AID Recognizes Structured DNA for Class Switch Recombination

Highlights

- Structured substrates, such as G4 substrates, are preferred AID targets in vitro.

- A bifurcated substrate-binding surface supports structured-substrate recognition.

- G4 substrates induce AID oligomerization upon binding.

- Disrupting structured-substrate recognition or AID oligomerization compromises CSR.

Authors

Qi Qiao, Li Wang, Fei-Long Meng, Joyce K. Hwang, Frederick W. Alt, Hao Wu

Correspondence
wu@crystal.harvard.edu

In Brief

Qiao et al. demonstrated that structured substrates, like G4 and branched DNA, are preferred AID targets in vitro. A bifurcated substrate-binding surface in AID structure supports structured-substrate recognition. G4 substrates mimicking the Ig S regions also induce cooperative AID oligomerization. Disrupting structured-substrate recognition or AID oligomerization both compromise CSR.

Qiao et al., 2017, Molecular Cell 67, 361–373

August 3, 2017 © 2017 Elsevier Inc.
http://dx.doi.org/10.1016/j.molcel.2017.06.034
AID Recognizes Structured DNA for Class Switch Recombination

Qi Qiao,1,2,5 Li Wang,1,2,5 Fei-Long Meng,1,2,3,4 Joyce K. Hwang,1,2,3 Frederick W. Alt,1,2,3 and Hao Wu1,2,6,*

1Program in Cellular and Molecular Medicine, Boston Children’s Hospital
2Department of Biological Chemistry and Molecular Pharmacology
3Howard Hughes Medical Institute
Harvard Medical School, Boston, MA 02115, USA
4Present address: Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-yang Road, Shanghai 200031, China
5These authors contributed equally
6Lead Contact
*Correspondence: wu@crystal.harvard.edu
http://dx.doi.org/10.1016/j.molcel.2017.06.034

SUMMARY

Activation-induced cytidine deaminase (AID) initiates both class switch recombination (CSR) and somatic hypermutation (SHM) in antibody diversification. Mechanisms of AID targeting and catalysis remain elusive despite its critical immunological roles and off-target effects in tumorigenesis. Here, we produced active human AID and revealed its preferred recognition and deamination of structured substrates. G-quadruplex (G4)-containing substrates mimicking the mammalian immunoglobulin switch regions are particularly good AID substrates in vitro. By solving crystal structures of maltose binding protein (MBP)-fused AID alone and in complex with deoxycytidine monophosphate, we surprisingly identify a bifurcated substrate-binding surface that explains structured substrate recognition by capturing two adjacent single-stranded overhangs simultaneously. Moreover, G4 substrates induce cooperative AID oligomerization. Structure-based mutations that disrupt bifurcated substrate recognition or oligomerization both compromise CSR in splenic B cells. Collectively, our data implicate intrinsic preference of AID for structured substrates and uncover the importance of G4 recognition and oligomerization of AID in CSR.

INTRODUCTION

Antibody diversification is a central process in adaptive immunity that produces antigen-specific, high-affinity antibodies to combat millions of antigens. In B cells, immunoglobulin (Ig) genes undergo two DNA-alteration events to enhance the specificity and functionality of antibodies: somatic hypermutation (SHM) and class switch recombination (CSR). Until now, activation-induced cytidine deaminase (AID) is the single enzyme that is known to initiate both SHM and CSR. AID belongs to the APOBEC cytidine deaminase family. It converts deoxyctydine into deoxyuridine on single-stranded DNA (ssDNA) substrates in vitro and in vivo but does not show detectable activity on RNA substrates (Bransteitter et al., 2003). Although AID exhibits sequence similarity to APOBEC cytidine deaminases, its critical function in antibody diversification, especially in CSR, cannot be substituted by APOBEC proteins.

Across the genome, AID predominantly mutates immunoglobulin genes at the variable (V) and switch (S) regions. In SHM, V region sequencing data showed that a WRCH (W = A/T, R = A/G, and H = A/C/T) hotspot motif exhibits higher AID-induced mutation probability than average, by about 2- to 4-fold (Larijani et al., 2005; Rogozin and Diaz, 2004). In CSR, mammalian S regions lie between sets of constant region exons and are enriched in the AGCT sequence, a palindromic example of WRCH. Deamination at both strands of DNA may contribute to double-strand breaks (DSBs) by co-opted repair pathways (Han et al., 2011; Yeap et al., 2015). Joining of AID-initiated DSBs replaces the IgM heavy chain constant region with that of other isotypes to achieve class switching (Matthews et al., 2014a). Patients with AID mutations produce mainly low-affinity IgM antibodies with impairment in CSR and/or SHM, a syndrome known as hyper-IgM immunodeficiency (Xu et al., 2012). AID off-target mutations as well as the subsequent DSBs and chromosomal translocations often promote tumorigenesis, in particular for many types of leukemia and lymphoma (Xu et al., 2012).

The mechanism of AID targeting has been a long-standing mystery. Genome sequencing data showed that AID mutates Ig regions at a 10⁻⁴–10⁻³ per base, per generation frequency (McKean et al., 1984; Rajewsky et al., 1987), which is a million-fold higher than the 10⁻⁹ genomic basal mutation frequency. Numerous models of AID targeting have been proposed, which both propel the field forward and leave remaining questions unanswered. Chromatin accessibility and transcription that provides ssDNA substrates for AID to act is a known prerequisite, but it does not explain AID targeting specificity to limited genomic regions. Convergently transcribed regions in certain super-enhancer loci appear to correlate with enhanced AID off-target mutations (Meng et al., 2014), but it is not clear whether Ig regions undergo convergent transcription, and not
all super-enhancer loci with convergent transcription are AID targets (Qian et al., 2014). The originally proposed hotspot model (Pham et al., 2003) also does not appear sufficient, as ssDNA substrates containing hotspot motifs do not show advantage in recruiting AID in vitro (Larijani et al., 2007), and hotspot motif distribution in the genome does not correlate with the AID off-targeting spectrum in vivo (Duke et al., 2013). In AID recruiter models, various proteins, including replication protein A (RPA) (Chaudhuri et al., 2004), Sp5 (Pavri et al., 2010), and 14-3-3 (Xu et al., 2010), have been proposed in guiding AID to genes. However, the genome distribution of these recruiters is not unique to Ig regions. A more recent recruiter model proposed that transcribed G-repeat rich switch RNAs form G-quadruplex (G4) structures, bind AID, and guide AID to genes for CSR (Zheng et al., 2015). However, the mechanism of transfer of AID between G4 RNA and target DNA has not been addressed.

Here, we present a comprehensive biochemical and structural study on AID, which provides unexpected insights on AID targeting specificity. Through protein engineering, we produced a fully functional monomeric AID with intact activity in vitro and in cells. Various forms of DNA, we found that structured substrates containing multiple ssDNA overhangs, like G4 and branched DNA, are preferred by AID in binding and deamination over linear ssDNA substrates in vitro. This observation may form the basis for the frequent targeting of AID to mammalian Ig switch regions containing high density of G-repeat sequences. We determined the crystal structures of maltose binding protein (MBP)-fused AID and its complex with cytidine (C), deoxycytidine (dC), and deoxycytidine monophosphate (dCMP). These structures not only explain the discrimination between DNA and RNA in AID catalysis but also reveal a bifurcated substrate-binding surface, which strongly supports that one AID recognizes two adjacent ssDNA overhangs from one structured substrate to achieve high affinity. In addition, we observed that G4 structured substrates induce AID cooperative oligomerization, which may promote clustered mutations in Ig S regions. Structure-guided mutagenesis revealed that both the bifurcated substrate binding surface and the putative oligomerization interface are essential for CSR, elucidating recognition of structured substrates as an important AID-targeting mechanism to Ig S regions.

RESULTS

Active AID Monomer from Protein Engineering

Although endogenous AID extracted from B cells exhibited near-monomer molecular mass (Chaudhuri et al., 2003), previously published data (Larijani et al., 2007) and our results showed largely heterogeneous and poorly active aggregates of recombinant wild-type (WT) AID (Figures 1A, 1B, S1A, and S1B). To obtain physiologically relevant recombinant AID for functional elucidation, we performed rounds of protein engineering. We found that the His-MBP-fused double-mutant H130A/R131E produced a low amount of monomer AID, and further N- and C-terminal tail (CTT) truncations improved the monomer yield, resulting in constructs that we named AID.mono+CTT and AID.mono, respectively (Figures 1A, 1B, S1A, and S1B). Both monomeric and aggregated AID fractions were purified and compared in deamination assays, revealing that monomeric AID with and without CTT exhibited much higher deamination activity on ssDNA than aggregated fractions of WT AID, AID.mono+CTT, and AID.mono (Figure 1C). Compared to previously reported AID turnover rates (~10 fmol/min/μg; King et al., 2015), the activities of monomeric AID.mono+CTT and AID.mono were roughly 1,000 fold higher (~106 fmol/min/μg). These data suggest that aggregation renders AID largely inactive. Consistent with the lack of sequence conservation of H130 and R131 among AID species (Figure S1C), the H130A/R131E mutant behaved as WT in mutation frequency by an SHM-mimic rifampicin resistance (RifR) assay (Petersen-Mahrt et al., 2002; Figure 1D) and fully rescued CSR when reconstituted into AID-deficient ex vivo CSR-activated splenic B cells (Figures 1E and S1F). Notably, all AID constructs tested in RifR and CSR assays contain only the internal mutations, with no fusion tag, and no truncations of the nuclear transportation signals at the N or C terminus. Because AID.mono+CTT and AID.mono both showed robust deamination activity and the His-MBP tag did not affect this activity (Figures S1D and S1E), we mainly used the His-MBP-fused AID.mono with higher yield in subsequent biochemical characterizations.

Linear and G4 Structured Substrates from Ig S Regions

Genome sequencing has shown that the mammalian Ig S regions contain abundant tandem G repeats interspersed by AGCT hotspots and are heavily targeted by AID during CSR (Figure S2A; Yu et al., 2003). It has been proposed that, during transcription, G4 structures form on the G-repeat non-template strand and contribute to R-loop stability (Duquette et al., 2004). Consistently, we found that authentic non-template S$^m$ fragment (64 nt) and a single G-repeat substrate both spontaneously assembled into G4 structures (Figures 2A and 2B). The G4 assembly was validated by the fluorescence enhancement of a G4-specific dye and the disruptive effect of LiCl (Figure 2B; Bardin and Leroy, 2008). Gel electrophoresis showed that the 64-nt S$^m$ fragment formed homogeneous oligomers, likely representing mixed inter- and intramolecular G4s (Figure S2B). Differently, the single G-repeat substrate formed homogeneous intermolecular G4 that could be separated from linear ssDNA by size-exclusion chromatography and gel electrophoresis (Figures 2C, 2D, and S2B). Dimethyl sulfate protection footprinting indicated that all Gs in the GGGGTG motif were involved in the G4 assembly (Figure S2C; Sun and Hurley, 2010). Circular dichroism (CD) spectroscopy showed that the G-repeat substrates were predominantly parallel, instead of anti-parallel, G4 structures (Figure S2D; Vorlikova et al., 2012). Notably, the purified linear and G4 fractions of the single G-repeat substrate share the identical DNA sequence and only differ in structure. The same size-exclusion purification procedure was used in preparing other linear and G4 structured substrates containing a single G repeat for in vitro studies.

AID, but Not APOBECs, Prefers Binding and Deaminating G4 Structured Substrates over Linear Substrates

Despite the identical DNA sequence, the purified linear and G4 substrates exhibited distinct behavior in AID binding and deamination assays. The G4 structured substrates displayed ~10-fold
higher AID binding affinity ($K_D = 0.1–0.2 \mu M$) than the linear substrates of the same sequence ($K_D = 1.5–7.1 \mu M$; Figure 3A). The difference is irrespective of the hotspot or the direction of the ssDNA overhang relative to the G-repeat sequence (Figure 3A). By dissecting the G4 substrate structure, we found that AID did not bind to the core structure but rather required at least 5-nt single-stranded overhangs for optimal interaction (Figure 3B). Previously, binding of AID to switch region RNA G4 transcripts has been observed (Zheng et al., 2015). Interestingly, we found that the binding of AID to RNA G4 is equal to DNA G4, with similar affinities and requirement for single-stranded overhangs (Figure 3C). Therefore, AID appears to recognize single-stranded overhangs adjacent to a G4 core, with little dependence on their sequence, orientation, and whether they are DNA or RNA.

Consistently, in vitro deamination assays showed that AID.mono with and without CTT exhibited more robust activity on the G4 structured substrates than the linear substrates, despite the identical DNA sequence (Figures 3D and S3A).
preference remained irrespective whether the substrates contain a hotspot (AGCT) or cold spot (TTCT; Figures 3D and S3A). Remarkably, in a competition assay, linear ssDNA with or without hotspots did not affect G4 structured substrate deamination even at 100-fold molar excess but severely inhibited linear substrate deamination at 10-fold molar excess (Figure 3E). In comparison, two AID homologs APOBEC3A and APOBEC3G did not show a catalytic preference for G4 structured substrate (Figure 3F), suggesting that the G4 structure preference may be a unique feature for AID specific functions, like CSR.

**Location-Dependent Deamination by AID on G4 Structured Substrate**

To probe how AID performs catalysis on G4 structured substrates, we designed a series of G4 substrates with a hotspot (AACT) or a cold spot (TTCT) located at different positions of an ssDNA overhang. We found that the peak deamination activity was achieved when the target deoxycytidine was placed at the third position 3’ to the G4 core (Figures 3G and S3B). AID-mediated deamination steadily decayed when the deoxycytidine was moved away from the third position, albeit still significantly higher than that for the linear substrate (Figure 3G). Peak activities at the third position were similar between hotspot- and cold-spot-containing substrates, suggesting that the G4 structure might override the hotspot preference in AID targeting. In positions away from the third position, deamination activity on hotspot substrates roughly doubled that on cold spot substrates, recapitulating the observed hotspot preference in vitro and in cells (Larijani et al., 2005; Pham et al., 2003). Notably, in Ig S regions, a deoxycytidine often exists exactly at the third position from the G-repeat motif (GGGGTG; Figure S2A), an ideal site for AID deamination, as suggested by our results.

Because of the significant cooperativity observed in binding assays using G4 structured substrates (Hill coefficient n \( \geq 2; \) Figure 3A), we suspected that AID-AID interaction might occur upon G4 binding. The hypothesis was confirmed by reconstituting AID.mono/G4 complex in vitro, which mainly eluted from the void position of a size-exclusion column with an apparent measured molecular mass of \( \sim 1.3 \) MDa (Figure S3C) and was observed as large oligomers under electron microscopy (Figure S3D). To discern the consequence of AID oligomerization, we designed a similar series of G4 substrates with the target
Figure 3. AID Preferentially Binds and Deaminates G4 Structured Substrates

(A) Electrophoretic mobility shift assay (EMSA) curves showing the significantly higher AID binding affinity of G4 fractions than linear fractions with identical primary sequences.

(B) $K_d$ calculated by EMSA showing that ssDNA overhangs in G4 substrates are required for AID binding. Affinities increased with overhang length and plateaued at 5 nt.

(C) EMSA showing that AID binds to RNA G4 similarly as to DNA G4.

(D) In vitro deamination assays showing that AID has a higher deamination activity on G4 substrates than on linear substrates, with or without hotspots. Experiments used 0.1 mM AID and 1 mM DNA.

(E) Competition assays showing that excess linear substrates did not compete with G4 substrates. Experiments used 1 mM AID, 1 mM substrate DNA, and up to 100 mM competitor ssDNA. Reaction time was 10 min.

(F) In vitro deamination assay showing that Apobec3A and 3G did not exhibit G4 preference. Experiments used 0.1 mM APOBEC protein and 1 mM substrate DNA.

(G) In vitro deamination assays showing that the peak activity of AID appears when the substrate nucleotide is at the third position 30 to the G4 core. Experiments used 0.1 mM AID and 1 mM DNA. Reaction time was 10 min.

(H) In vitro deamination assays showing that AID oligomerization causes clustered mutations. Experiments used 0.2 μM DNA and 0.2, 0.4, or 0.8 μM AID mono. Reaction time was 2 min.

Data in (A)–(C), (G), and (H) are represented as mean ± SD from three independent measurements. See also Figure S3.
C located at up to 18 nt away from the G4 core (Figure S3E), which showed a similar peak of deamination at the third position at a low AID concentration (Figure 3H, blue trace). However, when we increased the AID concentration to induce oligomerization, we observed that the cytidine sites away from the G4 core were more efficiently deaminated (Figure 3H, green and yellow traces), suggesting that AID oligomerization on G4 may spread the mutations to more distal sites. The peaks of deamination are separated by ~6 nt (Figure 3H), which is consistent with the 5-nt minimal ssDNA length for AID binding that we identified earlier (Figure 3B).

Structures of AID and Its Complex with Substrates

Because oligomerization of AID.mono upon binding to a G4 substrate resulted in a heterogeneous oligomeric complex, we screened additional mutations on surface hydrophobic resi-
dues that could disrupt AID oligomerization to facilitate crystalli-
ization. We found that the F42E/F141Y/F145E triple mutation plus shortening of the linker to the MBP tag rendered AID.mono entirely monomeric as measured by multi-angle light scattering (MALS) (Figures 1A and S4A). The new construct that we named AID.crystal maintained G4 preference in vitro (Figure S4B) and formed a homogeneous AID4/G4 complex without further oligomerization, as determined by MALS (Figure 4A). Details of the stoichiometry of the complex suggested that each AID is capable of binding two ssDNA overhangs in the G4 substrate. Confirma-
tively, a designed branched substrate with only two ssDNA over-
hangs displayed a 1:1 interaction with AID.crystal (Figure 4A).

Interestingly, AID.mono bound and deaminated better on the two-overhang branched substrate in comparison with a single-
overhang substrate, but the binding did not show cooperativity (Hill coefficient n = 1.0; Figures 4B and S4C) as for G4 substrates (Figure 3A).

The MBP-fused AID.crystal and its catalytically dead mutant E58A were crystallized in complex with G4 DNA or branched DNA but did not crystallize alone. However, despite confirmed presence of DNA in the crystals (Figure S4D), only fragmented DNA density was visible, which appeared to mediate crystal packing. Indeed, AID also co-crystallized with blunt-ended double-stranded DNA (dsDNA) that it did not bind in solution (Figure 4B); the dsDNA stacked in the crystal lattice and likely neutralized repulsion between highly positively charged AID with isoelectric point (pI) of ~9.0 (Figure S4E). Upon trying many different substrates (Table S1), we obtained in total seven structures of AID, alone and in complex with cytidine (C), 2'-deoxycytidine (dC), and 2'-deoxyctydine-5'-monophos-
phate (dCMP) at a highest resolution of 2.4 Å (Figures 4C; Table S2). Importantly, mutations in AID.crystal at residues F42, F141, and F145 all localize on the opposite side of the active site defined by the bound dCMP (Figure 4D).

In contrast to the recently reported crystal structure of an AID-
APOBEC3A hybrid (AIDv) alone (Pham et al., 2016), our AID/ dCMP complex captured using the E58A mutant revealed a deep substrate channel and the direction of ssDNA binding (Figure 4E). Because of the replacement of residues 7–36 of AID with those of APOBEC3A, the AIDv structure exhibits dis-
rupted shape and charge distribution at the substrate channel (Figures 4F and S4F). The active site is comprised of the catalytic proton-donating residue E58 and the Zn2+ ion coordinated by H56, C87, C90, and usually a fourth ligand, e.g., a water in the dCMP complex or a cacodylic acid from an Apo-AID crystalliza-
tion condition (Figures 4G and S4G). No significant conforma-
tional changes were observed at the active site between WT and E58A structures (Figure S4H).

Mechanisms of Cytosine Recognition and DNA/RNA Differentiation

The dCMP-defined substrate channel is mainly formed by the α1-β1, β2-α2, and the β4-α4 loops (Figure 4E), among which the β4-α4 loop was previously designated as the recognition loop important for hotspot specificity (Kohli et al., 2009; Wang et al., 2010). Although there are limited global conformational differ-
ences between structures of AID alone and its complexes, the side chain of F115 in the β4-α4 recognition loop is antiparallel to the conserved Y114 in all complex structures (Figure S4I), whereas a stacking conformation between F115 and Y114 was also observed in Apo-AID structures (Figure S4I), suggesting that the flip of F115 may be induced or stabilized by substrate binding.

The cytosine is cradled by aromatic residues H56, W84, and Y114 (Figure 4G) and precisely positioned by interactions with surrounding residues. The atom N4 forms hydrogen bonds to the Zn2+-coordinating water and the carbonyl oxygen of S85, whereas the atom O2 hydrogen bonds with the hydroxyl of T27 (Figure 4G). If we superimpose the E58-containing Apo-AID structure, N4 also interacts with the side chain of E58 (Figure 4G). In contrast, the protein uracil possesses an O4 instead of an N4 and cannot form the stabilizing hydrogen bonds, consistent with the lack of ligand density in AID co-crystallized with uridine (Table S1). The core arrangement of the AID active site is akin to that of human cytidine deaminase (CDA) despite a different structural fold (Figure S4J), with similar Zn2+ coordination and location of the catalytic Glu (Figure S4K). This structural observation suggests that AID uses a similar deamination mechanism, in which the E58 side chain interacts with N3 of the pyrimidine ring to facilitate nucleophilic attack at C4 by the Zn2+-activated water for deamination (Chaudhuri and Alt, 2004).

Previous data showed that AID does not deaminate RNA substrates (Branstetter et al., 2003). Particularly, replacing the target dC with C on an otherwise ssDNA substrate abol-
ished AID-induced deamination (Nabel et al., 2013). Support-
ing this observation, we only captured significant electron density of dCMP, but not CMP, in the catalytic center (Figure S4L), in spite of similar binding behaviors of AID for DNA and RNA in vitro (Figures 3A and 3C). In the AID/dCMP complex, R25 interacts with the 5'-phosphate and Y114 interacts with the 5'-phosphate, whereas N51 hydrogen bonds with the 3'-OH (Figure 4G). Compared to the Apo-AID structure, both R25 and N51 undergo significant side chain adjustments upon sub-
strate binding (Figure 4H). Proximity of the carbonyl oxygen of R25 to the C2' of the deoxyribose indicates a steric hindrance if the deoxyribose is replaced by a ribose (Figures 4G and 4H). Without the 5'-phosphate, the AID structure in complex with C or dC either showed different sugar position or much weaker density (Figures S4L and S4M), suggesting that the 5'-phosphate is essential for fixing dCMP orientation in the catalytic center for DNA/RNA differentiation. The interactions
Figure 4. Structures of AID and Its Complex with dCMP

(A) Gel filtration chromatography with in-line MALS showing that AID.crystal binds G4 DNA in 2:1 ratio and branched DNA in 1:1 ratio. Measured and calculated molecular masses are labeled.

(B) EMSA curves showing enhanced AID binding affinity for branched substrate with two overhangs (red), in comparison to that with one overhang (linear substrate, black) or no overhang (dsDNA, blue). Data are represented as mean ± SD from three independent measurements.

(C) Ribbon diagram of human AID in rainbow color showing the secondary structures and catalytic residues near active site Zn²⁺. W, water.

(D) Locations of F42, F141, and F145 mutated in AID.crystal on the face of the crystal structure opposite to the active site.

(E) Surface charge distribution of the AID/dCMP structure and pocket prediction revealed a substrate binding channel that passes through the active site (green mesh).

(F) Comparison between the AID-APOBEC3A hybrid AIDv (PDB: 5JJ4) and AID.crystal showed distinct surface charge distribution at the substrate channel.

(G) Substrate dCMP in AID (E58A) catalytic center showing the interactions with surrounding residues. The E58 side chain is taken from the WT Apo-AID structure.

(H) Alignment between Apo- and dCMP-bound AID structures showing the movement of R25 and N51 upon substrate recognition.

(I) Mutations associated with the hyper-IgM syndrome mapped to AID structure.

See also Figure S4 and Tables S1–S3.
we observed in the complex structure explain many hyper-IgM syndrome mutations (Figure 4I; Table S3) and previously reported disruptive mutations, such as R25A/D, T27A, N51A, and Y114A (Basu et al., 2005; King et al., 2015; Shivarov et al., 2008).

**A Bifurcated Substrate-Binding Surface Explains G4 Preferences and Is Critical for CSR**

As described earlier, our in vitro data suggested that one AID interacts with two ssDNA overhangs (Figure 4A). Supporting this observation, we found that, apart from the substrate channel, the AID structure contains an additional positively charged surface at helix α6, which we named the “assistant patch” (Figure 5A). Together with the substrate channel, the AID structure suggested a bifurcated substrate-binding surface wedged by a negatively charged β4-α4 loop (Figure 5A). This structural feature is surprisingly analogous to branched nucleic acid recognition by the T4 RNase H (Devos et al., 2007) and Cas9 (Jiang et al., 2016; Figures S5A and S5B). Thus, we propose a simultaneous recognition mechanism for two overhangs in structured substrates, such as G4 in AID targeting, in which one ssDNA overhang passes through the active site and an adjacent ssDNA binds at the assistant patch to enhance affinity. Sequence alignment shows that the positively charged residues in the bifurcated substrate-binding surface are highly conserved in AID across species, but this conservation is absent in APOBEC homologs (Figure 5B), explaining the lack of G4 preference in APOBEC3A and APOBEC3G (Figure 3F).

To validate the bifurcated substrate-binding surface, we mutated either the substrate channel or the assistant patch. For the substrate channel, the triple mutation on K22, R24, and R25 and the double mutation on R50 and R52 abolished AID deamination activity on DNA substrates with either one or two ssDNA overhangs (Figure 5C). As a control, a triple mutation on residues K34, R77, and R107, which are localized far away from the substrate-binding surface, did not affect AID activity (Figure 5C). For the assistant patch, mutations on R171, R174, R177, and R178 specifically compromised AID deamination activity on substrate with two ssDNA overhangs, without significantly altering AID activity on one ssDNA overhang substrate (Figure 5D), confirming its assistant role in recognizing structured substrates with multiple overhangs.

Remarkably, when reconstituted into AID-deficient splenic B cells, the assistant patch mutants showed completely
abolished CSR activity (<1% of the WT; Figures 5E and S5C). Among the mutants, R174S was previously reported from patients with hyper-IgM syndrome (Durandy et al., 2006; Honjo et al., 2012; Table S3). Moreover, the AID-APOBEC3A hybrid

AIDv (Pham et al., 2016) was also completely deficient in conducting CSR (Figures 5E and S1A), likely due to the disrupted substrate channel orientation (Figures 4F and S4F). Of note, all AID constructs tested in the CSR assay were with intact N-terminal nuclear localization sequence (NLS) and CTT, supporting that the functional deficiencies were solely contributed by the internal mutations. Collectively, we demonstrate that the integrity of the bifurcated substrate-binding surface, including both the substrate channel and the assistant patch, is essential for AID function in CSR.

**Putative AID Oligomerization on G4 Contributes to CSR**

Compared to AID.mono, the three additional mutations in AID.crystal (F42/F141/F145) localize on the opposite side of the proposed substrate-binding surface (Figure 4D). Interestingly, these mutations impaired CSR by ~60% in comparison to the WT without affecting AID deamination activity in vitro and in the SHM-mimic RifR assay (Figures 1D, 1E, and S4B), suggesting that CSR may require more AID properties than SHM. DNA binding assay showed that the oligomerization-deficient AID.crystal exhibited no cooperativity (Hill coefficient n = 1.0; Figure S5D) or formed large oligomers (Figure S3D) upon G4 binding, in comparison to the highly cooperative AID.mono (Figure 3A). Additionally, AID.crystal exhibited much faster dissociation from G4 structured substrates than
WT-like AID.mono as measured by bio-layer interferometry, irrespective of the DNA sequence (Figures 6A and 6B). Based on these data, we hypothesize that AID oligomerization upon G4 structure binding contributes to CSR by enhancing AID recruitment and accumulation to Ig S regions. It should be noted that AID.mono only cooperatively oligomerizes on G4, but not on a simpler structured substrate, such as branched DNA, that may be formed by local secondary structures (Figure 4B). The AID oligomerization may be further facilitated by the CTT (Mondal et al., 2016) and result in previously observed processivity (Pham et al., 2003).

### Comparison of Substrate Recognition by AID and APOBECs

Two recent papers reported crystal structures of APOBECs in complex with ssDNAs, including those of the non-catalytic APOBEC3G N-domain (A3G-N) (Xiao et al., 2016), APOBEC3A (A3A), and APOBEC3B containing a replaced z1-β1 loop from A3A (A3B-A chimera; Shi et al., 2017). Consistent with its lack of an active site, A3G-N interacts with ssDNA at a surface shifted from the analogous AID active site (Figure S6A). The A3A and A3B-A chimera structures displayed the same binding mode to U-shaped ssDNA (Figure 6C), among which the cognate C at the center binds similarly as dCMP in AID (Figure 6D). However, the remaining ssDNA clashes with the AID surface and does not follow the observed substrate channel of AID (Figures 6D and S6B).

To further validate the observed substrate channel in AID, we modeled the binding of an AGCTT hotspot substrate using the bound dCMP position as an anchor (Figure 6E). We found that the β4-z4 recognition loop stayed in the conformation observed in the dCMP complex structure (Figure S4I) and that conformation rearrangement is only required at the z1-β1 loop to accommodate the substrate at the −1 and −2 positions (Figures 6E and S6C). No gross conformational changes are needed for the superficial binding at +1 and +2 positions (Figure 6E). The modeling exercise therefore supports the observed substrate channel. We suspect that, if a U-shaped substrate channel exists in AID as in A3A and A3B, it will likely not be able to support simultaneous binding of another ssDNA at the assistant patch, which may be the reason for the difference between substrate recognition by AID and APOBECs (Figure 6F) and for the CSR deficiency of AIDv (Figure 5E).

### DISCUSSION

In this study, we combined biochemical, biophysical, and structural approaches to elucidate AID targeting mechanisms, especially in CSR. By using the fully functional AID.mono, we clearly demonstrated that G4 substrates mimicking the Ig S regions are preferred AID targets in vitro. Different from the previously proposed hotspot hypothesis, our data indicate that the AID preference for G4 substrates is predominantly due to their bundled ssDNA overhangs structure rather than any primary sequence motif. By solving structures of the MBP-fused AID.crystal alone and in complex with dCMP, we observed a bifurcated substrate-binding surface, which strongly supports a model of one AID recognizing two adjacent ssDNA overhangs in a structured substrate, such as G4, to achieve high affinity (Figure 7). Although AID can recruit either DNA or RNA in a
sequence-independent manner, the positioning of dCMP in the active site suggests that only deoxycytidine, but not cytidine, can flip into the catalytic center to be deaminated.

Other than Ig S regions, it has been shown that G4 structures may be prevalent in other regions of mammalian genomes (Chambers et al., 2015; Maizels and Gray, 2013). Especially, G-rich regions in AID off-target genes, like c-MYC and BCL6, have been proposed to be targeted by AID (Duquette et al., 2005, 2007). After inspecting many identified AID off-target genes, we found that these genes often contain G-repeat (GGG) enriched regions, particularly in their non-template strands, which may lead to G4 assembly during transcription and direct AID recruitment (Figure S6D; Table S4). However, more experimental data and genome sequence analysis may be required to establish the correlation between G-repeat motif distribution and AID targeting.

Additionally, our data suggest that G4 DNA-binding-induced cooperative AID oligomerization may contribute to its accumulation in Ig S regions, promoting high-density mutations, DSBs, and their joining during CSR (Figure 7). In comparison, SHM can be induced by CSR-deficient AID variants in cells (Hwang et al., 2015; Pham et al., 2016) and by APOBEC3 homologs during retroviral infection (Halemano et al., 2014), suggesting that mutations in Ig V regions may not require all the AID features employed in CSR. Because chromatin immunoprecipitation sequencing data showed that Ig V regions may not bind AID as stably as S regions (Matthews et al., 2014b), and the assistant patch mutant R174S also causes SHM deficiency (Durandy et al., 2006; Honjo et al., 2012), we speculate that Ig V regions may only transiently recruit AID using local secondary structures like branched DNA, which do not induce AID oligomerization and stable association.

Although our biochemical data and the bipartite DNA binding surface in AID structure strongly suggest a model of how AID recognizes structured substrates (Figure 6F), a definitive complex structure that contains fully characterized substrate conformation is still lacking, despite extensive efforts using both crystallography and electron microscopy (EM). Co-crystallization of purified AID/G4, AID/branched DNA, and AID/linear DNA complexes did not yield visible density for the substrates, likely due to crystal packing effects. Alternatively, we tried co-crystallization using 1- to 5-nt ssDNA fragments together with a 12-bp dsDNA (Table S1), but any ssDNA fragment longer than 1 nt did not reveal any density in the active site. Neither did soaking of various substrates into pre-formed AID crystals (Table S1). We further used EM on the AID/G4 complex, but the flexibility of the complex caused issues in the reconstruction (Figure S6E). Thus, the exact molecular basis for our model remains to be determined by future structural studies of AID/DNA complexes. Due to the close correlation of AID activity with B cell lymphoma and other types of cancers, the AID structures will also provide templates for potential therapeutic intervention against this important cytidine deaminase.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING

**METHOD DETAILS**

- Protein Engineering and Purification
- Ex Vivo CSR Assay
- Rifampicin Resistance (Rif\(^{R}\)) Assay
- In Vitro Deamination Assay
- In Vitro EMSA
- G4 and Branched DNA Purification and Characterization
- AID/DNA Complex Crystallization and Structure Determination
- Molecular Mass Measurement by Multi-Angle Light Scattering (MALS)
- Electron Microscopy

**QUANTIFICATION AND STATISTICAL ANALYSES**

- In Vitro EMSA and Deamination Assay Quantification
- Data Analysis in Flow Cytometry

**DATA AND SOFTWARE AVAILABILITY**

- Accession Numbers

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes six figures and four tables can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2017.06.034.

**AUTHOR CONTRIBUTIONS**

H.W. supervised the project. Q.G. and L.W. designed the experiments and performed protein engineering, in vitro assays, structure determination, and analysis. F.-L.M. performed ex vivo CSR assay, and F.W.A. supervised the effort. H.W. and Q.G. wrote most of the manuscript with help from J.K.H. and F.W.A. for the introduction. All authors contributed to data analysis and critical interpretation of results and approved the manuscript.

**ACKNOWLEDGMENTS**

We thank Ermelinda Damko and Devendra Srivastava for their earlier work on this project; Ming Tian, Zhou Du, Leng Swee Yap, Jiayi Hu, and Junchao Dong for discussions; Yang Li for EM data analysis; Xia Xie for assistances in CSR assay; Rida Mourtada in Dr. Loren D. Walensky’s lab and Kelly Arnett in Harvard Medical School Center for Macromolecular Interactions for their assistance with CD spectroscopy; Sukumar Narayanasami and Surajit Banerjee of NE-CAT at the Advance Photon Source for their assistance on data collection; and Maria Ericsson and Louise Trakimas of Harvard Medical School EM facility for assistance on EM imaging. This work was supported by a Cancer Research Institute Irvington Postdoctoral Fellowship (to Q.G.), a Lymphoma Research Foundation Fellowship (to F.-L.M.), NIH R01 AI077595 (to F.W.A.), and NIH F30 AI114179 (to J.K.H.). F.W.A. is an Investigator of the Howard Hughes Medical Institute.

Received: March 1, 2017

Revised: May 3, 2017

Accepted: June 27, 2017

Published: July 27, 2017

**REFERENCES**


### STAR★METHODS

#### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APC Rat anti-Mouse IgG1</td>
<td>BD Biosciences</td>
<td>560089; RRID: AB_1645625</td>
</tr>
<tr>
<td>Anti-AID</td>
<td>Chaudhuri et al., 2003</td>
<td>N/A</td>
</tr>
<tr>
<td>Anti-GFP</td>
<td>MBL</td>
<td>Code # 598; RRID: AB_10597267</td>
</tr>
<tr>
<td>Anti-CD40</td>
<td>eBioscience</td>
<td>16-0402-86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Bacterial and Virus Strains</strong></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E.coli BL21 (DE3)</td>
<td>Agilent Technologies</td>
<td>200131</td>
</tr>
<tr>
<td>E. coli KL16</td>
<td>Coli Genetic Stock Center</td>
<td>CGSC# 4245</td>
</tr>
<tr>
<td>E. coli BW310</td>
<td>Coli Genetic Stock Center</td>
<td>CGSC# 6078</td>
</tr>
<tr>
<td>GIBCO Sf9 cells</td>
<td>Thermo Scientific</td>
<td>11496-015</td>
</tr>
<tr>
<td>DH10Bac</td>
<td>Thermo Scientific</td>
<td>10361-012</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Chemicals, Peptides, and Recombinant Proteins</strong></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SYBR Gold Nucleic Acid Gel Stain</td>
<td>Thermo Scientific</td>
<td>S-11494</td>
</tr>
<tr>
<td>40% Acrylamide/Bis Solution, 19:1</td>
<td>Bio-Rad</td>
<td>161-0144</td>
</tr>
<tr>
<td>SIGMAFAST Protease Inhibitor Cocktail Tablets, EDTA-Free</td>
<td>Sigma-Aldrich</td>
<td>S8830-20TAB</td>
</tr>
<tr>
<td>Rifampicin</td>
<td>Sigma-Aldrich</td>
<td>R3501-250MG</td>
</tr>
<tr>
<td>Dimethyl sulfate</td>
<td>Sigma-Aldrich</td>
<td>D186309-100ML</td>
</tr>
<tr>
<td>Piperidine</td>
<td>Sigma-Aldrich</td>
<td>411027-100ML</td>
</tr>
<tr>
<td>Celfectin II Reagent</td>
<td>Invitrogen</td>
<td>Cat#10362-100</td>
</tr>
<tr>
<td>Recombinant Mouse Interleukin-4/IL-4</td>
<td>Novoprotein</td>
<td>Cat#CK15</td>
</tr>
<tr>
<td>Lipopolysaccharides from Escherichia coliO111:B4</td>
<td>Sigma Aldrich</td>
<td>L2630-100MG</td>
</tr>
<tr>
<td>Uracil-DNA Glycosylase</td>
<td>NEB</td>
<td>M0280S</td>
</tr>
<tr>
<td>2′-deoxycytidine</td>
<td>Sigma-Aldrich</td>
<td>D3897-500MG</td>
</tr>
<tr>
<td>2′-deoxycytidine 5′-monophosphate</td>
<td>Sigma-Aldrich</td>
<td>D7625-100MG</td>
</tr>
<tr>
<td>Cytidine</td>
<td>Sigma-Aldrich</td>
<td>C4654-1G</td>
</tr>
<tr>
<td>Cytidine 5′-monophosphate</td>
<td>Sigma-Aldrich</td>
<td>C1006-1G</td>
</tr>
<tr>
<td>Uranyl formate</td>
<td>Electron Microscopy Sciences</td>
<td>22450</td>
</tr>
<tr>
<td>PreScission protease</td>
<td>GE Healthcare</td>
<td>27-0843-01</td>
</tr>
<tr>
<td>LB Broth</td>
<td>RPI</td>
<td>L24066-5000.0</td>
</tr>
<tr>
<td>HeClone SFX-Insect Media</td>
<td>Thermo Scientific</td>
<td>SH30278LS</td>
</tr>
<tr>
<td>Grace’s Insect Medium, Supplemented</td>
<td>Thermo Scientific</td>
<td>11605-102</td>
</tr>
<tr>
<td>Fetal Bovine Serum</td>
<td>Sigma-Aldrich</td>
<td>TMS-013-B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Critical Commercial Assays</strong></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EasySep Mouse B Cell Isolation Kit</td>
<td>STEMCELL Technologies</td>
<td>Cat#19854</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Deposited Data</strong></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AID.crystal, co-crystallized with G4 DNA</td>
<td>Protein Data Bank</td>
<td>PDB: 5W02Z</td>
</tr>
<tr>
<td>AID.crystal E58A, co-crystallized with Branched DNA and cacodylic acid</td>
<td>Protein Data Bank</td>
<td>PDB: 5W0R</td>
</tr>
<tr>
<td>AID.crystal E58A, co-crystallized with dsDNA and cytidine</td>
<td>Protein Data Bank</td>
<td>PDB: 5W1C</td>
</tr>
<tr>
<td>AID.crystal E58A, co-crystallized with dsDNA and dCMP</td>
<td>Protein Data Bank</td>
<td>PDB: 5W0U</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Recombinant DNA</strong></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEX6P1</td>
<td>GE Healthcare</td>
<td>Cat#23954648</td>
</tr>
<tr>
<td>pTrc99A</td>
<td>DNA Resource Core</td>
<td>N/A</td>
</tr>
<tr>
<td>pFastBac</td>
<td>Thermo Scientific</td>
<td>10584-027</td>
</tr>
<tr>
<td>All synthesized DNA/RNA</td>
<td>Integrated DNA Technologies</td>
<td>N/A</td>
</tr>
</tbody>
</table>

(Continued on next page)
CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Hao Wu (wu@crystal.harvard.edu).

METHOD DETAILS

Protein Engineering and Purification
Constructs of human AID (UniProt: Q9GZX7) were generated in the pFastBac vector (Thermo Fisher Scientific) and expressed in Sf9 insect cells for 48 hr using recombinant baculoviruses. To overcome AID aggregation in recombinant expression, we created numerous AID constructs and found that adding the maltose binding protein (MBP) solubility tag, modifying the N-terminal nuclear localization sequence (NLS) and removing the C-terminal tail (CTT) resulted in improved yield, but retained AID aggregation. By screening non-conserved AID surface residue mutations, we identified the AID mutant H130A/R131E. We named this construct AID.mono because it contained a distinct monomeric fraction in addition to the aggregated faction, in contrast to complete aggregation of wild-type (WT) AID. When necessary, the MBP-tag was removed by incubating with the PreScission Protease (GE Healthcare) at 1/100 ratio at 4°C overnight. Adding C-terminal tail back (AID.mono+CTT) decreased the monomer yield by 95%. In AID.crystal, the linker between AID and MBP was shortened to facilitate crystallization. The proteins were affinity-purified using amylose resin (New England Biolabs), followed by chromatography using Superdex 200 10/300 GL, monomer fraction was further purified by HiTrap Heparin HP (GE Healthcare) and another Superdex 200. Final gel filtration buffer contains 20 mM Bis-Tris at pH 6.8, 200 mM NaCl and 1 mM tris(2-carboxyethyl)phosphine (TCEP).

Human Apobec3A (1-199) and APOBEC3G (197-380) was cloned into a pGEX6p-1 vector and expressed in Bl21 (DE3) strain. The N-GST fusion proteins were purified by Glutathione Sepharose 4B resin (GE Healthcare), and the tag was removed by on-column Precision Protease treatment. APOBEC proteins were further purified over a Superdex 200.

Ex Vivo CSR Assay
The AID constructs used in this assay contained point mutations in AID.mono (H130A/R131E), AID.crystal (H130A/R131E plus F42E/F141Y/F145E), assistant patch residues, or AIDv. The ex vivo CSR assay was performed as described previously (Cheng et al., 2009).
Briefly, AID-deficient mouse splenic B cells were stimulated with IL-4 (Novoprotein) and anti-CD40 (eBioscience) to induce CSR to IgG1. One day after stimulation, WT or mutant AID together with GFP via an internal ribosome entry site (IRES) was retrovirally delivered into the cells. The percentage of GFP-positive cells that underwent switching to IgG1 was taken as the level of CSR rescue. The data for each mutant were either from 3 or 6 mice.

**Rifampicin Resistance (Rif³) Assay**

WT and mutant AID were cloned into the pTrc99A vector (pTrc99A-AID) and the rifampicin resistance (Rif³) assay was performed as previously described (Wang et al., 2009). Briefly, *E. coli* strain KL16 (Hfr [PO-45] relA1 spoT1 thi-1) and its UDG-deficient derivative BW310 bacteria transformed with pTrc99A-AID plasmids were grown overnight to saturation in LB medium supplemented with ampicillin (100 μg/ml) and isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM), and plated on LB low-salt agar containing ampicillin (100 μg/ml) and rifampicin (50 μg/ml). Mutation frequency was measured by determining the median number of colony-forming cells that survived selection per 10⁴ viable plated from 12 independent cultures. The identity of mutations was determined by sequencing the relevant section of rpoB (typically from 25 to 200 individual colonies) after PCR amplification using oligonucleotides 5’-TTGGCGAAATGGCGGAAAACC-3’ and 5’-CACCAGCAGGATACCCCTGCTG-3’ (synthesized by Integrated DNA Technologies).

**In Vitro Deamination Assay**

Each 10 μl reaction contained 0.1 or 1 μM AID, and 1 μM substrate DNA labeled by 6-carboxyfluorescein (FAM) at 5’ end (synthesized by Integrated DNA Technologies) in 20 mM HEPES at pH 7.5, 100 mM KCl and 1 mM DTT. Following incubation at 37°C for the indicated length of time, each reaction was raised to 95°C for 10 min and flash cooled to inactivate AID and resolve DNA structures. Sufficient amount of UDG (New England Biolabs, 5 unit with each unit catalyzing 60 pmol/min at 37°C) was then added and the mixture was incubated at 37°C for 1 h. Lastly, NaOH was added to a final concentration of 0.15 M, and treated at 95°C for 15 min to break abasic sites. Urea gel was used to separate the product from the substrate. Experiments in Figures 5G and 5H used SYBR Gold staining to save cost. Images were taken on Image Scanner FLA-9000 (Fujifilm) using excitation wavelength of 495 nm and emission wavelength of 519 nm. Quantification was performed using the Image Lab (Bio-Rad) and Multi Gauge (Fujifilm) software.

**In Vitro EMSA**

Up to 25 μM AID was titrated into 10 nM FAM labeled DNA followed by 10%-12% acrylamide:biacrylamide (19:1) native gel to separate free DNA and the AID/DNA complex. Quantified free DNA amounts at different AID concentrations were used to calculate the dissociation constant (Kd) and the Hill coefficient (n) of the AID/DNA interaction. 100 nM unlabeled TGGGGT₁₅ and TGGGGT₁₀ were used in experiments of Figure 3B to eliminate any effect from the fluorophore and its linker to DNA. SYBR Gold was used for staining. Images were taken on Image Scanner FLA-9000 (Fujifilm). The Multi Gauge (Fujifilm) software was used for quantification and the Origin software (OriginLab Corporation) was used in curve fitting.

**G4 and Branched DNA Purification and Characterization**

To generate mostly G4 structures, S4 fragment and G-repeat DNAs were dissolved in 20 mM HEPES at pH 7.5, 100 mM KCl and 1 mM DTT, incubated at 95°C for 5 min and slowly cooled down to room temperature. In contrast, to generate mostly linear DNA, heat denaturation at 95°C for 10 min followed by flash cooling on ice was used to eliminate DNA structures. A Jasco J-815 Circular dichroism (CD) spectropolarimeter was used in determining the parallel and anti-parallel conformation of the G4 assembly (Vorlickova et al., 2012). When preparing large amounts of G4 structures, up to 1M KCl was used. The size exclusion column Superdex 75 or 200 (GE Healthcare) was used to separate G4 and linear DNA fractions. Notably, the FAM fluorescence intensity was quenched.

**AID/DNA Complex Crystallization and Structure Determination**

Gel filtration purified AID.crystal (WT and E58A)/branched composites containing DNA with the sequences of GTTTCAAGGCCAGCTTTT and TTTTTCTCGGCTTGTAAC were crystallized using 0.05 M sodium cacodylate at pH 5.5, 20 mM MgCl₂, 10 mM CaCl₂, 10 mM spermidine and 5% PEG3350 at 16°C.

When using dsDNA as a crystallization chaperone, AID.crystal (E58A) was mixed with dsDNA with the sequences of GTTCAAGGCGAAGCTTTT and TTTTTCTCGGCTTGTAAC were crystallized using 0.05 M sodium cacodylate at pH 5.5, 20 mM MgCl₂, 10 mM CaCl₂, 10 mM spermidine and 5% PEG3350 at 16°C.

When using dsDNA as a crystallization chaperone, AID.crystal (E58A) was mixed with dsDNA with the sequences of GTTCAAGGCGAAGCTTTT and TTTTTCTCGGCTTGTAAC were crystallized using 0.05 M sodium cacodylate at pH 5.5, 20 mM MgCl₂, 10 mM CaCl₂, 10 mM spermidine and 5% PEG3350 at 16°C. Many other short oligos were also tried in co-crystallization, but did not show ordered nucleotide density in the structures (Table S1).

The AID/G4 structure was solved by molecular replacement calculations in Phaser using the crystal structure of MBP (PDB: 3VOW) (Kitamura et al., 2012). Subsequent structures were determined by molecular replacement using the AID structure as a model. Phenix, CCP4 and Coot were used in model building and refinement (Adams et al., 2017).
et al., 2010; Emsley et al., 2010; Winn et al., 2011). The POCASA 1.1 server was used for pocket prediction (Yu et al., 2010). The YASARA server was used for energy minimization of the substrate manually docked to AID (YASARA Biosciences). Pymol was utilized for molecular visualization and structure display (Schrödinger).

Molecular Mass Measurement by Multi-Angle Light Scattering (MALS)
AID alone and AID-DNA complexes were reconstituted and purified as described above. The complex peak fractions containing ~0.2 mg protein were loaded onto a Superdex 200 gel filtration column coupled to a three-angle light scattering detector (mini-DAWN TRISTAR) and a refractive index detector (Optilab DSP) (Wyatt Technology). Data analysis was carried out using ASTRA V.

Electron Microscopy
AID and AID/DNA complex were applied to carbon-coated grids and negatively stained with 1% uranyl formate (Electron Microscopy Sciences). Samples were examined in a Tecnai G2 Spirit BioTWIN electron microscope (FEI) at an accelerating voltage of 80 keV and a nominal magnification of × 49,000.

QUANTIFICATION AND STATISTICAL ANALYSES

In Vitro EMSA and Deamination Assay Quantification
Images were obtained by instruments described in Method Details. The band areas were boxed and backgrounds were subtracted. The band quantification results were averaged from three repeats and the error bars were shown as standard deviations.

Data Analysis in Flow Cytometry
Cell population was identified in FlowJo. Unpaired t test was used and two-tailed P value was calculated using GraphPad Prism6. The class switch recombination (CSR) results showed average from three biological replicates, with bars indicated SD.

DATA AND SOFTWARE AVAILABILITY

Accession Numbers
The accession numbers for the data reported in this paper are PDB: 5W0Z (AID.crystal, co-crystallized with G4 DNA), 5W0R (AID.crystal E58A, co-crystallized with Branched DNA and cacodylic acid), 5W1C (AID.crystal E58A, co-crystallized with dsDNA and cytidine), and 5W0U (AID.crystal E58A, co-crystallized with dsDNA and dCMP).