCl, 0.1 % Triton X-100, 10 mM NaF, 10 mM NaH_2PO_4, 10 mM $Na_4P_2O_7$, 1 mM EDTA, 10 mM dithiothreitol, 50 μg/ml RNAseA, 10% glycerol and protease inhibitor) and lysed by sonication. The crude lysate, containing soluble GST-AIDv was collected after centrifugation at 12,000 rpm for 30 min followed by incubation with Glutathione Sepharose resin (GE healthcare) at 4 °C overnight. After extensive washing with PBS buffer, 120 units of Thrombin (Calbiochem) were incubated for 6 hr at room temperature to cleave the GST-tag. AIDv was eluted and concentrated using an Ultracel-10K filter unit (Millipore) and applied to Superdex 75 gel filtration column (GE Healthcare), equilibrated with GF-buffer (20 mM Tris pH 7.5, 500 mM NaCl, 2 mM dithiothreitol and 0.5 mM EDTA). Fractions containing chromatographically homogeneous monomeric AIDv were pooled, dialyzed in GF buffer containing 10% glycerol, and stored at −80 °C.

MPB-AIDv(Δ15) expression vector was constructed by cloning the coding sequence of AIDv(Δ15) into pMALX plasmid [36]. MPB-AIDv(Δ15) mutants were constructed by site directed mutagenesis using QuickChange II site-directed mutagenesis kit (Agilent). *E. coli* cells (CSH50 strain) harboring the MPB-AIDv(Δ15) or mutant expressing plasmid were grown at 30°C until OD600 = 0.5 and the protein was expressed for 4 hours at 18°C. Cells harvested from 5 liters of culture were resuspended in 250 ml of lysis buffer (20 mM Tris, pH 7.5, 0.5 M NaCl, 1 mM DTT, 2 mM EDTA and 1 mM PMSF) and lysed by French press. After centrifugation at 15,000 × g for 30 min, the crude lysate was incubated with 12 ml amylose resin (New Englend Biolabs) at 4°C. MPB-AIDv(Δ15) was washed then eluted

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

using the lysis buffer containing 40 mM maltose and applied to Superdex 75 gel filtration column (GE Healthcare), equilibrated with GF-buffer (20 mM Tris pH 7.5, 500 mM NaCl, 1 mM dithiothreitol and 1 mM EDTA and 5 mM maltose). Fractions containing monomeric MPB-AIDv(Δ15) were pooled, concentrated to ~20 – 30 mg/ml and stored at −80°C in the presence of 5% glycerol.

2.2. Crystal growth and structural determination

Original crystals of the MBP-AIDv(Δ15) were grown at 4° C in space group P32 using the sitting drop vapor diffusion technique by mixing 325 nl of 15 mg/ml protein in 20 mM Tris pH 7.5, 500 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol and 5 mM Maltose with 225 nl of reservoir solution containing 5% 2-n-propanol, 50 mM MES pH 6.5, and 100 mM calcium acetate. For data collection, the crystal was transferred to a cryo-solution consisting of 74% reservoir and 26% ethylene glycol, then flash frozen in liquid nitrogen. A 2.8 Å data set was collected from a crystal grown in space group $P2_1$ that was obtained at 4°C by mixing the protein with a 3 fold molar excess of DNA (5'ATCGAGGTGCG/iSp18/CGCACC-3') then diluting down to a solution consisting of 25 mM MES pH 6.0, 75 mM NaCl, 1 mM EDTA, 5 mM Maltose and 5% glycerol followed by concentrating the protein to 9.5 mg/ml. Crystals were grown using the sitting drop vapor diffusion technique by mixing 400 nl of protein:DNA solution with 250 nl of reservoir solution consisting of 15% PEG8000, 100 mM MES pH 6.5, and 200 mM calcium acetate. For data collection crystals were transferred to a cryo-solution consisting of 85% reservoir solution and 15% ethylene glycol then flash frozen in liquid nitrogen. Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory at a wavelength of 1.0 Å. All data were processed with HKL2000 [37]. To solve the structure with molecular replacement a model for MBP from PDBID_code 3DM0 and a modified model of Apo3G PDBID_code 3E1U for AID were used against data collected from the P32 crystal. The structure was subsequently refined at 3.1 Å resolution. A partially refined version of this model was then used for molecular replacement with the data collected from the $P2_1$ crystal at 2.8 Å resolution. All molecular replacements and refinements were carried out in Phenix [38, 39] and model building to the electron density (Supplementary Fig. 5) in COOT [40, 41]. The $P2_1$ data set consists of 3 molecules of MBP-AIDv(Δ15) in the asymmetric unit with a maltotetraose bound to each MBP. The structure possesses good geometry with 96.5% of the residues in the Ramachandran favored region with 0.1% outliers [42].

2.3. Analysis of AID/APOBEC deamination specificity and processivity on lacZ target

AID/APOBEC deamination specificities were determined by measuring mutational spectra in a *LacZα* reporter ssDNA gapped region of M13mp2 dsDNA, as described previously [10, 43]. M13mp2 gapped DNA (500 ng) was incubated with native AID, AIDv(Δ15), AIDv, Apo3A, Apo3G, or AIDv(Δ15) mutants, in a 30 μl reaction volume, for 10 min at 37 °C. The ratio of the enzyme to DNA was chosen so that the fraction of mutant phage to wild type phage is ~ 5%, which ensures that each mutant phage reflects the action of a single AID/APOBEC molecule acting on M13 DNA. The reactions were quenched by a double extraction with phenol:chloroform:isoamyl alcohol (25:24:1). The DNA product was transfected into *ung⁻ E. coli* competent cells and plated on α-complementation host cells.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Conversions of C→U on the DNA substrate were detected as C→T mutations in a *LacZα* target gene. To measure relative deamination activity of AIDv(Δ15) mutants, 4 μg of purified AIDv(Δ15) or each of the AIDv(Δ15) mutants were incubated with 500 ng of the gapped substrate for 10 min at 37 °C. Relative activities were calculated as the ratio of the frequency of M13 mutant plaques for each AIDv(Δ15) mutant to the frequency of M13 mutant plaques for AIDv(Δ15). Deamination spectrum for each AID/APOBEC enzyme was obtained by sequencing ~ 48 deaminated clones and mutability indexes were calculated as described previously [10, 43, 44].

3. RESULTS

3.1. AID crystal structure

AID, when expressed in its native form, is tightly associated with non-specific inhibitory RNAs [9], which interfere with the ability to isolate a soluble protein at a high concentration. The association of AID with RNAs is caused by the positively charge N-terminus containing a cluster of Arg and Lys residues. To obtain a "better behaved" AID construct we introduced 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 strand β1 (Fig. 1a) which reduced the charge of AID from +11.9 to +3.4 at neutral pH. The resulting highly soluble AID variant, AID_v, containing the N-terminus similar to Apo3A, was readily expressed and purified as monomer from *Sf*9 and *E. coli* (see details in Methods). To enhance crystallization, a 15 residue C-terminal truncation of AID_v, AID_v(Δ15), was generated. Although these C-terminal residues are critical for AID function in CSR [45, 46], they are not required for SHM [45–47] and do not affect AID's preference for WRC motifs [44]. As described in detail below in Results, we have verified that AID_v(Δ15) retains all hallmark biochemical properties of AID *in vitro*, as well as the important ability to induce SHM in Ramos B-cells with mutation spectra similar to native human AID (Supplementary Fig.1).

AID_v(Δ15) was crystallized as an MBP-fusion protein in space group $P2_1$ with three molecules in the asymmetric unit that maintain very similar structures (RMSDs 0.27–0.38 Å on ~170 Cα's) (see Fig.1b, c and Supplementary Fig. 2). These crystals were grown in the presence of ssDNA overhang substrate, but no DNA was observed in the crystal structure. AID maintains the canonical APOBEC fold with 6 α-helices surrounding a central 5-stranded β-sheet. At the heart of the active site are four highly conserved residues. Three of these residues (His56, Cys87 and Cys90) are bound to a catalytic zinc ion and the forth, Glu58 has been proposed to serve as a proton shuttle during catalysis [48]. A water molecule is also bound to the Zn ion at the likely position of the amine group for the substrate dC (Fig. 1d).

A superposition of the crystal structure of the dCMP deaminase from bacteriophage S-TIM5 (pdb 4P9C) with dUMP bound at the active site [49]onto AID suggests a probable binding position for the dC nucleotide substrate with the carbonyl product of dU near the position of the Zn chelating water in AID_v(Δ15) (Fig. 1d). In the S-TIM5 crystal structure Tyr116 forms a perpendicular base stack with the uridine and the O2 and O4 atoms of the base form hydrogen bonds with the backbone amides of Ala68 at the top of helix 2 and Cys95. In addition, the 3′OH of the ribose forms a hydrogen bond with Asn43 [49]. These interactions

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

are likely preserved in AID with residues Tyr114, Val57, Cys87 and Asn51 (Fig. 1d, Supplementary Fig. 3), but in AID the substrates 3′OH would be a bridging oxygen in the phosphodiester backbone.

Loop 7 of AID consists of a unique sequence from residues Leu113 to Pro123 that have been shown to dictate the substrate specificity of APOBEC proteins [50–52] (Fig. 1d, e). It has previously been demonstrated that swapping loops between APOBEC proteins switches the specificity of these enzymes [50–52]. The crystal structure reveals that Leu113 and Cys116 cap a hydrophobic pocket of AID consisting or residues from β-strand 5 and α-helices 4 and 6. Cys116 is found packed between Leu113 and Pro123 helping to stabilize the conformation of the loop. In addition Glu122 is found in a position to potentially form a hydrogen bond with the backbone amide of Tyr88 (average distance between 3 molecules in the asymmetric unit 3.2 Å), which could also help stabilize this conformation (Fig. 1e).

The major structural differences between our crystal structure of $AID_v(\Delta 15)$ and those of APOBEC Apo3A [34] and Apo3G [35] are at loop 3 and the substrate specificity loop 7 (Fig. 2a – d). The specificity loop 7 of $AID_v(\Delta 15)$ shares only one conserved residue with the others (Tyr114) and is two residues longer (Supplementary Fig. 4). After the conserved Tyr, the structures of the loop deviate with Phe115 stacking against Tyr114 and side-chains from E117, R119, and K120 all located on the same side of the loop (Fig 1d, 2b, 2d). The significant difference in the sequence and structures of this loop likely confer the specificity differences of these enzymes. Although not sequentially conserved, the carboxylate moiety of residues Asp133 from Apo3A and Asp317 from Apo3G is structurally conserved with that of Glu122, and in both Apo3A and Apo3G these residues form hydrogen bonds with the amide at the top of helix 3.

3.2. Structural determinants for AID WRC deamination specificity

While a co-crystal of an AID/APOBEC bound to DNA remains elusive, the superimposed dUMP molecule bound to the active site of S-TIM5 deoxycytidylate deaminase suggests how the ssDNA substrate might bind to AID. Producing an electrostatic surface [53] of $AID_v(\Delta 15)$ shows that the superimposed dUMP fits cleanly in the catalytic pocket without manipulation (Fig. 3a). The 3′-side of substrate DNA is likely to extend from the 3'-bridging oxygen of the ribose associated with the positively charged crest of the pocket formed by residue N51 and exits adjacent to loop 3, while the 5'-side of DNA starts from the 5’-phosphate extending away from the catalytic pocket toward a cleft between loop 7 and loop 1 (Fig. 3a, yellow lines). The putative DNA path across the active site of $AID_v(\Delta 15)$ (Fig. 3a) suggests that beside the substrate specificity loop 7, loop 1 may also be involved in interacting with the 5' adjacent nucleotides on the other side of the positively charged cleft.

3.3. Catalytic motif selectivity

To verify the crystal structure of $AID_v(\Delta 15)$ is a good surrogate for that of AID, we have determined the motif selectivity of $AID_v(\Delta 15)$ and AID_v in comparison with native AID, Apo3A and Apo3G. *LacZα* mutational spectra were measured for native AID and AID_v lacking MBP, expressed in *Sf*9 insect cells, and for $AID_v(\Delta 15)$ expressed in *E. coli* as an MBP fusion protein (Fig. 4a). The spectra for $AID_v(\Delta 15)$ and AID_v show basically similar

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

spatial distributions of mutations at hot WRC motifs (Fig. 4, red bars), cold SYC motifs (Fig. 4, blue bars) and neutral motifs (Fig. 4, green bars). While the three AID spectra favor mutations in hot motifs and decrease mutations in cold motifs, the native AID spectrum exhibits a more prominent distribution of hot spot mutations and less pronounced distribution of cold spot mutations compared to both AID variants (Fig. 4). However, in contrast to the three AID spectra, the spectral data for Apo3A contains numerous mutations at AID cold spots, several AID neutral spot mutations, and a few AID hot spot mutations (Fig. 4), while Apo3G strongly favors cold spot mutations (Fig. 4). A clearly distinctive feature is the dense distribution of mutations over the entire *LacZα* target for the three AID enzymes (Fig. 4), compared to the sparse numbers of mutations for Apo3A and Apo3G (Fig. 4), owing presumably to the strong inhibition of C deamination in the presence of proximal purines for both APOBECs [22–25].

These qualitative observations of comparative motif specificity can be converted to quantitative measurements by determining a mutability index (MI) for each enzyme obtained from the *LacZα* spectral data. The MI can be used to quantify catalytic selectivity for hot WRC, cold SYC and neutral motifs. MI is calculated as the ratio of the number of times a given trinucleotide motif within a segment of DNA (365 nt *LacZα* target) contains a mutation, divided by the number of times the motif would be expected to be mutated for a mechanism with no sequence bias [10, 16].

We have measured the average mutability index (MI) for the purified native forms of human AID, Apo3A, Apo3G, and for two variant forms of AID, $AID_v(\Delta 15)$ and AID_v (Table 1). A comparison of native AID, and both AID variants, with Apo3A and Apo3G confirms that the WRC hot motifs for AID (Average MI ~ 1.1 – 1.8) are cold for Apo3A (Average MI = 0.2) and Apo3G (Average MI = 0.1), and that SYC cold motifs for AID (Average MI ~ 0.5 – 1.0) are hot for Apo3A (Avergae MI = 2.1) and Apo3G (Average MI = 3.2) (Table 1). The spectral and MI data for the native forms of AID, Apo3A, and Apo3G measured here agree with previous data obtained for each of the native enzymes [10, 22–25]. We conclude that the variant forms of AID, containing a modified N-terminal region, retain AID's distinctive catalytic specificity favoring WRC motifs and disfavoring SYC motifs (Table 1). A detailed tabulation of MI, was made for all 16 trinucleotide motifs at 106 target C spanning the entire *LacZα* (Supplementary Table 2).

3.4. Effects of active site pocket, specificity loop, and "substrate contact" mutants on processivity

We turn now to the effects on catalytic processivity caused by mutations in the specificity loop (loop 7, Fig. 1d), near the active site (Fig. 1d), and at presumed DNA contact points (Fig. 3a,b). Catalytic processivity is defined as the number of C deaminations (scored as C→T mutations in *LacZα*) that occur during an encounter of a single AID molecule with one ssDNA substrate, prior to dissociation and subsequent reassociation with another substrate [10– 12]; we have shown at single molecule resolution that AID remains bound to a ssDNA substrate for a median time of ~ 4.5 min [54]. Our experimental conditions (AID/DNA ratio, incubation time) are adjusted so that ~ 5 % of the DNA substrates are mutated (white or light blue M13 plaques), which ensures that essentially all of the mutated

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

DNA clones are acted on by a single molecule of AID, in accordance with Poisson statistics. A detailed description of the “single-hit” assay and analysis has been reported previously [10–12].

We’ve investigated the effects on processivity involving three classes of AID mutants, “active site pocket”, “specificity loop”, and presumed “substrate contact” mutants. Based on the crystal structure of $AID_V(\Delta 15)$, Ala substitutions were introduced into $AID_V(\Delta 15)$ at two highly conserved sites near the active center (N51 and Y114) (Figs. 1d, and 3a), at 6 sites in the substrate specificity loop 7 (F115, C116, E117, R119, K120, E122) (Fig. 1d), and at 3 sites located on helix α6 (R171, Q175, R178), which, based on our structural model, could potentially interact with the 5'-side of the DNA substrate (Fig. 3a, b). In addition, K52 from loop 3, presumably located on the 3'-side of the dC substrate, was also mutated. The active site pocket N51A mutant having an average of 3.7 mutations/clone is 2.7-fold less processive than its parent $AID_V(\Delta 15)$ with 10 mutations/clone, while an adjacent K52A mutation reduces processivity by about 1.7-fold (Table 2). A mutation at the edge of the specificity loop, Y114A, reduces processivity by about 4.3-fold, which is comparable to active site mutant N51A; both mutants are roughly 20-fold less active than $AID_V(\Delta 15)$, which is likely responsible for these mutants having the fewest mutations/clone among the twelve evaluated (Table 2). Based on the crystal structure, it appears that Y114 is positioned optimally to stabilize dC in the active site (Fig. 1d,e). The rest of the specificity loop mutants (positions 115 – 122) have little effect on processivity, although R119A does show about a 1.5-fold increase in processivity (~ 15 mutations/clone). Resides117 – 120 of loop 7 are “flipped-away” with respect to the active site (Fig. 1d).

The presumed substrate contact mutants on helix α6 at residues R171, Q175, R178 and K52 from loop 3, show about a 2-fold lower processivity (Table 2). Since these amino acids are remote from the active site (Fig. 3), 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 [10–12, 54].

3.5. Impact of HIGM-2 mutations on AID structure and functions

Based on the crystal structure of AID we are now able to better understand how point mutations in AID may lead to phenotypic disease states associated with HIGM-2 [55]. Each of the 23 missense mutations identified in patients with HIGM-2 syndrome is displayed on the $AID_V(\Delta 15)$ structure (Fig. 3c).

Mutations of H56Y, E58K, C87R, and C87S are all of residues found at the catalytic core of the enzyme. Both His56 and C87 bind the catalytic zinc ion while E58 is the proposed catalytic proton shuttle for the deaminase reaction. Mutations M6T, F11L, F15L, W80R, L98R, L106P, A111E, I136K, M139T, M139V and F151S all alter residues that are tightly packed in or against hydrophobic pockets. Mutations in these residues likely affect the structural integrity of the protein. The L113P mutation may also affect the structural integrity of the protein, but L113 is also located at the beginning of the substrate specificity loop where such a mutation could- alter the conformation of this loop and disrupt substrate binding. Mutations at S83P and S85N, in addition to potential steric effects, likely destabilize the enzyme as the hydroxyls from both residues are involved in forming

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

hydrogen bonds with amides from neighboring residues. Similarly, the side chain of R112, located right before the substrate binding loop, points back into the core of the molecule forming hydrogen bonds with backbone carbonyls from Asn14 and Asn16. The mutations R112C and R112H would disrupt these seemingly important interactions. The most likely mutations that would directly affect substrate binding are R24W from loop 1 and R174S from helix 6 as these arginine residues are located in an area consistent with ssDNA binding. Mutations at these residues as well as S43P are located on the surface of the protein and thus could potentially disrupt protein-protein interactions important for either SHM or CSR.

4. DISCUSSION

We have obtained a crystal structure for a human AID variant, $AID_v(\Delta 15)$, at a resolution of 2.8 Å (Fig. 1b–d). When expressed in Ramos B-cells, $AID_v(\Delta 15)$ catalyzes SHM showing a similar increase in IgV_H hypermutation frequency, accompanied by a similar C→T base substitution mutational spectrum, compared with the expression of native human AID (Supplementary Fig. 1). $AID_v(\Delta 15)$ and native AID exhibit similar catalytic properties favoring deamination in WRC over SYC motifs (Fig. 4, Table 1). By mirroring these key criteria, SHM and mutational spectra *in vivo*, and catalytic deamination specificity *in vitro*, our structure for $AID_v(\Delta 15)$ can be used as a surrogate for native AID to investigate the molecular basis for AID's unique mutational signature.

Catalytic specificity is perhaps the most distinctive feature that sets AID apart from other APOBEC-family members. AID exhibits a moderately strong catalytic preference for converting C→U in 5′-WRC trinucleotide motifs accompanied by a moderately strong aversion against acting in SYC motifs [10, 43](Table 1). In contrast, Apo3G is most active by far in 5′-CCC (Table 1). Apo3A, Apo3B, and Apo3F also tend to favor 5′-YYC, but in a less restrictive manner compared to Apo3G. Each of these APOBEC dC deaminases discriminates strongly against WRC motifs. These reciprocal AID vs. APOBEC motif specificity constraints can be represented quantitatively by their respective MI values: AID (WRC) ~ 1.8, AID (SYC) ~ 0.5; Apo3A (WRC) ~ 0.2, Apo3A (SYC) ~ 2.1; Apo3G (WRC) ~ 0.1, Apo3G (SYC) ~ 3.2 – (Table 1). The crystal structures for the APOBEC enzymes are indeed all quite similar, sharing a conserved fold and a loop (loop 7) of similar size, located before helix 4, that determines the enzymes specificity for ssDNA.

The crystal structure for AID reveals the specificity loop is decidedly different. The specificity loop 7 for AID is significantly larger compared to the APOBEC structures, and it extends out from the active site (Figs. 1, 2, 3). These two distinguishing features likely account for the different catalytic motif specificities allowing AID to accommodate two purines on the 5'-side adjacent to the substrate dC (WRC motif) without stringently discriminating against pyrimidines. The shorter loops of the other APOBEC enzymes on the other hand strongly favor two smaller pyrimidines bases at these positions (SYC motif). Since the pocket for AID can accommodate nearest-neighbor Pur or Pyr, that might explain why AID exhibits a wide range of C deamination patterns, which include WRC, SYC and neutral trinucleotide motifs [10, 43](Table 1). A comparison of MI ratios can be used to illustrate this point: the ratio of hot WRC/cold SYC MI values for AID ~ 3.6, which is 36-fold greater than Apo3A ~ 0.1, and 120-fold greater than Apo3G ~ 0.03 (Table 1). As yet

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

there are no structural data with AID bound to DNA. Nevertheless, the $AID_v(\Delta 15)$ structure with a superimposed dUMP, based on a co-crystal structure of S-TIM5 dCMP deaminase, suggests a plausible ssDNA binding path paved with positively charged amino acids (Fig. 3a, yellow lines). The structure also suggests how missense mutations that span almost the entire gene for AID (Fig. 3c) contribute to loss of catalytic function that results in human HIGM-2 syndrome, a defect in CSR causing an accumulation of IgM and absence of IgG, IgA, and IgE isotypes [55].

On the biochemical side of the ledger, AID scans ssDNA processively in short slides/hops [10⁻12, 54]. Yet, despite exhibiting distinct catalytic preferences, AID shows no discernible difference in binding to ssDNA having hot motifs, cold motifs, or even no motifs, i.e., the absence of C [44, 56]. Even its hottest motifs, AAC, AGC, are largely ignored by AID, with deaminations occurring from 1 – 7% per motif encounter [11, 12, 57]. The $AID_v(\Delta 15)$ structure contains a positively charged channel on both sides of the active site that could perhaps serve as a ssDNA substrate binding surface (Fig. 3a) on which AID can move along freely, pausing infrequently to act on WRC, and even less frequently at SYC. Our data show that mutations at four residues along the surface, R171A, Q175A, and R178A helix 6 and K52A on loop 3, cause a 2-fold reduction in AID processivity, which is consistent with these presumed AID-ssDNA contacts being reduced by a decrease in positive charge. An intuitive picture would have AID scanning essentially unconstrained along ssDNA in short back-and-forth slides/hops. This type of inefficient, haphazard catalytic process would seem ideal for an enzyme whose central function is to generate mutational diversity to guarantee Ab diversity. On the other side of the coin, however, this same type of stochastic recurrent scanning in non-Ig genomic ssDNA appears responsible for B-cell lymphomas, which contain a clear spatial signature of AID in the form of *kataegis*, mutational clusters concentrated in WRC motifs [17].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Alice Landolph for purification of AIDv and AIDv(Δ15) proteins. 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.]. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences Contract W-31-109-Eng-38.

Abbreviations

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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Apo3G APOBEC3G

HIGM-2 Hyper IgM syndrome type 2

REFERENCES

1. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell. 2000; 102:553–563. [PubMed: 11007474]
2. Di Noia JM, Neuberger MS. Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem. 2007; 76:1–22. [PubMed: 17328676]
3. Peled JU, Kuang FL, Iglesias-Ussel MD, Roa S, Kalis SL, Goodman MF, Scharff MD. The biochemistry of somatic hypermutation. Annu Rev Immunol. 2008; 26:481–511. [PubMed: 18304001]
4. Stavnezer J, Guikema JE, Schrader CE. Mechanism and regulation of class switch recombination. Annu Rev Immunol. 2008; 26:261–292. [PubMed: 18370922]
5. Chelico L, Pham P, Petruska J, Goodman MF. Biochemical basis of immunological and retroviral responses to DNA-targeted cytosine deamination by activation-induced cytidine deaminase and APOBEC3G. J Biol Chem. 2009; 284:27761–27765. [PubMed: 19684020]
6. Chahwan R, Edelmann W, Scharff MD, Roa S. Mismatch-mediated error prone repair at the immunoglobulin genes. Biomed Pharmacother. 2011; 65:529–536. [PubMed: 22100214]
7. Pasqualucci L, Bhagat G, Jankovic M, Compagno M, Smith P, Muramatsu M, Honjo T, Morse HC 3rd, Nussenzweig MC, Dalla-Favera R. AID is required for germinal center-derived lymphomagenesis. Nat Genet. 2008; 40:108–112. [PubMed: 18066064]
8. Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, Kuppers R, Dalla-Favera R. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature. 2001; 412:341–346. [PubMed: 11460166]
9. Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc. Natl. Acad. Sci. USA. 2003; 100:4102–4107. [PubMed: 12651944]
10. Pham P, Bransteitter R, Petruska J, Goodman MF. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature. 2003; 424:103–107. [PubMed: 12819663]
11. 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. 2011; 286:24931–24942. [PubMed: 21572036]
12. 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. 2013; 288:29786–29795. [PubMed: 23979486]
13. 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. 2009; 106:8629–8634. [PubMed: 19443686]
14. Rogozin IB, Kolchanov NA. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim. Biophy. Acta. 1992; 1171:11–18.
15. Dorner T, Foster SJ, Brezinschek H-P, Lipsky PE. Analysis of the targeting of the hypermutational machinery and the impact of subsequent selection on the distribution of nucleotide changes in human VHDJH rearrangements. Immunol Rev. 1998; 162:161–171. [PubMed: 9602362]
16. Shapiro GS, Aviszus K, Murphy J, Wysocki LJ. Evolution of Ig DNA Sequence to Target Specific Base Positions Within Codons for Somatic Hypermutation. J. Immunol. 2002; 168:2302–2306. [PubMed: 11859119]
17. Qian J, Wang Q, Dose M, Pruett N, Kieffer-Kwon KR, Resch W, Liang G, Tang Z, Mathe E, Benner C, Dubois W, Nelson S, Vian L, Oliveira TY, Jankovic M, Hakim O, Gazumyan A, Pavri

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

R, Awasthi P, Song B, Liu G, Chen L, Zhu S, Feigenbaum L, Staudt L, Murre C, Ruan Y, Robbiani DF, Pan-Hammarstrom Q, Nussenzweig MC, Casellas R. B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. Cell. 2014; 159:1524–1537. [PubMed: 25483777]

18. Lada AG, Dhar A, Boissy RJ, Hirano M, Rubel AA, Rogozin IB, Pavlov YI. AID/APOBEC cytosine deaminase induces genome-wide kataegis. Biol Direct. 2012; 7:47. discussion 47. [PubMed: 23249472]
19. Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, Rada C, Stratton MR, Neuberger MS. DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. Elife. 2013; 2:e00534. [PubMed: 23599896]
20. Conticello SG. The AID/APOBEC family of nucleic acid mutators. Genome Biol. 2008; 9:229. [PubMed: 18598372]
21. Chiu YL, Greene WC. Multifaceted antiviral actions of APOBEC3 cytidine deaminases. Trends Immunol. 2006; 27:291–297. [PubMed: 16678488]
22. Love RP, Xu H, Chelico L. Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A. J Biol Chem. 2012; 287:30812–30822. [PubMed: 22822074]
23. 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. 2013; 288:29294–29304. [PubMed: 23979356]
24. Suspene R, Sommer P, Henry M, Ferris S, Guetard D, Pochet S, Chester A, Navaratnam N, Wain-Hobson S, Vartanian JP. APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. Nucleic Acids Res. 2004; 32:2421–2429. [PubMed: 15121899]
25. Chelico L, Pham P, Calabrese P, Goodman MF. APOBEC3G DNA deaminase acts processively 3' -->5' on single-stranded DNA. Nat Struct Mol Biol. 2006; 13:392–399. [PubMed: 16622407]
26. Chen KM, Harjes E, Gross PJ, Fahmy A, Lu Y, Shindo K, Harris RS, Matsuo H. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. Nature. 2008; 452:116–119. [PubMed: 18288108]
27. Holden LG, Prochnow C, Chang YP, Bransteitter R, Chelico L, Sen U, Stevens RC, Goodman MF, Chen XS. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. Nature. 2008; 456:121–124. [PubMed: 18849968]
28. 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]
29. Byeon IJ, Ahn J, Mitra M, Byeon CH, Hercik K, Hritz J, Charlton LM, Levin JG, Gronenborn AM. NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. Nat Commun. 2013; 4:1890. [PubMed: 23695684]
30. Shandilya SM, Nalam MN, Nalivaika EA, Gross PJ, Valesano JC, Shindo K, Li M, Munson M, Royer WE, Harjes E, Kono T, Matsuo H, Harris RS, Somasundaran M, Schiffer CA. Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. Structure. 2010; 18:28–38. [PubMed: 20152150]
31. Kitamura S, Ode H, Nakashima M, Imahashi M, Naganawa Y, Kurosawa T, Yokomaku Y, Yamane T, Watanabe N, Suzuki A, Sugiura W, Iwatani Y. The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. Nat Struct Mol Biol. 2012; 19:1005–1010. [PubMed: 23001005]
32. Bohn MF, Shandilya SM, Albin JS, Kouno T, Anderson BD, McDougle RM, Carpenter MA, Rathore A, Evans L, Davis AN, Zhang J, Lu Y, Somasundaran M, Matsuo H, Harris RS, Schiffer CA. Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. Structure. 2013; 21:1042–1050. [PubMed: 23685212]
33. Shi K, Carpenter MA, Kurahashi K, Harris RS, Aihara H. Crystal Structure of the DNA Deaminase APOBEC3B Catalytic Domain. J Biol Chem. 2015; 290:28120–28130. [PubMed: 26416889]
34. Bohn MF, Shandilya SM, Silvas TV, Nalivaika EA, Kouno T, Kelch BA, Ryder SP, Kurt-Yilmaz N, Somasundaran M, Schiffer CA. The ssDNA Mutator APOBEC3A Is Regulated by Cooperative Dimerization. Structure. 2015; 23:903–911. [PubMed: 25914058]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

35. Lu X, Zhang T, Xu Z, Liu S, Zhao B, Lan W, Wang C, Ding J, Cao C. Crystal structure of DNA cytidine deaminase ABOBEC3G catalytic deamination domain suggests a binding mode of full-length enzyme to single-stranded DNA. J Biol Chem. 2014; 290:4010–4021. [PubMed: 25542899]
36. Moon AF, Mueller GA, Zhong X, Pedersen LC. A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. Protein Sci. 2010; 19:901–913. [PubMed: 20196072]
37. Otwinowski Z, Minor W. Processing of X-ray Diffraction Data Collected in Oscillation Mode, Methods in Enzymology. Macromolecular Crystallography, Pt A. 1997; 276:307–326.
38. McCoy AJ. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallogr D Biol Crystallogr. 2007; 63:32–41. [PubMed: 17164524]
39. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]
40. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
41. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]
42. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010; 66:12–21. [PubMed: 20057044]
43. Bransteitter R, Pham P, Calabrese P, Goodman MF. Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. J Biol Chem. 2004; 279:51612–51621. [PubMed: 15371439]
44. 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. 2012; 287:28007–28016. [PubMed: 22715099]
45. 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. 2003; 4:843–848. [PubMed: 12910268]
46. 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. 2003; 12:501–508. [PubMed: 14536088]
47. McBride KM, Barreto V, Ramiro AR, Stavropoulos P, Nussenzweig MC. Somatic Hypermutation Is Limited by CRM1-dependent Nuclear Export of Activation-induced Deaminase. J Exp Med. 2004; 199:1235–1244. [PubMed: 15117971]
48. Betts L, Xiang S, Short SA, Wolfenden R, Carter CW Jr. Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex. J Mol Biol. 1994; 235:635–656. [PubMed: 8289286]
49. Marx A, Alian A. The first crystal structure of a dTTP-bound deoxycytidylate deaminase validates and details the allosteric-inhibitor binding site. J Biol Chem. 2015; 290:682–690. [PubMed: 25404739]
50. 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. 2009; 284:22898–22904. [PubMed: 19561087]
51. Carpenter MA, Rajagurubandara E, Wijesinghe P, Bhagwat AS. Determinants of sequence-specificity within human AID and APOBEC3G. DNA Repair (Amst). 2010; 9:579–587. [PubMed: 20338830]
52. 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. 2010; 207:141–153. [PubMed: 20048284]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

53. Walsh I, Minervini G, Corazza A, Esposito G, Tosatto SC, Fogolari F. Bluues server: electrostatic properties of wild-type and mutated protein structures. Bioinformatics. 2012; 28:2189–2190. [PubMed: 22711791]

54. 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. 2015; 6:10209. [PubMed: 26681117]

55. Durandy A, Peron S, Fischer A. Hyper-IgM syndromes. Curr Opin Rheumatol. 2006; 18:369–376. [PubMed: 16763457]

56. Larijani M, Petrov AP, Kolenchenko O, Berru M, Krylov SN, Martin A. AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. Mol Cell Biol. 2007; 27:20–30. [PubMed: 17060445]

57. Mak CH, Pham P, Afif SA, Goodman MF. Random-walk enzymes. Phys Rev E Stat Nonlin Soft Matter Phys. 2015; 92:032717. [PubMed: 26465508]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

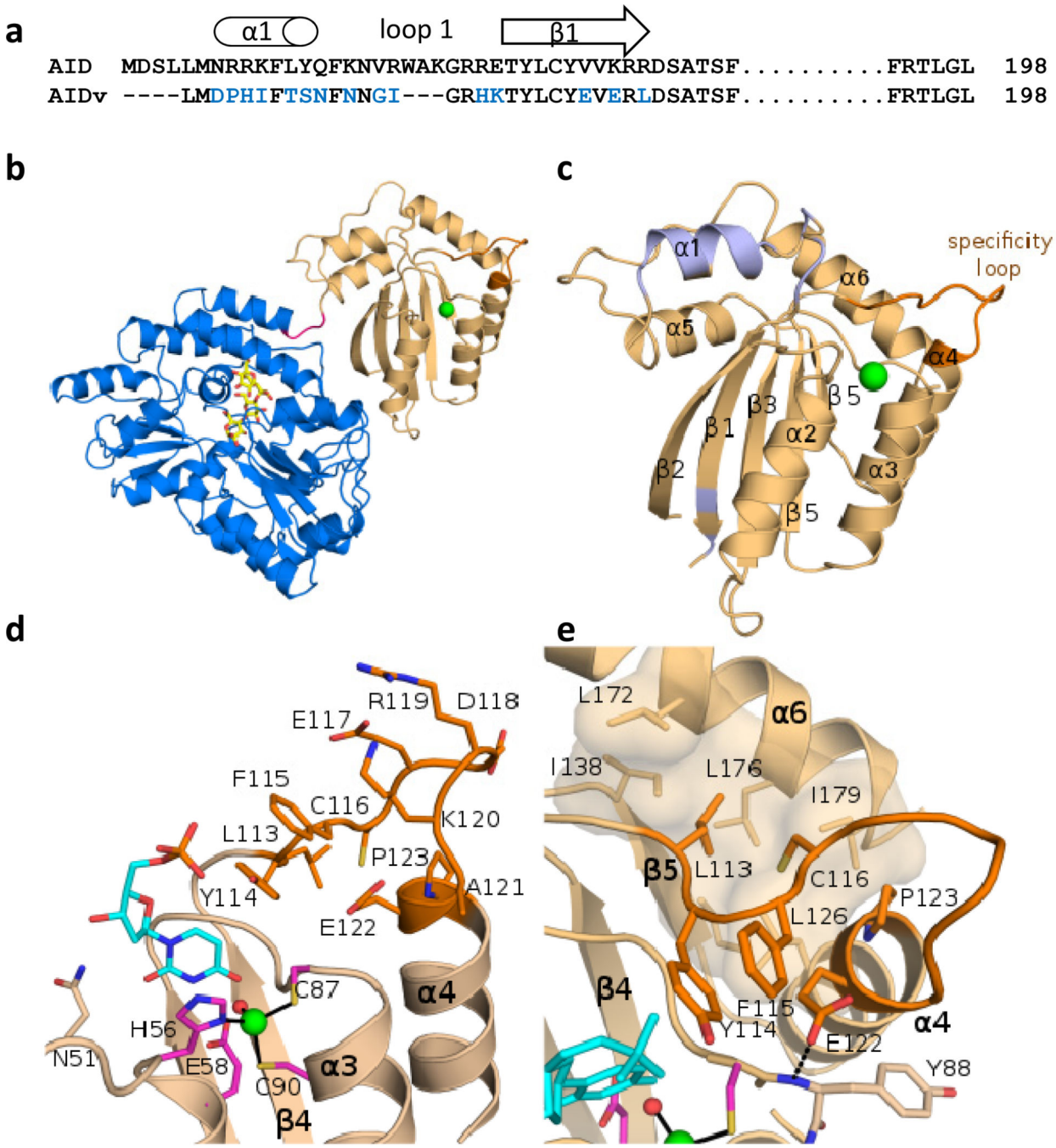

Figure 1. AID crystal structure

a) Sequence differences at the N-terminus between AID and AIDv. **b**) Ribbon diagram of the crystal structure of the MBP-AIDv(Δ15) fusion protein. The MBP is colored metallic blue (with bound maltotetraose in yellow) and the AID is colored peach. **c**) Residues that were modified in AID for solubility purposes are colored cornflower blue. **d**) Active site region of AIDv with a dUMP (cyan) superimposed in to the active site from a crystal structure of deoxycytidylate deaminase from bacteriophage S-TIM5 (pdb 4P9C). Active site residues chelating the Zn ion and the proposed proton shuttle (E58) are shown in magenta. A

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

water molecule coordinating the Zn ion is also shown (red). **e**) Interactions of the substrate binding loop. Residues Tyr114 base stacks with Phe115 and is in position to form a perpendicular base-stacking arrangement with the substrate dC. Leu113 and Cys116 cap off a hydrophobic pocket consisting of residues from beta strand 5, helix 4, and the C-terminal helix 6. Cys116 is sandwiched between residues L113 and Pro123 that define the ends of the substrate binding loop. Glu122 forms a potential hydrogen bond with the backbone amide from Tyr88 at the top of helix 3. These interactions may help maintain the loops conformation. In all figures the Zn ion is colored green and the substrate loop is colored orange.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

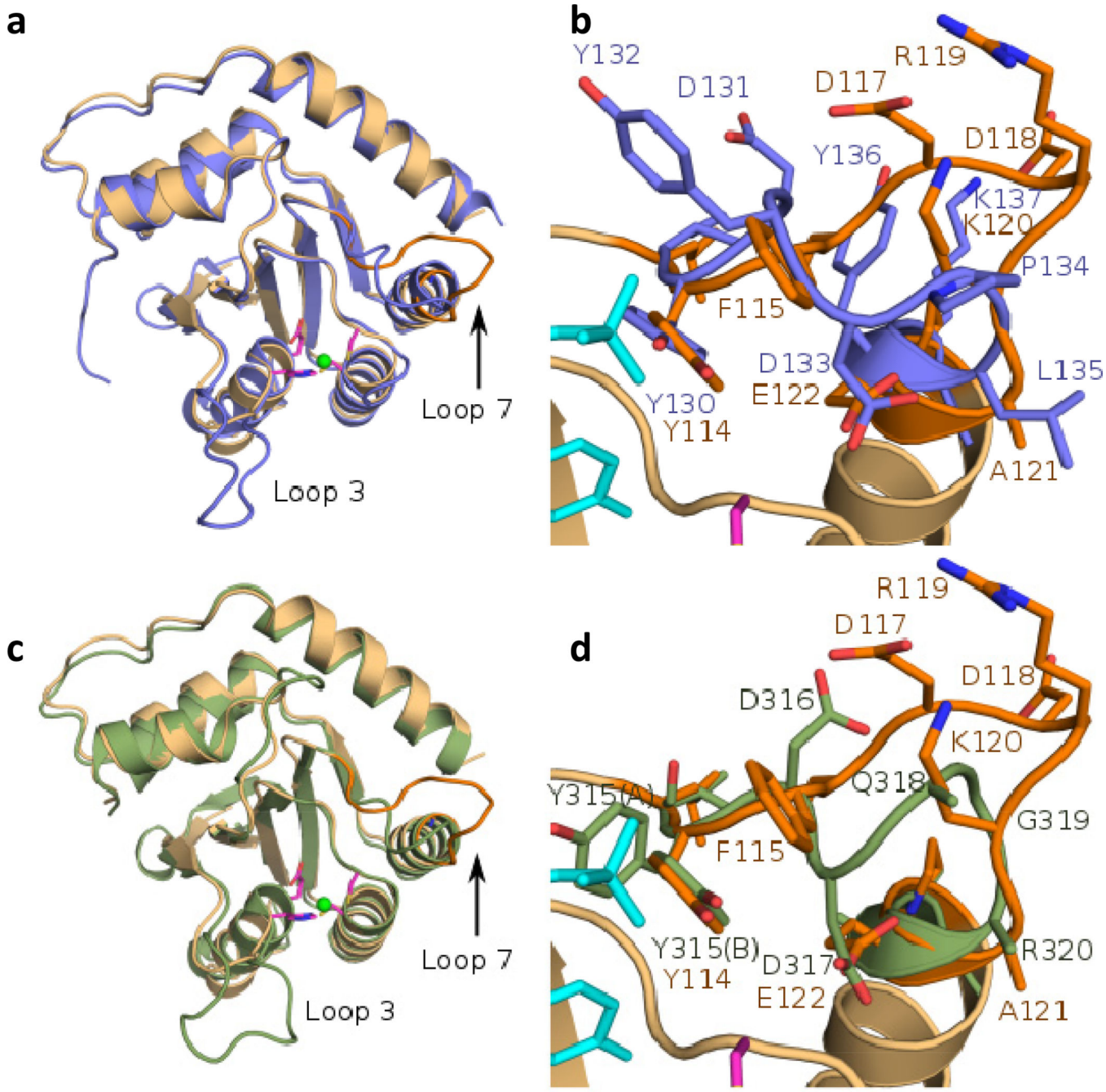

Figure 2. Comparisons of the crystal structure of AID to Apo3A and Apo3G

a) Global superposition (RMSD 1.1 Å over 168 Cα) of the crystal structure of AID (peach) and Apo3A (cornflower blue, pdb 4XXO). **b**) Differences in the conformations and sequence within the substrate binding loop of AID and Apo3A. **c**) Global superposition (RMSD 1.1 Å over 166 Cα) of the crystal structure of AID and Apo3G (green, pdb 4ROV). **d**) Differences in the conformations and sequence within the substrate binding loop of AID and Apo3G. Both Apo3A and Apo3G contain an Asp residue (Asp131, Asp317 from Apo3A and Apo3G respectively) that form a similar backbone interaction with the top of

helix 3 as Glu122 but reside in a different portion of the loop. Zn ion and active site residues from AID are displayed in green and magenta and the superimposed dUMP from S-TIM5 is shown in cyan.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

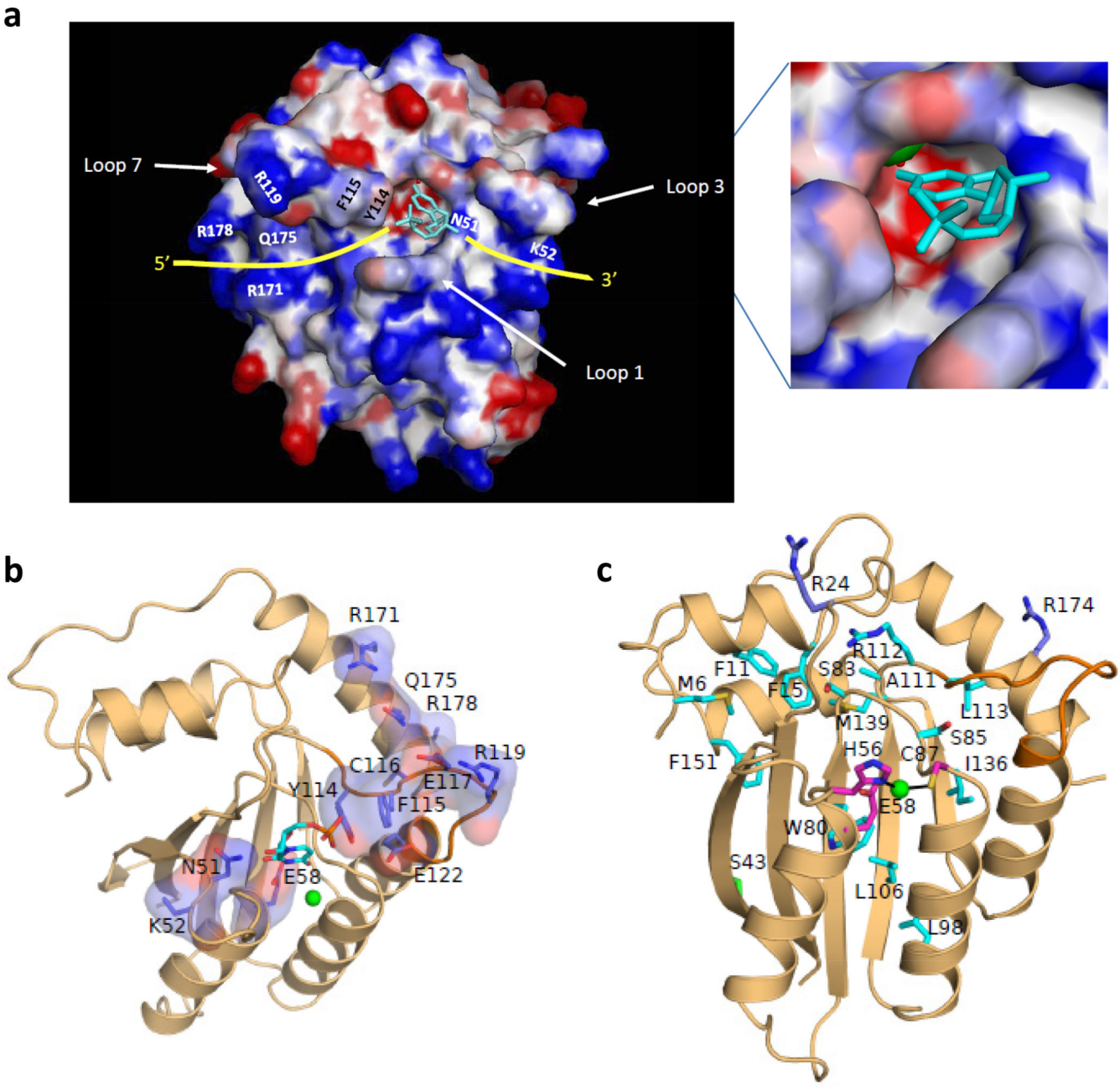

Figure 3. Potential residues responsible for interaction with ssDNA

a) Electrostatic surface representation of AIDv. Arrows indicate position of loop 1, loop 3 and specificity loop 7. The superimposed dUMP from S-TIM5 structure suggests a probable path for the 5' and 3' sides of substrate DNA (yellow). *Inset*, zoom-in the active site with the superimposed dUMP. Zn ion is shown in green and a water molecule coordinating the Zn ion is shown in red. **b**) Selected residues surrounding the active site that could potentially interact with ssDNA that were mutated in this study. The dUMP from S-TIM5 is superimposed in the active site for reference. **c**) Positions of the HIGM-2 disease state mutations found in AID. Residues whose mutations likely affect the stability of AID are shown in cyan. Residues that may interact with the DNA or form protein-protein interactions

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

are shown in purple and S43 which may be involved in protein-protein interactions is shown in green. Zn ion is shown in green and catalytic and Zn ion binding residues are shown in magenta.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

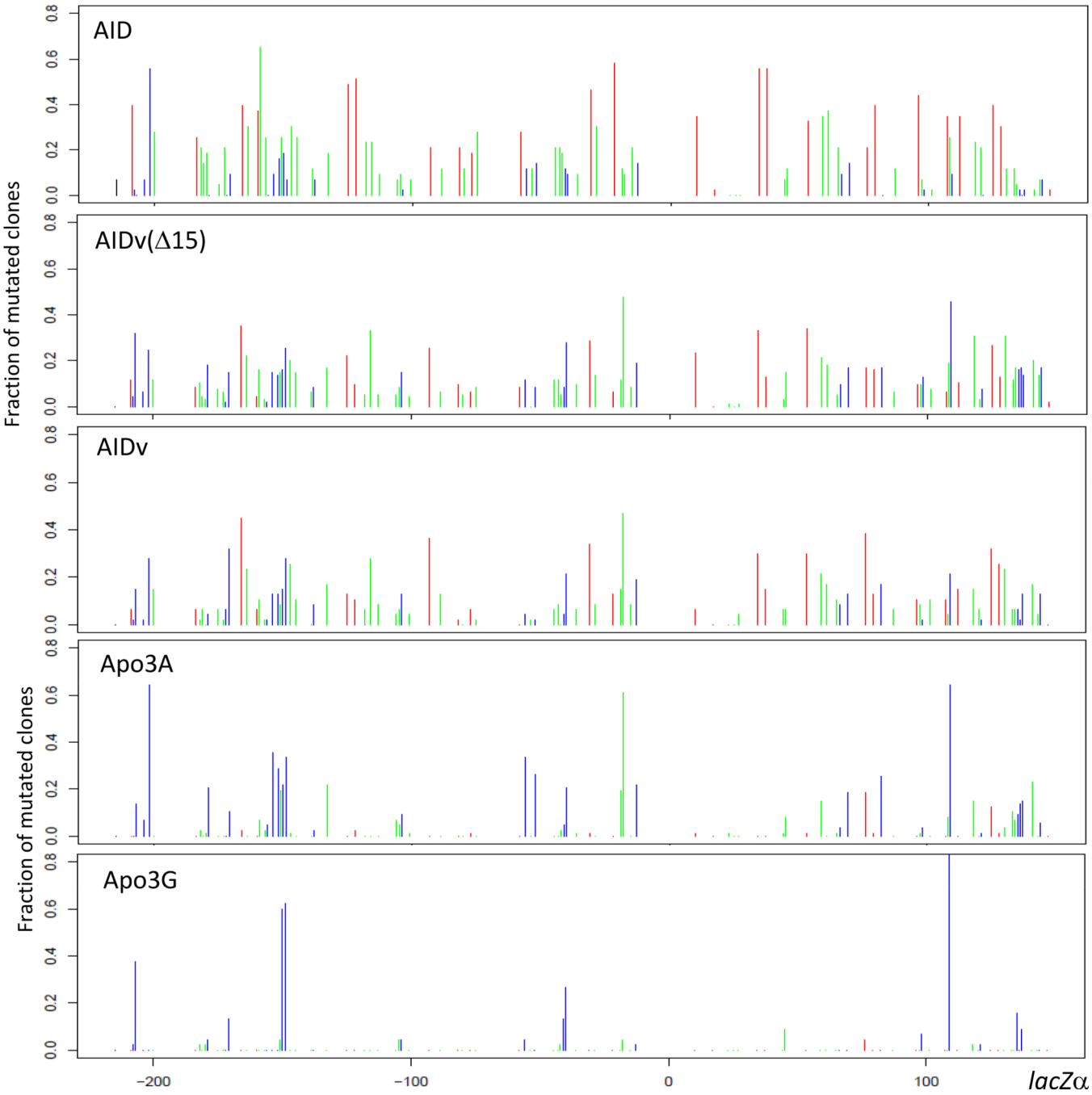

Figure 4. C → U deamination spectra of AID, AIDv(Δ15), AIDv, Apo3A and Apo3G

a) Deamination spectra of purified *Sf*9 expressed AID, AIDv, Apo3A, Apo3G and *E. coli*-expressed AIDv(Δ15). Each *colored bar* represents a fraction of mutated clones with a C → U deamination at the indicated position on the *lacZa* target sequence (−217 to +149). *Red bars* identify C deaminations occurring in 5'-WRC hot motifs, *blue bars* represent 5'-SYC cold motifs, and *green bars* represent neutral motifs (i. e., neither WRC nor SYC).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Comparison of mutability indexes for AID, AIDv(Δ15), AIDv, Apo3A and Apo3G

Mutability Index (MI) *					
Motif (5'-3')	AID	AIDv(Δ15)	AIDv	Apo3A	Apo3G
Hotspot					
AAC	1.7	0.8	1	0	0
AGC	1.5	0.8	1.1	0.1	0
TAC	1.8	1.7	2.6	0.5	0.4
TGC	2.4	1.2	1.7	0.4	0
WRC average	1.8 ± 0.4	1.1 ± 0.4	1.6 ± 0.8	0.2 ± 0.2	0.1 ± 0.2
Coldspot					
CCC	0.3	1.4	1.1	2.2	11.9
CTC	1.1	1.1	0.9	4.1	0.5
GCC	0.2	0.5	0.3	0.3	0.3
GTC	0.4	1	0.6	2	0
SYC average	0.5 ± 0.4	1.0 ± 0.4	0.7 ± 0.3	2.1 ± 1.6	3.2 ± 5.8
Mutability Index (MI) *					
Motif (5'-3')	AID	AIDv(Δ15)	AIDv	Apo3A	Apo3G
Hotspot					
AAC	1.7	0.8	1	0	0
AGC	1.5	0.8	1.1	0.1	0
TAC	1.8	1.7	2.6	0.5	0.4
TGC	2.4	1.2	1.7	0.4	0
WRC average	1.8 ± 0.4	1.1 ± 0.4	1.6 ± 0.8	0.2 ± 0.2	0.1 ± 0.2
Coldspot					
CCC	0.3	1.4	1.1	2.2	11.9
CTC	1.1	1.1	0.9	4.1	0.5
GCC	0.2	0.5	0.3	0.3	0.3
GTC	0.4	1	0.6	2	0

DNA Repair (Amst). Author manuscript; available in PMC 2017 July 01.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

SYC average	0.5 ± 0.4	1.0 ± 0.4	0.7 ± 0.3	2.1 ± 1.6	3.2 ± 5.8
Neutral					
ACC	1.2	0.8	0.3	0.5	1.0
ATC	0.9	1.5	0.9	1.5	0.5
CAC	0.9	0.9	1.0	0.0	0.0
CGC	1.2	0.8	0.6	0.0	0.0
GAC	0.5	0.3	0.7	0.1	0.0
GGC	0.7	0.5	0.7	0.0	0.0
TCC	0.6	1.6	1.2	2.7	1.5
TTC	0.6	0.9	1.3	1.7	0.0
Neutral average	0.8 ± 0.3	0.9 ± 0.4	0.8 ± 0.3	0.8 ± 1.0	0.4 ± 0.6

*The mutability index (MI) is defined as the number of times a given motif within a segment of DNA contains a mutation, divided by the number of times the motif would be expected to be mutated for a mechanism with no sequence bias[10, 16]. MI values were calculated from the spectra in Fig. 4. W = A/T, S= G/C, R = purine, Y = pyrimidine.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Effects of point mutations on AIDv(Δ15) activity and processivity

Enzyme	Activity*	Number of C→U	*p-value***
AIDv(Δ15)	1	10	
N51A	0.04	3.7	< 0.01
K52A	0.75	5.8	< 0.01
Y114A	0.05	2.3	< 0.01
F115A	0.44	5.7	< 0.01
C116A	0.35	9.5	0.68
E117A	0.13	8	0.14
R119A	0.69	15.2	< 0.01
K120A	1.1	11.7	0.23
E122A	0.24	7.1	0.06
R171A	0.17	5.3	< 0.01
Q175A	0.47	5.3	< 0.01
R178A	0.73	4.7	< 0.01

Deamination activity was measured using the *lacZ*a gap deamination assay.

*Relative activities were calculated as the ratio of the frequency of M13 mutant plaques for each AIDv(Δ15) mutant to the frequency of M13 mutant plaques for AIDv(Δ15).

***P*-values for comparison of mutants to AIDv(Δ15) processivity were calculated using an unpaired *t*-test.