

Supplementary Data

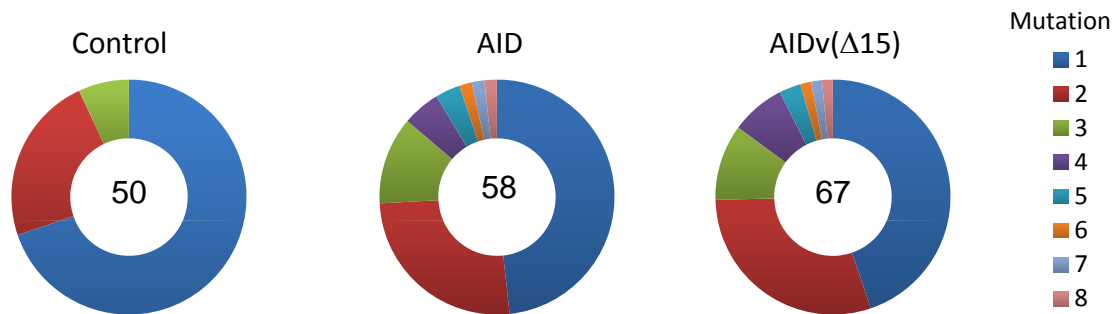
Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification

Phuong Pham^{1,#}, Samir A. Afif¹, Mayuko Shimoda^{2,3,4}, Kazuhiko Maeda^{3,4}, Nobuo Sakaguchi^{4,5}, Lars C. Pedersen^{6,#}, & Myron F. Goodman^{1,7*}

a

	Transfected Ramos B-cells		
	Control	AID	AIDv(Δ 15)
Number of sequenced IgV _H clones	50	58	67
Number of substitution mutations	16	64	75
Mutation frequency (x 10 ⁻³ per bp)	1.0	3.0	3.3
Transitions/Transversions (%)	63/38	52/48	63/37
Mutations at A/T sites (%)	19	18	15
Mutations at C/G sites (%)	81	83	85

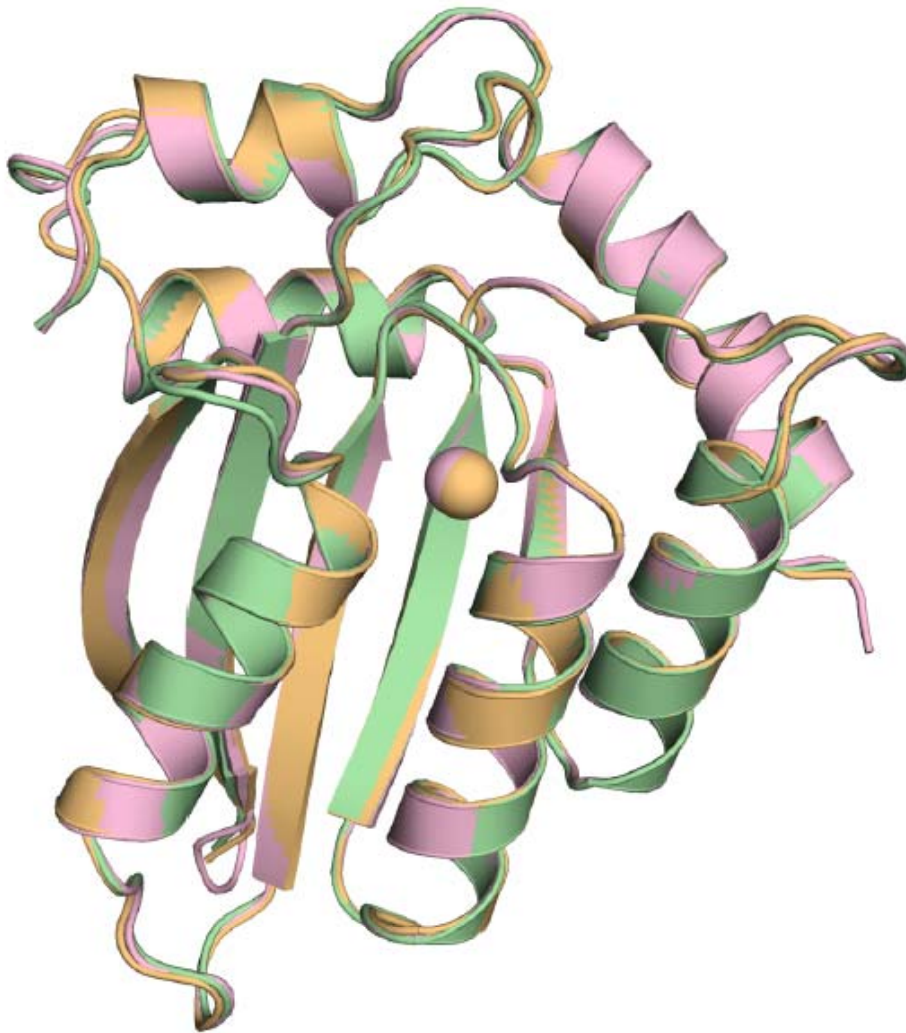
b



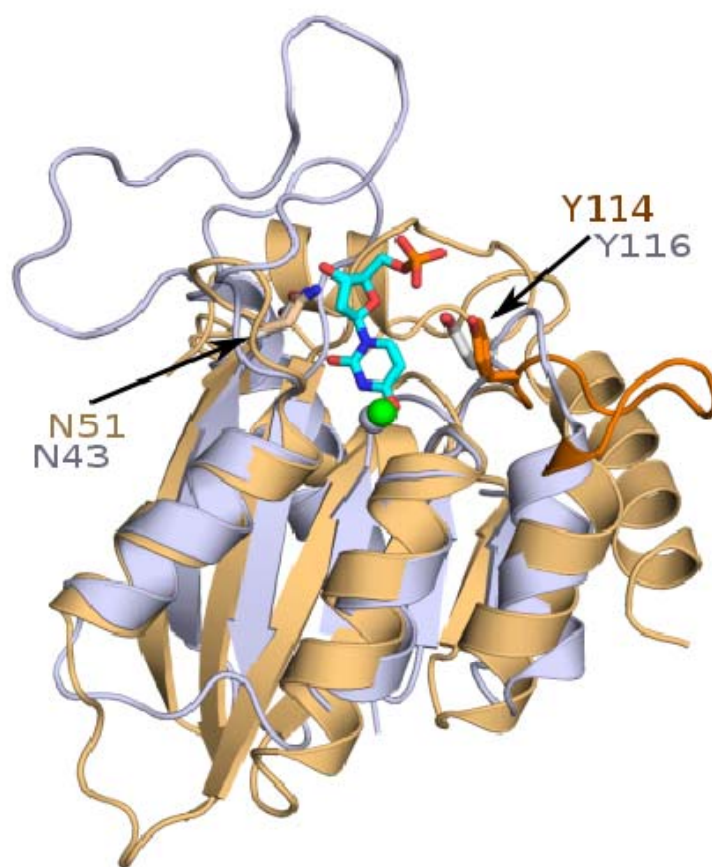
c

Control						AID						AIDv(Δ 15)					
To/From	A	C	G	T	Total	To/From	A	C	G	T	Total	To/From	A	C	G	T	Total
A		0	1	1	13%	A		3	3	2	13%	A		0	4	2	8%
C	0		1	4	31%	C	3		2	14	30%	C	1		6	18	33%
G	4	4		0	50%	G	14	15		5	53%	G	22	13		4	52%
T	0	1	0		6%	T	0	2	1		5%	T	0	3	2		7%

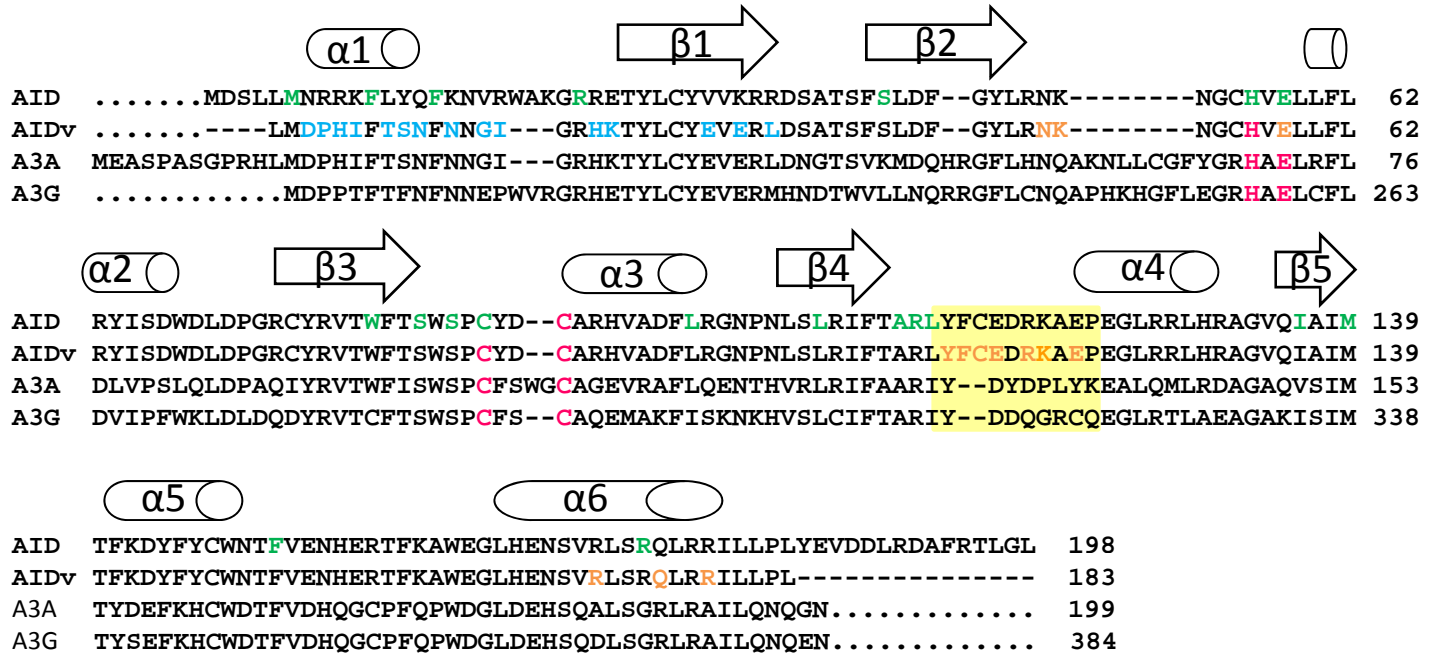
Supplementary Figure 1. AIDv(Δ 15)-induced IgV_H SHM in Ramos B-cells. **a)** Analysis of IgV_H SHM mutations induced by AID and AIDv(Δ 15). Retroviruses, carrying AID or AIDv(Δ 15) along with the mock control were used to infect Ramos B-cells. After culturing for 2 weeks, IgV_H region DNA from infected B-cells was subjected to DNA sequencing analysis. **b)** Distribution of sequenced IgV_H clones. The number in the center of the pie chart indicates the number of IgV_H-regions analyzed. Colored sectors indicate the distribution of clones with a given number of base substitution mutations from 0 to 8. **c)** Mutation matrices showing the types and numbers of mutations in the IgV_H-region.



Supplementary Figure 2. Superpositions of the three AID molecules in the crystal asymmetric unit. All molecules are very similar with mol A (green) and mol B (pink) having RMSD of 0.37 and 0.27 over 171 C α s respectively to mol C (peach).

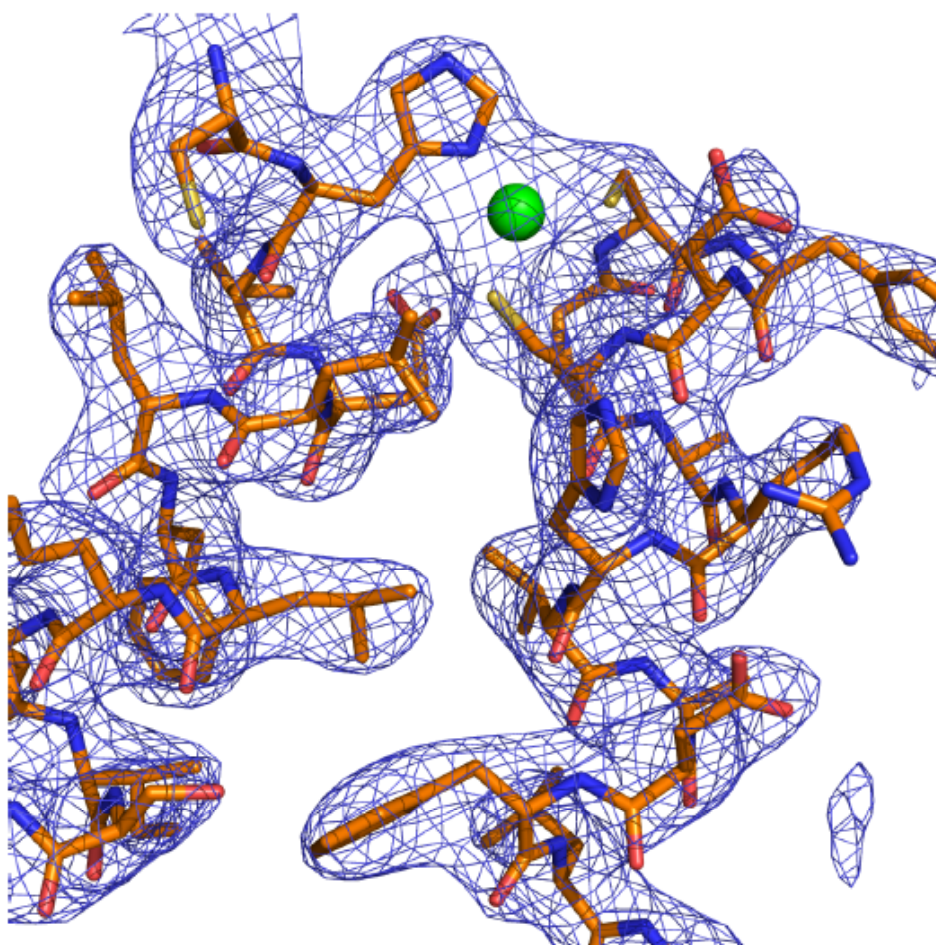


Supplementary Figure 3. Superposition of the crystal structure of deoxycytidylate deaminase from bacteriophage S-TIM5 (grey, pdb 4P9C) with AID (peach, specificity loop orange). Superposition is based on residues 67 to 73 and 93 to 100 of S-TIM5 and 56 to 62 and 85 to 92 of AID (RMSD 0.40 on Cas). dUMP S-TIM5 is colored cyan. The Zn atoms from AID and S-TIM5 are colored green and grey respectively. Sidechains N43 and Y116 from S-TIM5 (grey labels) that position the dUMP are shown along with the corresponding residues N51 and Y114 from AID (peach and orange labels).



Blue AID solubility mutations
 Green HIGM2 mutation sites on AID
 Orange site-directed mutations sites in this study
 Magenta Conserved active site residues
 Light peach background specificity loop

Supplementary Figure 4. Amino-acid sequence alignment of AID, AIDv($\Delta 15$), Apo3A and Apo3G



Supplementary Figure 5. 2Fo-Fc electron density contoured at 1.0 σ of helices 2 and 3 in the crystal structure of AIDv(Δ 15).

Supplementary Table 1. Data collection and refinement statistics for MBP-AIDv(Δ 15)

	Low Res	High Res
Data collection		
Space group	P32	P2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	111.27, 111.27, 200.04	106.79, 96.81, 108.69
α , β , γ (°)	90.0, 90.0, 120	90.0, 102.87, 90.0
Resolution (Å)	50.00-3.10 (3.15-3.10) *	50.00-2.80 (2.85-2.80)
<i>R</i> _{sym} or <i>R</i> _{merge}	0.05 (0.62)	0.07 (0.77)
<i>I</i> / σI	11.6 (2.5)	10.3 (2.0)
Completeness (%)	99.2 (99.0)	99.8 (100.0)
Redundancy	4.1 (4.1)	3.8 (3.8)
Refinement		
Resolution (Å)		50.0-2.8
No. reflections		52657
<i>R</i> _{work} / <i>R</i> _{free}		0.18/0.23
No. atoms		12734
Protein		12560
Ligand/ion		132/6
Water		36
<i>B</i> -factors		
Protein		
(molA/molB/molC)		74.7/82.0/66.1
Ligand/ion		56.7/90.4
Water		55.7
R.m.s. deviations		
Bond lengths (Å)		0.008
Bond angles (°)		0.817

*Values in parentheses are for highest-resolution shell. Both data sets were from single crystals

Supplementary Table 2. AID, AID_v($\Delta 15$), AID_v, Apo3A and Apo3G mutability indexes

Motif (5'-3')	Mutability Index (MI)*				
	AID	AID _v ($\Delta 15$)	AID _v	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 \pm 0.4	1.1 \pm 0.4	1.6 \pm 0.8	0.2 \pm 0.2	0.1 \pm 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 \pm 0.4	1.0 \pm 0.4	0.7 \pm 0.3	2.1 \pm 1.6	3.2 \pm 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 \pm 0.3	0.9 \pm 0.4	0.8 \pm 0.3	0.8 \pm 1.0	0.4 \pm 0.6