Supporting information

Structure-based design of first-generation small molecule inhibitors targeting the catalytic pockets of AID, APOBEC3A and APOBEC3B

Justin J. King[1,2*#,] Faezeh Borzooee[1,2#], Junbum Im[2], Mahdi Asgharpour[1,2], Atefeh Ghorbani[1,2], Cody P. Diamond[2], Heather Fifield[2], Lesley Berghuis[2], Mani Larijani[1,2*]

[1]Department of Molecular Biology and Biochemistry, Faculty of Science, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.

[2]Program in immunology and Infectious Diseases, Division of Biomedical Sciences, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland A1B 3V6, Canada.

*To whom correspondence should be addressed: Email: Mani_Larijani@sfu.ca, or Justin_King@sfu.ca

[#]These authors made equal contributions

Table of contents

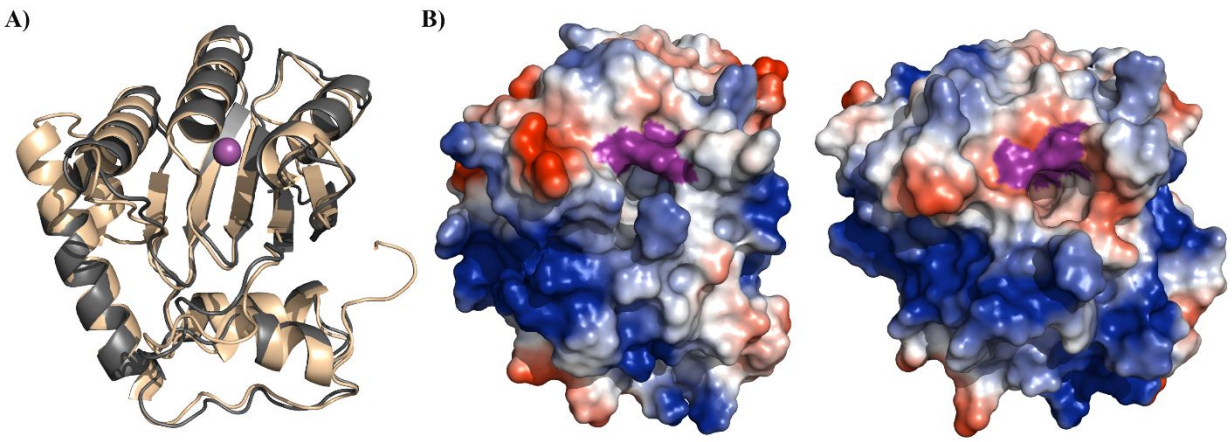

Figure S1

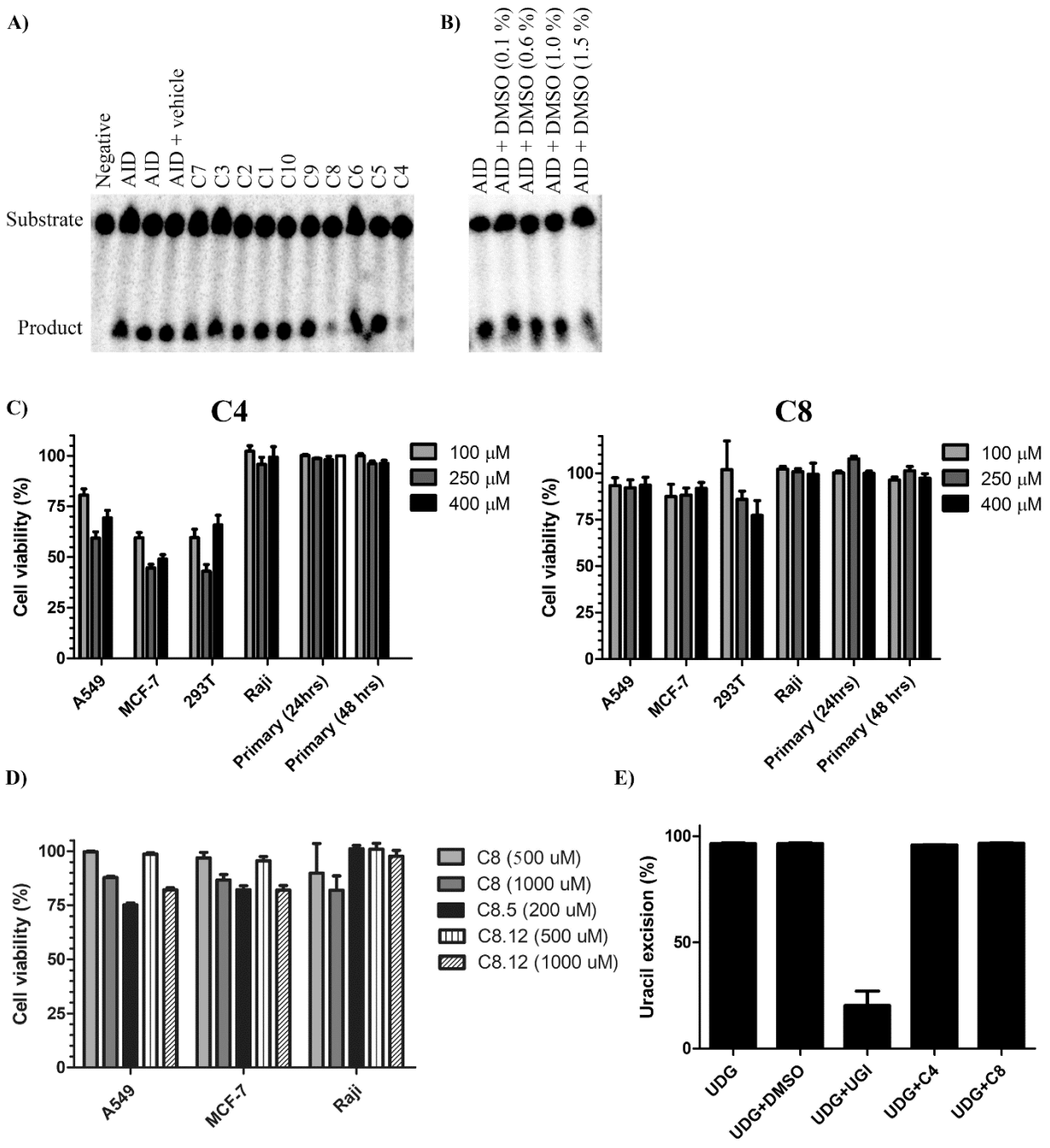
A)
Negative
AID
AID
AID + vehicle
C7
C3
C2
C1
C10
C9
C8
C6
C5
C4
Substrate
Product
B)
AID
AID + DMSO (0.1 %)
AID + DMSO (0.6 %)
AID + DMSO (1.0 %)
AID + DMSO (1.5 %)
C)
C4
C8
Cell viability (%)
100 μM
250 μM
400 μM
A549
MCF-7
293T
Raji
Primary (24hrs)
Primary (48 hrs)
D)
C8 (500 uM)
C8 (1000 uM)
C8.5 (200 uM)
C8.12 (500 uM)
C8.12 (1000 uM)
E)
Uracil excision (%)
UDG
UDG+DMSO
UDG+UGI
UDG+C4
UDG+C8

Figure S2

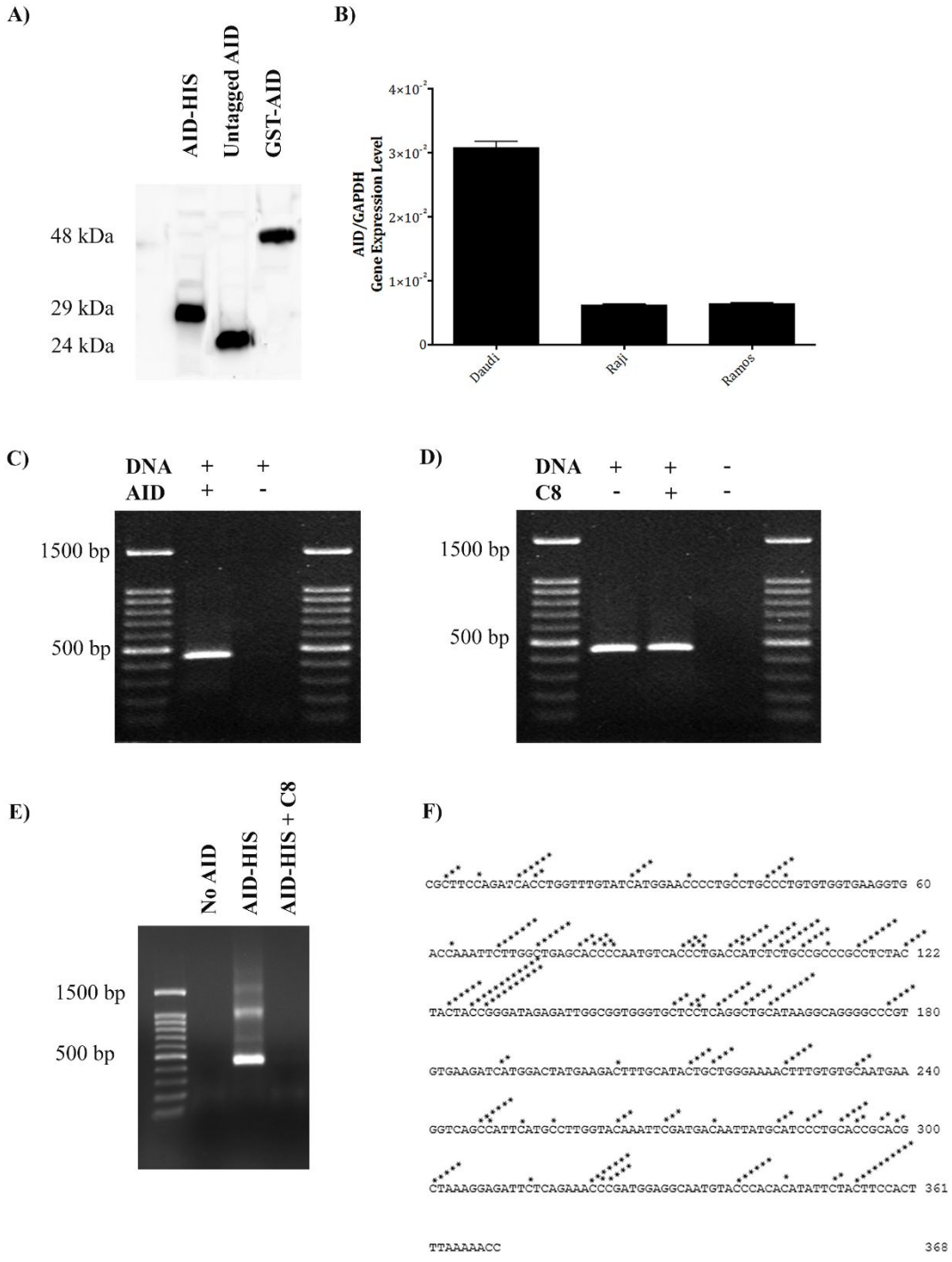
A)
AID-HIS
Untagged AID
GST-AID
48 kDa
29 kDa
24 kDa
B)
AID/GAPDH
Gene Expression Level
4×10⁻²
3×10⁻²
2×10⁻²
1×10⁻²
0
Daudi
Raji
Ramos
C)
DNA + +
AID + -
1500 bp
500 bp
D)
DNA + + -
C8 - + -
1500 bp
500 bp
E)
No AID
AID-HIS
AID-HIS + C8
1500 bp
500 bp
F)
CGCTTCCAGATCACCTGGTTTGTATCATGGAACCCCTGCCTGCCCTGTGTGGTGAAGGTG 60
ACCAAATTCTTGGCTGAGCACCCCAATGTCACCCTGACCATCTCTGCCGCCCGCCTCTAC 122
TACTACCGGGATAGAGATTGGCGGTGGGTGCTCCTCAGGCTGCATAAGGCAGGGGCCCGT 180
GTGAAGATCATGGACTATGAAGACTTTGCATACTGCTGGGAAAACTTTGTGTGCAATGAA 240
GGTCAGCCATTCATGCCTTGGTACAAATTCGATGACAATTATGCATCCCTGCACCGCACG 300
CTAAAGGAGATTCTCAGAAACCCGATGGAGGCAATGTACCCACACATATTCTACTTCCACT 361
TTAAAAACC 368

Figure S3

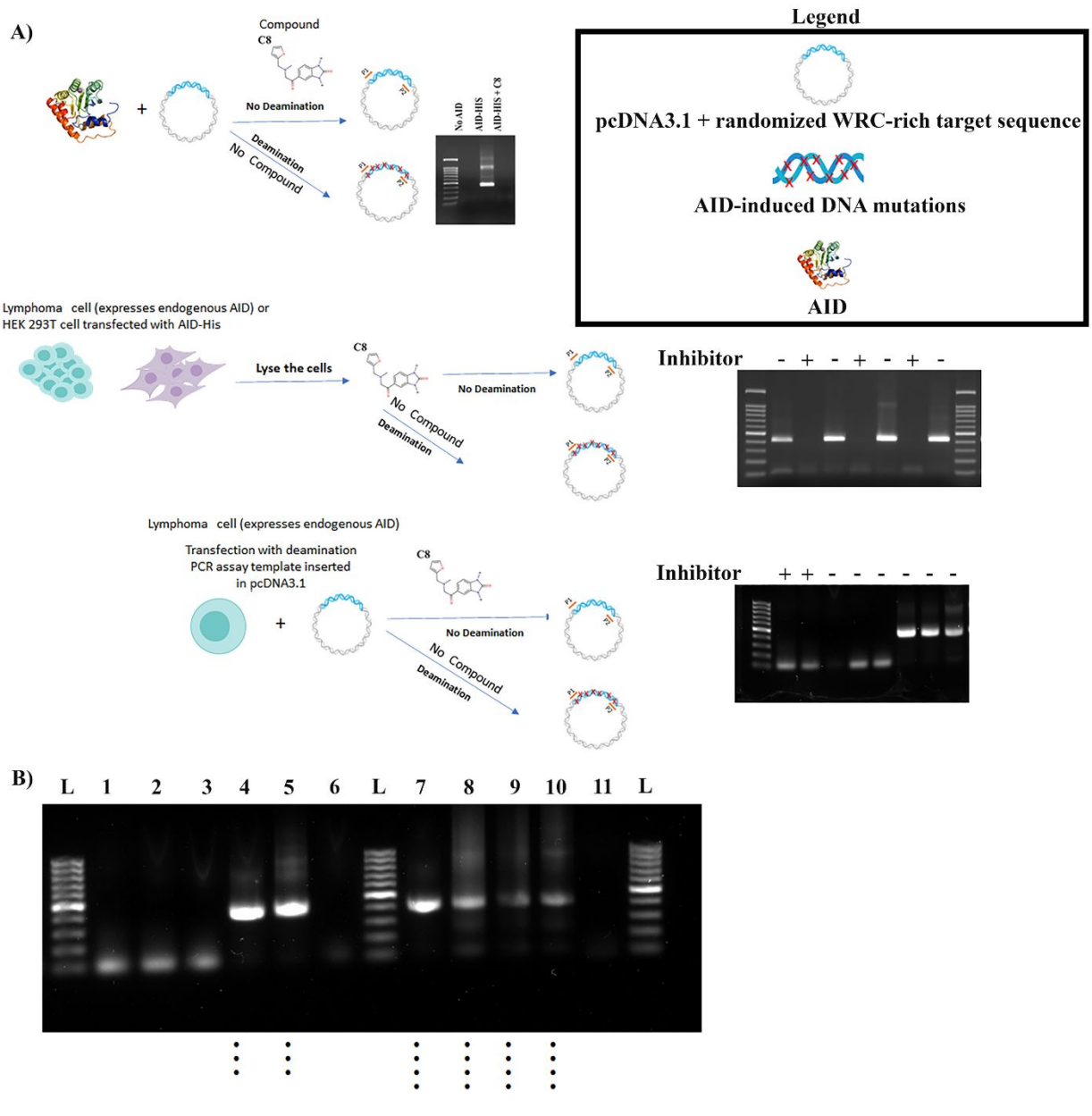
A)
Compound
C8
No Deamination
Deamination
No Compound
No AID
AID-HIS
AID-HIS + C8
Legend
pcDNA3.1 + randomized WRC-rich target sequence
AID-induced DNA mutations
AID
Lymphoma cell (expresses endogenous AID) or
HEK 293T cell transfected with AID-His
Lyse the cells
C8
No Deamination
No Compound
Deamination
Inhibitor - + - + - + -
Lymphoma cell (expresses endogenous AID)
Transfection with deamination
PCR assay template inserted
in pcDNA3.1
C8
No Deamination
No Compound
Deamination
Inhibitor + + - - - - - -
B)
L 1 2 3 4 5 6 L 7 8 9 10 11 L

Figure S4

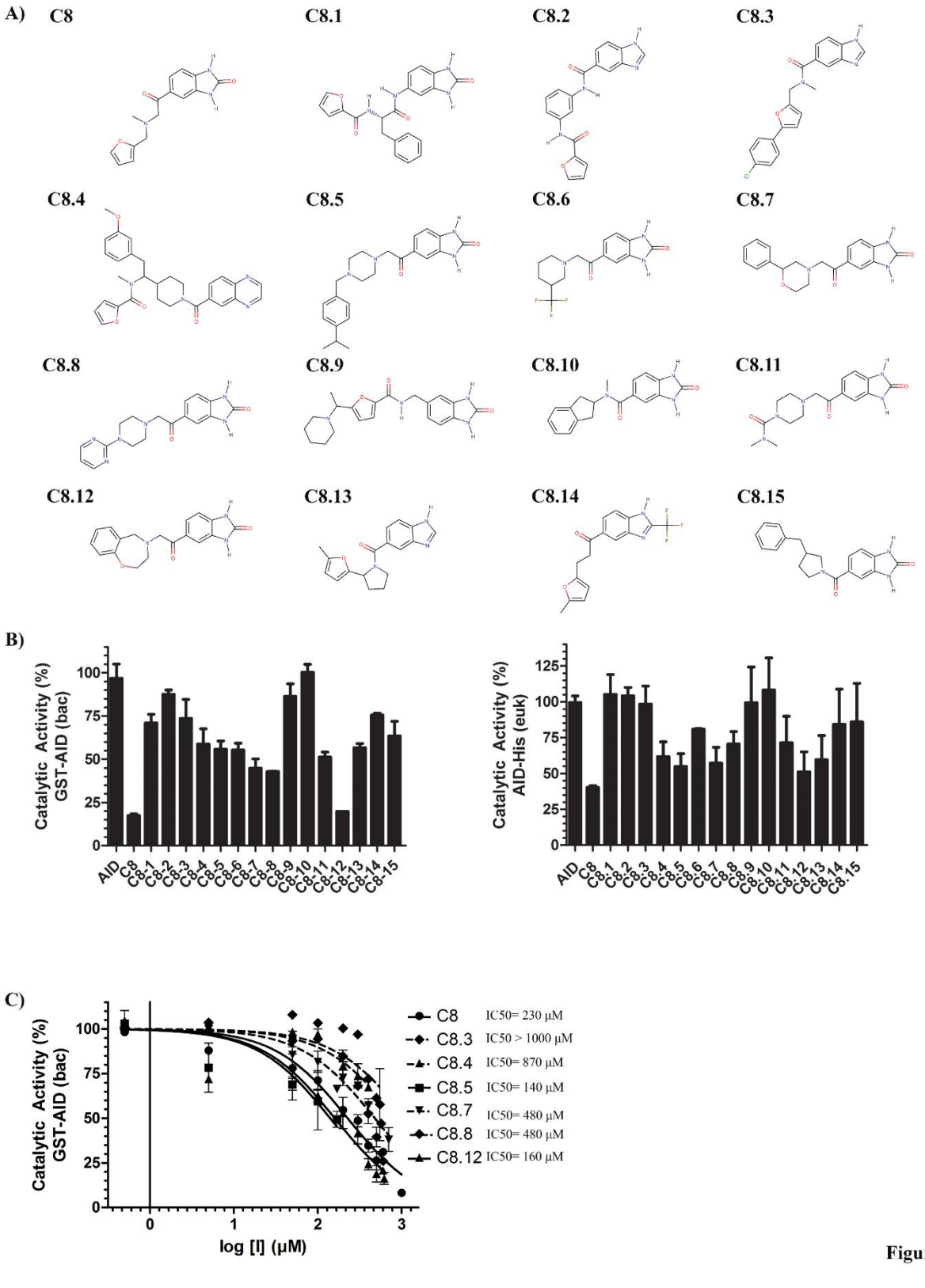
A)
C8
C8.1
C8.2
C8.3
C8.4
C8.5
C8.6
C8.7
C8.8
C8.9
C8.10
C8.11
C8.12
C8.13
C8.14
C8.15
B)
Catalytic Activity (%) GST-AID (bac)
Catalytic Activity (%) AID-His (euk)
AID
C8
C8.1
C8.2
C8.3
C8.4
C8.5
C8.6
C8.7
C8.8
C8.9
C8.10
C8.11
C8.12
C8.13
C8.14
C8.15
C)
Catalytic Activity (%) GST-AID (bac)
log [I] (μM)
C8 IC50= 230 μM
C8.3 IC50 > 1000 μM
C8.4 IC50= 870 μM
C8.5 IC50= 140 μM
C8.7 IC50= 480 μM
C8.8 IC50= 480 μM
C8.12 IC50= 160 μM

Figure S5

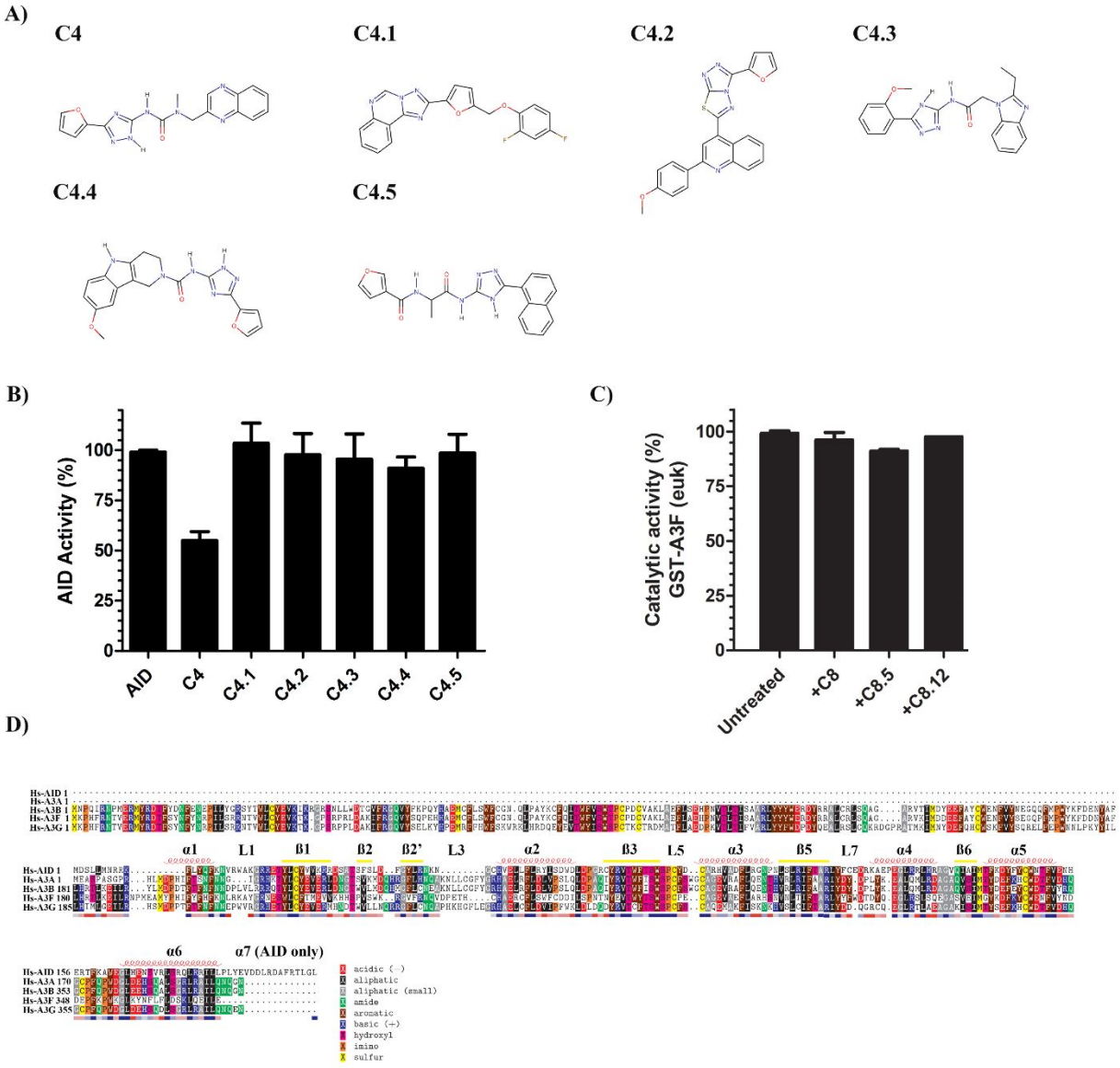
A)
C4
C4.1
C4.2
C4.3
C4.4
C4.5
B)
AID Activity (%)
100
75
50
25
0
AID
C4
C4.1
C4.2
C4.3
C4.4
C4.5
C)
Catalytic activity (%)
GST-A3F (euk)
100
75
50
25
0
Untreated
+C8
+C8.5
+C8.12
D)
Hs-AID 1
Hs-A3A 1
Hs-A3B 1
Hs-A3F 1
Hs-A3G 1
α1
L1
ß1
ß2
ß2'
L3
α2
ß3
L5
α3
ß5
L7
α4
ß6
α5
Hs-AID 1
Hs-A3A 1
Hs-A3B 181
Hs-A3F 180
Hs-A3G 185
α6
α7 (AID only)
Hs-AID 156
Hs-A3A 170
Hs-A3B 353
Hs-A3F 348
Hs-A3G 355
acidic ( )
aliphatic
aliphatic (small)
amide
aromatic
basic (+)
hydroxyl
imino
sulfur

Figure S6

AID bubble substrate

5'-GTC ACT GAA GCG GGA AGG GAC TG$^{T}_{G}$$^{\underline{TGC}}_{T}$$^{T}_{T}$CCG GGC GAA GTG CCG GGG CAG GAT CT-3'

3'-CAG TGA CTT CGC CCT TCC CTG AC$_{C}^{G}$$_{T}^{GGC}$$_{TTT}^{CCG}$ CTT CAC GGC CCC GTC CTA GA-5'

A3A, A3B and A3F substrate

5'-GAA GAG GTG TGT TGA TAA TAT AAG AGA TTC AAA AAA GAG AAA GAG GAA TGT TAA ATA-3'

A3G substrate

5'-GAA GAG GTG TGT TGA TAA TAT AAG AGA CCC AAA AAA GAG AAA GAG GAA TGT TAA ATA-3'

Figure S7

	AID 1		AID 2		AID 3		AID 4		AID 5	
ZINC ID	**Dock Rank (#)**	**Docking energy (kcal/mol)**	**Dock Rank (#)**	**Docking energy (kcal/mol)**	**Dock Rank (#)**	**Docking energy (kcal/mol)**	**Dock Rank (#)**	**Docking energy (kcal/mol)**	**Dock Rank (#)**	**Docking energy (kcal/mol)**
ZINC12323863 (C3)	**1**	-77	**1**	-73.9	**2**	-75.4	**2**	-87.8	**3**	-81.4
ZINC38767647	**2**	-70.2	**2**	-73.1	**1**	-78	**1**	-96.4	**1**	-85.7
ZINC64002748 (C8)	**3**	-65.9	**6**	-60.6	**ND**	ND	**6**	-81.9	**6**	-75.8
ZINC64470747	**4**	-64.7	**3**		**5**	-70	**3**	-86.4	**2**	-81.6
ZINC38532069 (C1)	**5**	-63.4	**ND**	ND	**22**	-53.3	**14**	-67.1	**16**	-59.7
ZINC38776626	**6**	-63.1	**4**	-64	**4**	-70.1	**4**	-85.1	**5**	-77.4
ZINC39411748	**7**	-63.1	**ND**	ND	**13**	-57	**18**	-65.1	**19**	-55.5
ZINC64470745	**8**	-62.9	**42**	-45.3	**15**	-55.8	**15**	-67.1	**14**	-59.7
ZINC63908937	**9**	-62.7	**5**	-60.6	**3**	-71.3	**5**	-84.4	**4**	-77.8
ZINC64470706	**10**	-61.8	**11**	-49	**12**	-57.3	**12**	-70.5	**11**	-63.4
ZINC63970468 (C7)	**11**	-60.5	**7**	-56	**8**	-66.9	**7**	-81.8	**7**	-73.1
ZINC38767171 (C10)	**12**	-59.6	**8**	53.8	**10**	-62.8	**8**	-78.4	**8**	-70.6
ZINC39729455	**13**	-59.4	**158**	-43.3	**19**	-55.2	**13**	-68	**12**	-60.2
ZINC40014267	**14**	-58.7	**69**	-44.6	**7**	-68.6	**10**	-71.1	**13**	-59.8
ZINC40014265	**15**	-58.4	**ND**	ND	**9**	-65.7	**9**	-74.5	**9**	-63.6
ZINC64470737 (C6)	**16**	-56.9	**144**	-43.4	**14**	-56.2	**17**	-66.4	**15**	-59.7
ZINC64470729	**17**	-56.1	**12**	-48.5	**11**	-59.7	**11**	-71	**10**	-63.5
ZINC97199736	**18**	-52.5	**ND**	ND	**ND**	ND	**ND**	ND	**ND**	ND
ZINC64012417 (C9)	**19**	-52.1	**ND**	ND	**18**	-55.4	**21**	-58	**23**	-47.9
ZINC39776444	**20**	-51.7	**251**	-42.5	**24**	-51	**20**	-59.2	**22**	-49.8
ZINC39709283	**23**	-50.8	**ND**	ND	**16**	-55.6	**19**	-61.4	**20**	-50.7
ZINC38768281	**160**	-46.8	**10**	-49	**25**	-50.9	**22**	-55.8	**25**	-44.8
ZINC00393537	**ND**	ND	**22**	-46.4	**17**	-55.5	**31**	-44.4	**18**	-56.2
ZINC39699427	**64**	-48	**142**	-43.5	**20**	-54.8	**16**	-66.6	**17**	-56.3
ZINC39953471	**ND**	ND	**ND**	ND	**21**	-54.6	**23**	-52.6	**28**	-43.7
ZINC64012414	**ND**	ND	**ND**	ND	**23**	-52.2	**24**	-52	**24**	-45.6
ZINC20255888	**ND**	ND	**ND**	ND	**ND**	ND	**25**	-47.1	**29**	-43.4
ZINC15757674	**ND**	ND	**ND**	ND	**ND**	ND	**26**	-47	**21**	-50.7
ZINC76563362	**49**	-48.3	**ND**	ND	**ND**	ND	**ND**	ND	**53**	-41.1
ZINC20580044 (C5)	**ND**	ND	**ND**	ND	**26**	-48.4	**ND**	ND	**211**	-39.3
ZINC21032580	**ND**	ND	**ND**	ND	**29**	-45.8	**ND**	ND	**324**	-38.8
ZINC38743557	**221**	-46.4	**ND**	ND	**297**	-41.1	**28**	-45.2	**76**	-40.6
ZINC38776749	**ND**	ND	**ND**	ND	**76**	-42.6	**29**	-44.5	**ND**	ND
ZINC38743555	**289**	-46.1	**ND**	ND	**ND**	ND	**32**	-44.1	**152**	-39.6
ZINC00424122 (C2)	**ND**	ND	**ND**	ND	**ND**	ND	**33**	-43.2	**62**	-40.8
ZINC38776242	**ND**	ND	**ND**	ND	**455**	-40.7	**34**	-42.4	**ND**	ND
ZINC48148145 (C4)	**323**	-46	**ND**	ND	**ND**	ND	**35**	-42.1	**ND**	ND
ZINC58327138	**ND**	ND	**ND**	ND	**ND**	ND	**67**	-40.6	**27**	-44.3
ZINC89851373	**ND**	ND	**ND**	ND	**ND**	ND	**ND**	ND	**30**	-43.3
ZINC73974308	**ND**	ND	**ND**	ND	**ND**	ND	**ND**	ND	**31**	-43.2

Table S1: 40 candidate compounds with the lowest binding energies to the catalytic pocket of AID

Supplementary Figure Legends

Figure S1. Comparison of the AID structure obtained by different methods. A. The crystal structure of near-native partially truncated AID (PDB: 5W1C) (gray colored) superimposed with a representative conformation of native AID published previously using our combined computational-biochemical method (beige colored). The crystal structure lacks the α7 C-terminal domain of AID and contains a mutation (E58A) in the active site. The purple sphere depicts the coordinated Zn in the active site. **B.** The electrostatic surface of the crystal structure of AID (left panel) and a representative AID conformation from the AID ensemble obtained using the combined computational-biochemical method (right panel). Positive, neutral, negative, and catalytic residues are represented by blue, white, red and purple surfaces, respectively.

Figure S2. Representative alkaline cleavage gel of AID inhibition by C1-C10 panel, MTT assay to evaluate toxicity of compounds and UDG inhibition assay. A. A representative preparation of bacterially-expressed and purified GST-AID incubated with C1-10, with AID in phosphate activity buffer used as positive control, AID + vehicle (DMSO) as the 100% bench-mark reaction, and substrates alone (no AID added) as negative control. Multiple independently purified preparations of AID (n=3-6) were used in each experiment. **B.** Since the candidate inhibitor compounds were dissolved in DMSO, AID activity was tested as a function of increasing DMSO concentrations. Results demonstrated that the DMSO concentrations used to solubilize the compounds do not impact AID activity negatively. **C.** MTT assay was performed on four cell lines and healthy primary cells. Cell lines included A549 (lung), MCF-7 (breast), 293 T (human embryonic kidney) and Raji (B cell lymphoma). Cells were incubated in triplicate with varying concentrations of C4 (left panel) and C8 (right panel). Untreated cells and vehicle treated cells were considered as negative controls. After 24 hours incubation at 37 °C, 10 µl of MTT reagent [3-(4,5-dimethylthiazol-2-yl)-2,5- diphenyltetrazolium bromide, 12 mM] was added followed by incubation for 4 hours and addition of 100 µl of SDS-HCl solution to each well, and incubation for 18 hours at 37 °C. Colorimetric evaluation was carried out using a spectrophotometer at 570 nm, and percentage of cell growth was calculated from the absorbance values of untreated and treated cells as % cell growth = (OD_{570} treated / OD_{570} untreated) × 100. To test toxicity on healthy primary cells, we tested peripheral blood monocytes (PBMC) from donors; 3×10^4 cells per well were transferred into same 96-well plates and treated with same concentrations of compounds and incubated for 24 and 48 h. C8 was non-toxic to all cells, but C4 exhibited some toxicity to 3/5 cell types tested. **D.** Since UDG is used downstream of AID in the alkaline cleavage assay, we sought to ensure that C4 and C8 are not acting through inhibition of UDG, using a UDG-inhibitor (UGI) as a positive control.

Figure S3: AID expression and deamination-specific PCR assay controls. A. Example western blot showing expression of eukaryotic-expressed untagged native AID compared with

bacterially-expressed and purified GST-AID, and eukaryotic-expressed AID-His in whole 293T cell lysate. **B.** Quantitative real-time PCR of AID expression levels relative to GAPDH expression levels in Daudi, Raji and Ramos cells indicating these cells express AID. **C**. Positive control reaction of AID-expressing 293T extracts incubated with substrate plasmid to demonstrate the deamination specificity of the PCR assay. **D.** Control PCR with C8 added after a purified AID + plasmid incubation to ensure C8 does not inhibit PCR. **E.** Deamination-specific PCR illustrating that like untagged AID, AID-His in 293T cell extracts is also inhibited by C8. **F.** Representative sequencing analysis of TA-cloned amplicons confirming the deamination specificity of the PCR assay used to detect AID activity in this experiment (mutated dC denoted by *).

Figure S4. Inhibition of endogenous intracellular AID detected by deamination-specific PCR. A. We used a modified version of the Deam-PCR assay: rather than the substrate plasmid being incubated in a cell-free reaction with purified AID (top row of assay schematic), or in a cell-free reaction with AID+ cell extracts (middle row of assay schematic) as in the classic version of the Deam-PCR, the substrate plasmid is transfected into AID-expressing cells, followed by Deam-PCR (bottom row of assay schematic). **B.** Raji cells were treated with C8, C8.5 (700 μM in media), vehicle (DMSO, 140 mM), or untreated, at the same time as being transfected with the deamination substrate plasmid. Recovered cell extracts containing plasmid DNA were then subject to Deam-PCR to probe for the presence of AID-targeted plasmids. Lanes 1-3 are the inhibitor-treated conditions, including Raji cells transfected with plasmid and treated with C8 (lanes 1 and 3) and C8.5 (lane 2). Lanes 4 and 5 are the equivalent inhibitor-untreated conditions, Raji cells transfected with the plasmid and treated with vehicle (140 mM DMSO), and untreated Raji cells transfected with plasmid substrate, respectively. Lane 6 is a negative control reaction of untransfected Raji cells, containing no plasmid substrate. Lane 7 is a positive control reaction of a cell-free Deam-PCR assay wherein Raji extract was incubated with the plasmid substrate. Lanes 8-10 are PCR controls reactions wherein the positive control reaction was mixed with 1 μL of the cell lysate template used in lanes 1-3 (the same volume of cell lysate used in the Deam-PCR of those lanes) in order to ensure that the lysate itself, containing the inhibitor, does not inhibit the Deam-PCR reaction. Lanes 8 and 9 contain lysates from the template for lanes 1 and 2, and lane 10 contains lysate from the template for lane 6. lane Lane 11 is a ddH_20 PCR negative control.

Figure S5. Structural analogues of C8 and inhibition of AID. A. The chemical structure of C8 and C8.1-C8.15 structural analogues. **B.** The catalytic activity of bacterially-expressed and purified GST-AID (left panel) and eukaryotic-expressed AID-His in 293T cell extracts (right panel) as a function of treatment with 15 structural analogues of C8. **C.** Catalytic activity of GST-AID as a function of log inhibitor concentration. In comparison to the parent C8, we found C8.5 and C8.12 were moderately similarly effective.

Figure S6. Ineffective structural analogues of C4, C8-analogue resistant A3F and AID/APOBEC3 sequence alignment. A. Chemical structure of C4 and C4 analogues. **B.** The catalytic activity of bacterially-expressed and purified GST-AID treated with C4 and C4.1-C4.5 analogues. None of the C4 analogues were effective inhibitors. **C.** GST-A3F was resistant to inhibition by C8, C8.5 and C8.12. **D.** Sequence alignment of AID, A3A, A3B, A3F and A3G, depicting secondary structure and position of secondary catalytic loops. The secondary structural elements above the alignments corresponds to the secondary structure of A3A (PDB: 2M65). The blue-red bar underneath the alignment corresponds to residue conservation, with blue being conserved and red representing divergent sequence.

Figure S7. AID, A3A, A3B, A3F and A3G oligo substrates. The nucleotide sequence of substrates used in inhibition assays for each enzyme.