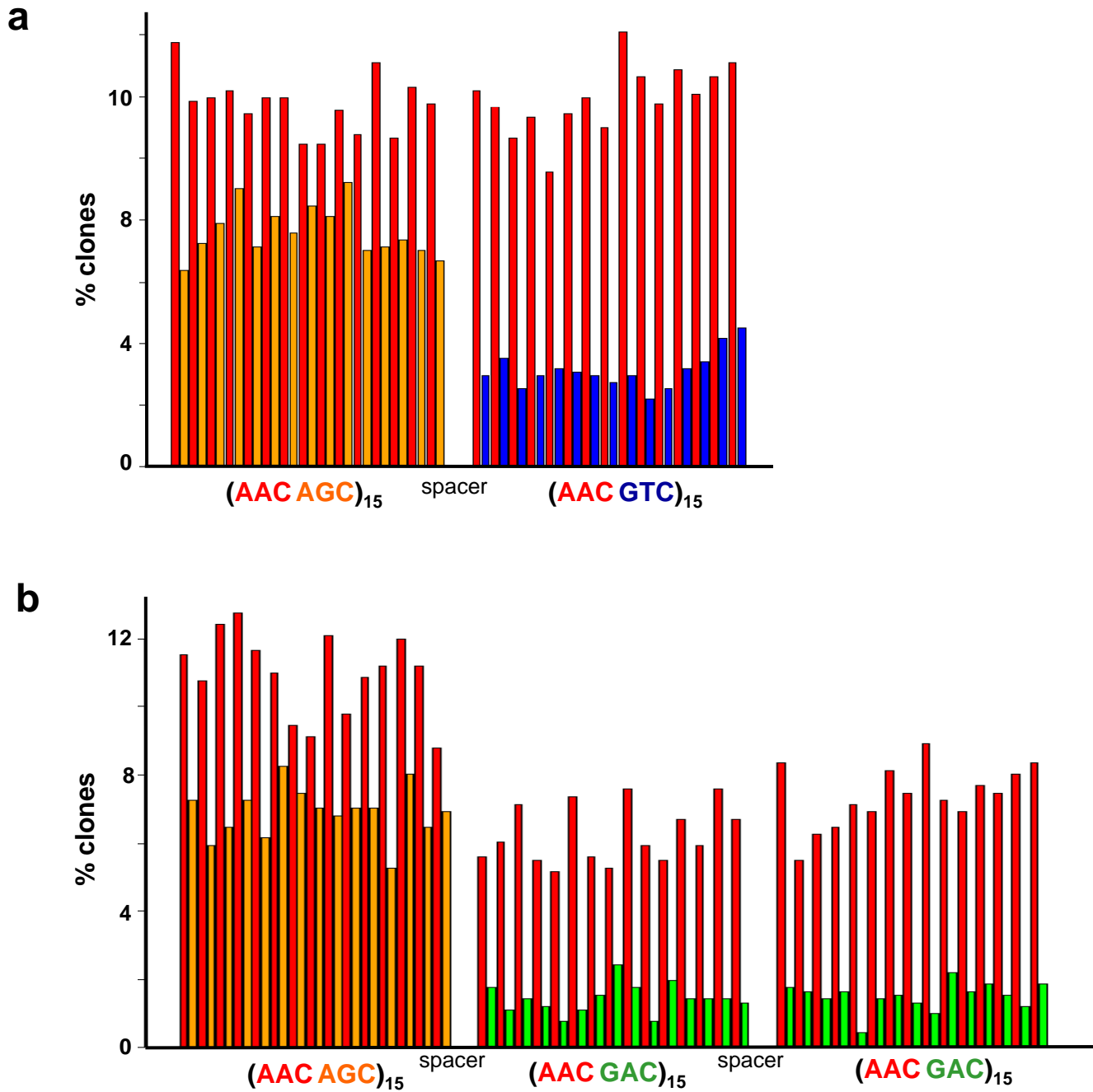


Supplemental Data

An analysis of a single-stranded DNA scanning process in which AID deaminates C to U haphazardly and inefficiently to ensure mutational diversity

Phuong Pham, Peter Calabrese, Soo Jung Park, and Myron F. Goodman

Departments of Biological Sciences and Chemistry, Molecular and Computational Biology Section,
University of Southern California, Los Angeles, California 90089-2910, USA



Supplemental Figure S1. Simulations of the modified random walk model based on 6 nt deamination motifs. **a**, Simulated mutation spectrum for $(\text{AAC AGC})_{15}$ - $(\text{AAC GTC})_{15}$ construct. **b**, Simulated mutation spectrum for $(\text{AAC AGC})_{15}$ - $(\text{AAC GAC})_{15}$ - $(\text{AAC GAC})_{15}$ construct. Parameters used for simulation are: Average sliding distance 30 nt, motif deamination probabilities: AGCAAC 3%, AACAGC 1.7%, GTCAAC 3%, AACGTC 0.75%, GACAAAC 1.5%, AACGAC 0.3%. C is the deamination target site.

Supplemental Table S1. Comparison of cluster patterns in the (AAC AGC)₁₅ cassette in the absence and presence of salt (150 mM NaCl)

Mutations/clone	Cluster	Mutation clusters (%)										
		1	2	3	4	5	6	7	8	9	10	
2	No salt (209)*	92.8	7.2									
2	150 mM NaCl (58)	91.6	8.4									
3	No salt (113)	87.6	11.4	1.0								
3	150 mM NaCl (26)	81.5	16.9	1.5								
4	No salt (64)	77.0	19.0	3.0	1.0							
4	150 mM NaCl (35)	74.3	20.0	3.8	1.9							
5	No salt (55)	71.1	23.9	2.5	2.0	0.5						
5	150 mM NaCl (10)	76.3	15.8	7.9	0.0	0.0						
6	No salt (37)	67.3	24.4	7.1	1.3	0.0	0.0					
6	150 mM NaCl (13)	76.7	18.3	3.3	1.7	0.0	0.0					
7	No salt (27)	62.1	25.0	11.3	1.6	0.0	0.0	0.0				
7	150 mM NaCl (6)	69.2	7.7	19.2	0.0	3.8	0.0	0.0				
8	No salt (19)	57.4	29.8	7.4	4.3	1.1	0.0	0.0	0.0			
8	150 mM NaCl (12)	58.2	27.3	3.6	3.6	7.3	0.0	0.0	0.0			
9	No salt (16)	48.7	28.9	11.8	6.6	2.6	1.3	0.0	0.0	0.0		
9	150 mM NaCl (3)	58.3	0.0	8.3	25.0	8.3	0.0	0.0	0.0	0.0		
10	No salt (12)	49.2	25.4	11.9	3.4	6.8	3.4	0.0	0.0	0.0	0.0	
10	150 mM NaCl (4)	63.6	18.2	9.1	0.0	4.5	0.0	4.5	0.0	0.0	0.0	

* Numbers in parentheses are the number of sequenced clones. Some of the rows show larger discrepancies than in **Tables 1** and **2** in the manuscript, which can be attributed to the small number of clones.

Supplemental Table S2. Distribution of distances between mutated motifs in the (AAC AGC)₁₅ cassette in the absence and presence of salt (150 mM NaCl)

		Distribution of distance between mutated motifs (%)								
Mutations/clone		0	1	2	3	4	5 to 10	11 to 15	16 to 20	21 to 29
2	No salt (209)*	2.4	11.0	5.7	9.1	4.8	27.8	19.1	12.4	7.7
2	150 mM NaCl (58)	8.6	6.9	3.4	6.9	10.3	25.9	19.0	10.3	8.6
3	No salt (113)	8.8	8.8	5.3	9.3	8.4	31.4	20.4	5.8	1.8
3	150 mM NaCl (26)	15.4	9.6	9.6	5.8	3.8	36.5	9.6	7.7	1.9
4	No salt (64)	14.6	14.6	6.8	8.9	6.3	31.8	16.1	1.0	0.0
4	150 mM NaCl (35)	16.2	17.1	7.6	10.5	9.5	28.6	8.6	1.9	0.0
5	No salt (55)	15.9	17.7	9.5	12.7	6.4	25.9	10.5	1.4	0.0
5	150 mM NaCl (10)	15.0	15.0	7.5	2.5	7.5	42.5	10.0	0.0	0.0
6	No salt (37)	15.1	20.5	8.6	11.4	16.2	23.8	3.8	0.5	0.0
6	150 mM NaCl (13)	15.4	12.3	12.3	18.5	7.7	30.8	3.1	0.0	0.0
7	No salt (27)	16.7	23.5	13.6	8.6	7.4	27.8	2.5	0.0	0.0
7	150 mM NaCl (6)	33.3	11.1	11.1	16.7	8.3	16.7	2.8	0.0	0.0
8	No salt (19)	21.8	21.8	11.3	12.0	12.0	20.3	0.8	0.0	0.0
8	150 mM NaCl (12)	16.7	32.1	7.1	15.5	11.9	16.7	0.0	0.0	0.0
9	No salt (16)	25.8	27.3	12.5	9.4	5.5	19.5	0.0	0.0	0.0
9	150 mM NaCl (3)	12.5	50.0	0.0	12.5	4.2	20.8	0.0	0.0	0.0
10	No salt (12)	30.6	25.9	12.0	13.9	5.6	12.0	0.0	0.0	0.0
10	150 mM NaCl (4)	22.2	27.8	27.8	13.9	2.8	5.6	0.0	0.0	0.0

* Numbers in parentheses are the number of sequenced clones. Some of the rows show larger discrepancies than in **Tables 1** and **2** in the manuscript, which can be attributed to the small number of clones.

Supplemental Table S3. Comparison of cluster patterns in three cassettes (AAC AGC)₁₅, (AAC GAC)₁₅ and (AAC GTC)₁₅

Mutations/clone	Cluster	Mutation clusters (%)										
		1	2	3	4	5	6	7	8	9	10	
2	(AAC AGC) ₁₅ (209)*	92.8	7.2									
2	(AAC GAC) ₁₅ (57)	94.4	5.6									
2	(AAC GTC) ₁₅ (30)	92.9	7.1									
3	(AAC AGC) ₁₅ (113)	87.6	11.4	1.0								
3	(AAC GAC) ₁₅ (23)	82.8	15.5	1.7								
3	(AAC GTC) ₁₅ (26)	85.1	13.4	1.5								
4	(AAC AGC) ₁₅ (64)	77.0	19.0	3.0	1.0							
4	(AAC GAC) ₁₅ (9)	75.0	21.4	3.6	0.0							
4	(AAC GTC) ₁₅ (15)	77.1	20.8	2.1	0.0							
5	(AAC AGC) ₁₅ (55)	71.1	23.9	2.5	2.0	0.5						
5	(AAC GAC) ₁₅ (5)	58.8	35.3	5.9	0.0	0.0						
5	(AAC GTC) ₁₅ (8)	50.0	36.4	4.5	0.0	9.1						
6	(AAC AGC) ₁₅ (37)	67.3	24.4	7.1	1.3	0.0	0.0					
6	(AAC GAC) ₁₅ (2)	66.7	33.3	0.0	0.0	0.0	0.0					
6	(AAC GTC) ₁₅ (9)	61.8	20.6	14.7	2.9	0.0	0.0					
7	(AAC AGC) ₁₅ (27)	62.1	25.0	11.3	1.6	0.0	0.0	0.0				
7	(AAC GAC) ₁₅ (3)	69.2	0.0	30.8	0.0	0.0	0.0	0.0				
7	(AAC GTC) ₁₅ (5)	45.0	40.0	10.0	5.0	0.0	0.0	0.0				
8	(AAC AGC) ₁₅ (19)	57.4	29.8	7.4	4.3	1.1	0.0	0.0	0.0			
8	(AAC GAC) ₁₅ (1)	25.0	50.0	25.0	0.0	0.0	0.0	0.0	0.0			
8	(AAC GTC) ₁₅ (9)	40.0	31.4	17.1	5.7	5.7	0.0	0.0	0.0			
9	(AAC AGC) ₁₅ (16)	48.7	28.9	11.8	6.6	2.6	1.3	0.0	0.0	0.0		
9	(AAC GAC) ₁₅ (0)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		
9	(AAC GTC) ₁₅ (8)	52.8	22.2	2.8	16.7	5.6	0.0	0.0	0.0	0.0		
10	(AAC AGC) ₁₅ (12)	49.2	25.4	11.9	3.4	6.8	3.4	0.0	0.0	0.0	0.0	
10	(AAC GAC) ₁₅ (2)	62.5	12.5	0.0	12.5	0.0	0.0	0.0	0.0	0.0	12.5	0.0
10	(AAC GTC) ₁₅ (3)	30.8	38.5	7.7	15.4	7.7	0.0	0.0	0.0	0.0	0.0	0.0

* Numbers in parentheses are the number of sequenced clones. Some of the rows show larger discrepancies than in **Tables 1** and **2** in the manuscript, which can be attributed to the small number of clones.

Supplemental Table S4. Distribution of distances between mutated motifs in three cassettes (AAC AGC)₁₅, (AAC GAC)₁₅ and (AAC GTC)₁₅

		Distribution of distance between mutated motifs (%)								
Mutations/clone		0	1	2	3	4	5 to 10	11 to 15	16 to 20	21 to 28
2	(AAC AGC) ₁₅ (209)*	2.4	11.0	5.7	9.1	4.8	27.8	19.1	12.4	7.7
2	(AAC GAC) ₁₅ (57)	1.8	8.8	3.5	5.3	5.3	36.8	19.3	12.3	7.0
2	(AAC GTC) ₁₅ (30)	3.3	10.0	0.0	16.7	0.0	36.7	20.0	10.0	3.3
3	(AAC AGC) ₁₅ (113)	8.8	8.8	5.3	9.3	8.4	31.4	20.4	5.8	1.8
3	(AAC GAC) ₁₅ (23)	6.5	17.4	4.3	23.9	4.3	28.3	10.9	4.3	0.0
3	(AAC GTC) ₁₅ (26)	3.8	17.3	1.9	11.5	0.0	42.3	17.3	5.8	0.0
4	(AAC AGC) ₁₅ (64)	14.6	14.6	6.8	8.9	6.3	31.8	16.1	1.0	0.0
4	(AAC GAC) ₁₅ (9)	18.5	11.1	7.4	11.1	3.7	33.3	11.1	3.7	0.0
4	(AAC GTC) ₁₅ (15)	8.9	17.8	2.2	13.3	8.9	42.2	6.7	0.0	0.0
5	(AAC AGC) ₁₅ (55)	15.9	17.7	9.5	12.7	6.4	25.9	10.5	1.4	0.0
5	(AAC GAC) ₁₅ (5)	10.0	30.0	0.0	30.0	0.0	25.0	0.0	0.0	5.0
5	(AAC GTC) ₁₅ (8)	12.5	43.8	3.1	15.6	3.1	18.8	3.1	0.0	0.0
6	(AAC AGC) ₁₅ (37)	15.1	20.5	8.6	11.4	16.2	23.8	3.8	0.5	0.0
6	(AAC GAC) ₁₅ (2)	0.0	30.0	10.0	30.0	0.0	20.0	10.0	0.0	0.0
6	(AAC GTC) ₁₅ (9)	15.6	28.9	6.7	17.8	0.0	26.7	4.4	0.0	0.0
7	(AAC AGC) ₁₅ (27)	16.7	23.5	13.6	8.6	7.4	27.8	2.5	0.0	0.0
7	(AAC GAC) ₁₅ (3)	11.1	33.3	5.6	27.8	5.6	16.7	0.0	0.0	0.0
7	(AAC GTC) ₁₅ (5)	6.7	43.3	6.7	16.7	6.7	20.0	0.0	0.0	0.0
8	(AAC AGC) ₁₅ (19)	21.8	21.8	11.3	12.0	12.0	20.3	0.8	0.0	0.0
8	(AAC GAC) ₁₅ (1)	14.3	42.9	14.3	28.6	0.0	0.0	0.0	0.0	0.0
8	(AAC GTC) ₁₅ (9)	12.7	46.0	4.8	22.2	1.6	12.7	0.0	0.0	0.0
9	(AAC AGC) ₁₅ (16)	25.8	27.3	12.5	9.4	5.5	19.5	0.0	0.0	0.0
9	(AAC GAC) ₁₅ (0)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	(AAC GTC) ₁₅ (8)	15.6	40.6	9.4	18.8	9.4	6.3	0.0	0.0	0.0
10	(AAC AGC) ₁₅ (12)	30.6	25.9	12.0	13.9	5.6	12.0	0.0	0.0	0.0
10	(AAC GAC) ₁₅ (2)	27.8	38.9	5.6	16.7	0.0	5.6	5.6	0.0	0.0
10	(AAC GTC) ₁₅ (3)	29.6	33.3	11.1	11.1	7.4	7.4	0.0	0.0	0.0

* Numbers in parentheses are the number of sequenced clones. Some of the rows show larger discrepancies than in **Tables 1** and **2** in the manuscript, which can be attributed to the small number of clones.