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*}

	Transfected Ramos B-cells			
	Control	AID	AIDv(∆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





	А	lDv	(Δ1	5)	
To	А	С	G	Т	Total
А		0	4	2	8%
С	1		6	18	33%
G	22	13		4	52%
Т	0	3	2		7%

Mutation

Control То А С G Т Total From 0 А 1 1 13% С 1 0 4 31% G 4 4 0 50% 0 Т 0 1 6%

AID To From С А G Т Total 2 3 А 3 13% 3 С 2 14 30% G 14 15 5 53% Т 0 2 1 5%

Supplementary Figure 1. AIDv(Δ15)-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.

С

а



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 C α s). 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).

	$(\alpha 1)$ $\beta 1$ $\beta 2$ $()$	
AID	MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFL	62
AIDv	LMDPHIFTSNFNNGIGRHKTYLCYEVERLDSATSFSLDFGYLRNKNGCHVELLFL	62
A3A	${\tt MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRHAELRFL}{}$	76
A3G	MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFL	263

	$(\alpha 2)$		$(\alpha 3)$	LB4	(α4	TR2>	
AID	RYISDWDLDP	GRCYRVTWFTSWSPC	YD <mark>C</mark> ARHVADFLRGI	NPNLSLRIFTARI	YFCEDRKAEP	EGLRRLHRA	GVQIAIM	139
AIDv	RYISDWDLDP	GRCYRVTWFTSWSPC	YDCARHVADFLRGI	NPNLSLRIFTARI	YFCEDRKAEP	EGLRRLHRA	GVQIAIM	139

A3A DLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIY--DYDPLYKEALQMLRDAGAQVSIM 153

A3G DVIPFWKLDLDQDYRVTCFTSWSPCFS--CAQEMAKFISKNKHVSLCIFTARIY--DDQGRCQEGLRTLAEAGAKISIM 338

-		
(α 5 ()	
	us (
<u> </u>	\sim	



AID	TFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL	198
AIDv	TFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPL	183
A3A	TYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN	199
A3G	TYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN	384

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(∆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 AlDv(Δ 15).

	Low Res	High Res
Data collection		
Space group	P32	P2 ₁
Cell dimensions		
a, b, c (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)
$R_{\rm sym}$ or $R_{\rm merge}$	0.05 (0.62)	0.07 (0.77)
Ι/σΙ	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
$R_{\rm work}$ / $R_{\rm free}$		0.18/0.23
No. atoms		12734
Protein		12560
Ligand/ion		132/6
Water		36
B-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 parenthese	es are for highest-resolution s	hell Both data sets were from single

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

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

	Mutability Index (MI)*				
Motif (5'-3')	AID	AIDv(∆15)	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
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

Supplementary Table 2. AID, AIDv(Δ 15), AIDv, Apo3A and Apo3G mutability indexes