

AID upmutants isolated using a high-throughput screen highlight the immunity/cancer balance limiting DNA deaminase activity

Meng Wang, Zizhen Yang, Cristina Rada & Michael S Neuberger

DNA deaminases underpin pathways in antibody diversification (AID) and anti-viral immunity (APOBEC3s). Here we show how a high-throughput bacterial papillation assay can be used to screen for AID mutants with increased catalytic activity. The upmutations focus on a small number of residues, some highlighting regions implicated in AID's substrate interaction. Many of the upmutations bring the sequence of AID closer to that of APOBEC3s. AID upmutants can yield increased antibody diversification, raising the possibility that modification of AID's specific activity might be used to regulate antibody diversification *in vivo*. However, upmutation of AID also led to an increased frequency of chromosomal translocations, suggesting that AID's specific activity may have been limited by the risk of genomic instability.

The immune system is unique in mammals in its use of active genomic mutation for a programmed physiological purpose. Within the adaptive immune system, the functionally rearranged immunoglobulin genes in activated B lymphocytes are targeted for deamination at cytosine residues by activation-induced deaminase (AID). Deamination within the immunoglobulin V gene triggers antibody gene diversification by somatic hypermutation and gene conversion, processes that underpin antibody affinity maturation. Deamination in the vicinity of the immunoglobulin switch (S) regions triggers the shift from the expression of IgM to that of one of the downstream isotypes (IgG, IgA or IgE)^{1,2}. Programmed DNA deamination is also used within the innate immune system, where deamination of retroviral replication intermediates by members of the apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3 (APOBEC3) family (which show sequence homology to AID) is associated with pathways of retroviral restriction^{3,4}.

Although such purposeful DNA deamination is beneficial to the host because it underpins major pathways in immunity, off-target action by the deaminases is likely to be harmful, leading to genomic instability and predisposition to cancer. Indeed, many of the chromosomal translocations and point mutations associated with proto-oncogene activation in tumors of the B cell lineage can be ascribed to events triggered by the action of AID⁵. Thus, the function of AID is likely to be regulated so as to achieve a balance between immunity and predisposition to cancer. Indeed, overexpression of AID in cell lines and in transgenic animals gives rise to increased genomic instability and tumor incidence, possibly because increased AID abundance results in an increased likelihood of it accessing inappropriate targets^{6–8}.

Here, to investigate whether it is possible to enhance antibody diversification by increasing the specific activity of AID, we use a high-throughput bacterial (*Escherichia coli*) papillation assay to screen for AID mutants showing increased activity. We use these mutants to show that the rate of antibody gene diversification is indeed limited by the specific activity of AID but show that, although the rate of antibody diversification can be enhanced using AID upmutants, this also leads to increased genomic instability. Intriguingly, relatively few specific amino acid substitutions are repeatedly identified when selecting AID upmutants (several of which are close to the catalytic site), with many of these upmutations bringing the sequence of AID closer to that of its APOBEC3 relatives. The results suggest that AID, which functions in the nucleus, may have experienced more stringent limitations to the upward evolution of its specific activity than its APOBEC3 relatives, which largely function in the cytosol.

RESULTS

Mutator screen for DNA deaminases

E. coli cells harboring a missense mutation within *lacZ* give rise to white colonies on MacConkey-lactose plates; within such white colonies, a few red microcolonies can often be discerned (called papilli; typically 0–2 per colony), which reflect spontaneously arising Lac⁺ revertants. Bacterial mutator clones with higher frequency of spontaneous mutation can be identified by an increased number of papilli.

Although papillation assays have been used to screen for *E. coli* mutants that are defective in some aspect of DNA repair^{9,10}, we wondered whether the assay could be adapted to screen for active

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, UK. Correspondence should be addressed to M.N. (msn@mrc-lmb.cam.ac.uk).

Received 11 March; accepted 22 May; published online 21 June 2009; doi:10.1038/nsmb.1623

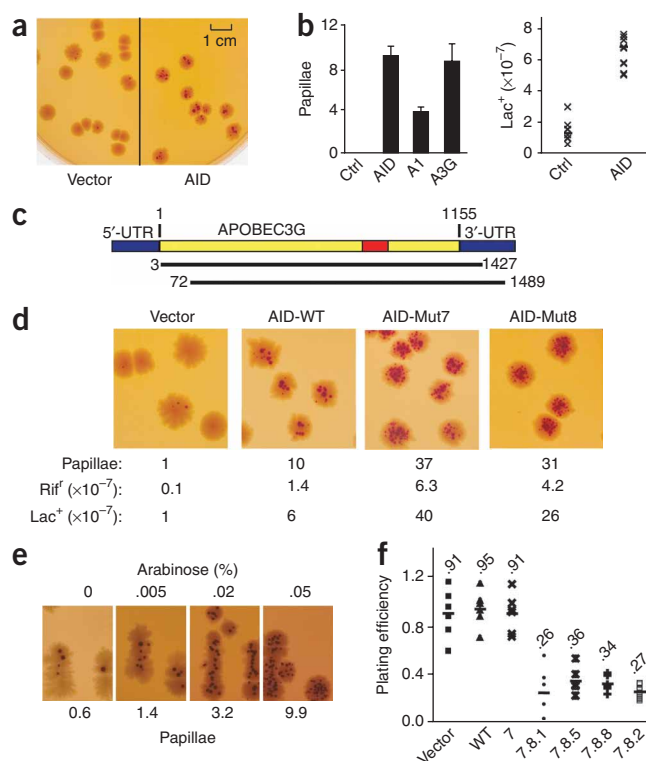


Figure 1 Papillation screen for active mutators. **(a)** Empty vector or AID-transformed *E. coli* strain CC102 was plated on MacConkey-lactose agar and grown at 37 °C for 4 d. **(b)** Frequency of reversion to Lac⁺ as determined by the average number of papillae per colony on MacConkey-lactose agar (left, histogram showing the mean and s.d. of three experiments) or by growth on minimal lactose plates (right, fluctuation analysis on six clones each) in CC102 transformants expressing vector control (Ctrl), AID, APOBEC1 (A1) or APOBEC3G (A3G). **(c)** Depiction of the two APOBEC3G cDNAs obtained by screening a human spleen cDNA library (50,000 colonies) for papillation on CC102. The wild-type full-length APOBEC3G mRNA is shown above and the structures of the two cDNAs below. The red shows the catalytic active site. The blue shows the untranslated region (UTR). Nucleotide residues are numbered relative to the start of the open reading frame (+1). **(d)** Comparison of papillation (and mutation frequencies on *lac* and rifampicin (Rif^r)) by wild-type AID and by upmutants Mut7 (K10E E156G) and Mut8 (K34E K160E). **(e)** Papillation by AID Mut1.1 expressed from plasmid pBAD30 as a function of the concentration (% (w/v)) of L-arabinose (inducer). **(f)** Bacterial titers in cultures grown to saturation in LB with ampicillin under conditions of IPTG induction, relative to the titers obtained from cultures grown in the absence of induction, for CC102 transformants expressing different AID upmutants. WT, wild type.

mutators. Strain CC102 carries a missense mutation in codon 461 of *lacZ*, with glutamate being substituted by glycine owing to a A-T to G-C transition mutation¹¹. If expression of AID in CC102 were to increase the rate of cytosine deamination at codon 461, this might be expected to increase the frequency of Lac⁺ revertants. This is indeed the case. AID-expressing transformants of CC102 give an increased frequency of papillation on MacConkey-lactose plates (**Fig. 1**). After 6 d of incubation, the number of papilli increased from 0–2 per colony to 8–10 per colony—this correlates with a more than three-fold increase in the frequency of Lac⁺ revertants in overnight cultures, as judged on minimal-lactose plates. Sequence analysis of six such Lac⁺ revertants confirmed that they had indeed arisen through reversion at codon 461. The AID-related deaminases APOBEC1 and APOBEC3G can also trigger increased papillation when expressed in CC102 (**Fig. 1b**).

To assess the robustness of the assay, we asked whether it could be used to isolate active mutators from a total splenic cDNA library. We introduced a human spleen cDNA library into CC102 and screened 50,000 colonies for enhanced papillation. We identified 36 possible candidates, which we then retested by streaking on MacConkey-lactose plates. Only two colonies were confirmed as giving increased papillation. Sequence analysis revealed that they carried distinct cDNAs derived from APOBEC3G (**Fig. 1c**), demonstrating that the assay can indeed be used as a high-throughput screen for active mutators. The most likely reason that the screen yielded APOBEC3G rather than AID relates to the fact that the cDNA library is derived from unstimulated human spleen (containing both B and T cells, where APOBEC3G is well expressed), whereas AID expression is largely restricted to activated B cells.

Selection for AID upmutators

To see whether we could isolate variants of human AID showing increased mutator activity in *E. coli*, we subjected a human AID cDNA

to PCR mutagenesis, cloned it into plasmid pTrc99a and transformed the plasmid into *E. coli* CC102. We then assessed the diversity of the library of AID mutants thereby generated by sequencing 48 randomly selected clones. These clones each carried 3–5 nucleotide substitutions per kilobase, with no two clones carrying an identical mutation. Screening of a total of 60,000 colonies from four independent PCR mutagenesis experiments yielded 13 AID transformants showing increased papillation on MacConkey-lactose plates (**Figs. 1d** and **2**). We then retransformed the AID-containing plasmids from nine of these mutants into *E. coli* strain KL16, and tested them for the frequency with which they yielded rifampicin-resistant colonies. All nine showed an increased frequency of mutation at the *rpoB* locus.

The AID cDNAs from two of the first-generation upmutants (Mut1 and Mut7) were then themselves subjected to PCR mutagenesis, and we obtained second-generation mutants showing enhanced papillation (**Fig. 2**). In fact, the high papillation shown by these second-generation mutants made it difficult to visually discern any additional increases in papillation. This meant that we needed to modify the assay in order to screen for further enhancement of mutator activity in a third round of mutation and selection. To this end, we moved AID Mut1.1 and 7.3 cDNAs to an arabinose-inducible expression vector such that the number of papilli obtained in CC102 transformants could be regulated by varying the concentration of arabinose in the medium (**Fig. 1e**). Screening for papillation under low (0.02%) arabinose, we obtained a third generation of AID upmutants, some of which gave a mutation frequency nearly 400 times greater than that of the wild-type AID, as judged by the frequency of mutation to rifampicin resistance (**Fig. 2**).

Several of the third-generation mutants seemed to cause toxicity in *E. coli*, as judged by the smaller colony size when grown under inducing conditions; this was accompanied by a reduced viable-cell count in bacterial cultures grown to saturation (**Fig. 1f**). This toxicity might have caused some highly papillating mutants to give anomalously low frequencies of mutation to rifampicin resistance (for example, Mut7.3.4; **Fig. 2**), with AID expression possibly being downregulated during overnight culture.

Hotspots for upmutation of human AID

Apart from the premature stop codon mutations identified in three of the AID upmutants (Mut5, Mut1.3 and Mut1.5), analysis of the sequences of the various AID upmutants revealed a marked preference

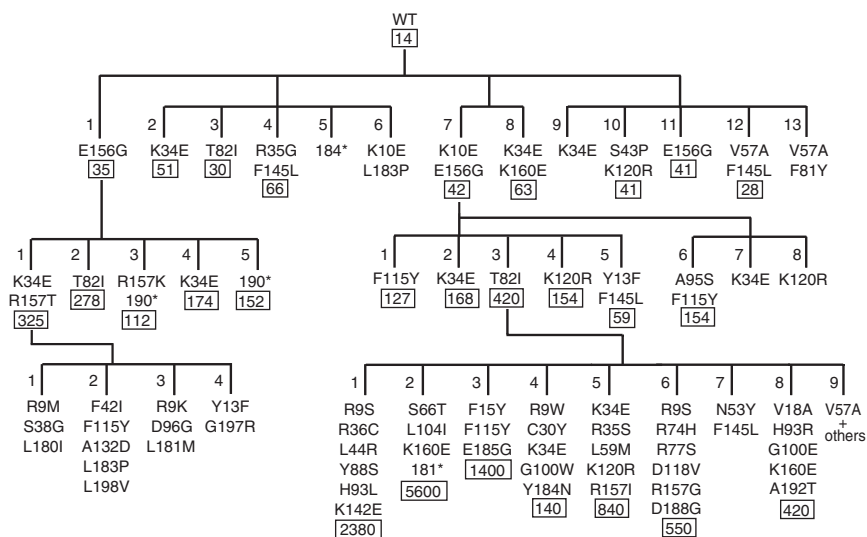


Figure 2 Dynasty of AID upmutants selected by papillation screen. Upmutants obtained in three successive rounds of mutagenesis, with mutants obtained from individual PCR-mutagenesis experiments grouped as families of siblings. The additional amino acid substitutions introduced in each round of mutagenesis are indicated, with the numbers below giving the mean frequency of mutation to rifampicin resistance relative to vector. * indicates a C-terminal truncation caused by introduction of a premature stop codon at the indicated codon. Individual mutants are numbered according to their dynastic origin. Thus, for example, Mut7 (K10E E156G) is the parent of Mut7.1 (K10E E156G F115Y).

the likely catalytic site (V57A and T82I), others are in a region equivalent to a portion of APOBEC3s that have previously been suggested to be involved in polynucleotide binding (F115Y and K120R)^{12–14}; several are clustered in regions whose function is unknown.

Upmutations increase AID's specific activity

We were interested in ascertaining whether the mutations in AID that increased its bacterial mutator activity acted by increasing the yield of soluble protein or by increasing its specific activity. Initial screening of the sonic extracts of a large number of upmutants did not reveal any that showed a substantial increase in the fractional yield of soluble protein, as judged by western blot analysis. However, when we generated glutathione *S*-transferase (GST) fusion proteins from human upmutants Mut1.1 and Mut7.3.6, we found a clear increase in specific activity, as judged by *in vitro* deamination assays performed on a single-stranded oligonucleotide substrate (Fig. 3b,c). The increased activity seems to reflect an increase in V_{max} rather than any major change in K_m . Thus, by performing the assay over a range of substrate concentrations, we estimate the K_m values of the wild-type

for certain amino acid substitutions. Thus, for example, the K34E, T82I and E156G substitutions (each of which is sufficient on its own to increase AID activity) have actually been selected in independent experiments. These mutations were not found among sequences of 48 random (that is, unselected) clones from the PCR-generated libraries, in which we observed a wide spectrum of mutations without indications of any major hotspots of the mutagenesis procedure itself. Thus, the repeat identification of a few amino acid substitutions suggests that there are a limited number of single amino acid substitutions in AID that yield increased papillation.

Although in some cases (especially in the third generation) the multiplicity of mutations introduced in a single round prevents unambiguous identification of those mutations responsible for the increased papillation, in many cases the relevant upmutation can be definitively identified because it constitutes the sole difference between a pair of differently papillating AID sequences or (less definitively) because it was independently obtained in multiple PCR reactions. The locations of such upmutations are depicted in Figure 3a: whereas some are located around the zinc-coordination motif in the vicinity of

Figure 3 Nature of the AID upmutations. (a) Primary sequence of human AID with upmutations identified in color. Mutations at residues in red constitute the sole difference between at least one pair of AID sequences resulting in a more than two-fold increase in mutation frequencies at *rpoB*. Mutations at residues in orange have been identified in multiple independent upmutants but in the presence of other substitutions. The box above or below the highlighted residues shows the identity of the substitution mutations and the frequency with which each substitution was detected in the total of nine independent libraries. Underlined letters identify residues where the corresponding position in fugu AID also seems to be a site of selected upmutation, as judged by the fact that it is either the sole mutation or present in multiple fugu upmutants. The zinc-coordination motifs (HVE, PCYDC) and regions of suggested polynucleotide contact (FCEDRKA)^{11–13} are highlighted by a blue box. (b) Deaminase activity of semipurified GST-AID fusion proteins (100 ng, 200 ng and 400 ng) was analyzed on an oligonucleotide substrate at the indicated time points. Protein abundance was monitored by western blot analysis using anti-AID antibody. (c) Quantification of extent of deamination over time. (d) The target specificity of the upmutants as judged by the spectrum of *rpoB* mutations in rifampicin-resistant (Rif^r) colonies. Transition mutations at any one of 11 C-G pairs within *rpoB* can give rise to Rif^r. The distribution of such mutations among Rif^r colonies is shown for AID upmutants Mut8 (orange), Mut1.1 (pink), Mut1.2 (blue), Mut7.3.5 (red) and Mut7.3.6 (green).

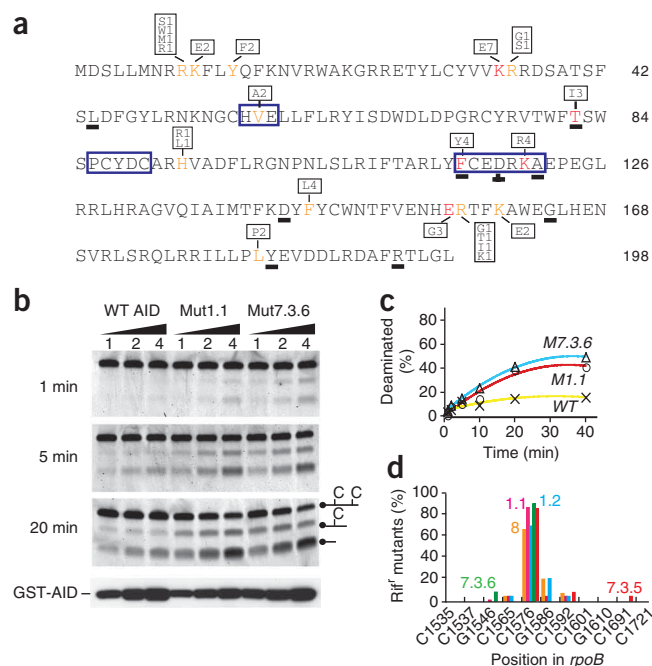
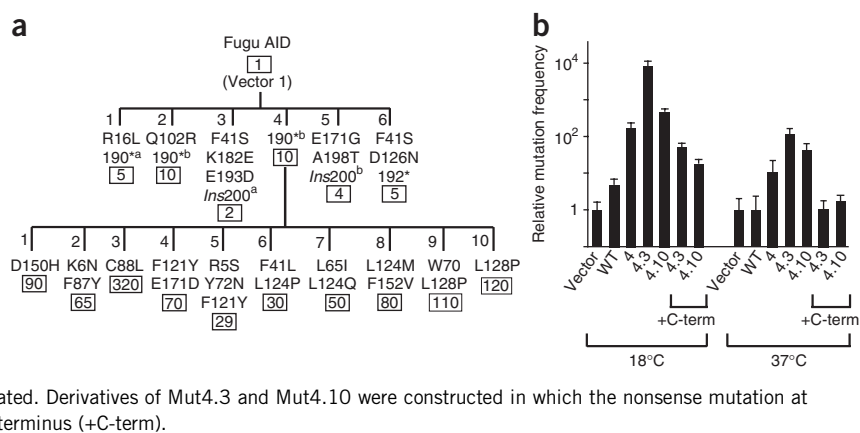


Figure 4 Upmutants of fugu AID. (a) Dynastry of upmutants of fugu AID selected at 37 °C. * indicates a C-terminal truncation caused by introduction of a premature stop codon; *^a and *^b indicate different single-nucleotide substitutions at codon 190 that cause the premature stop codon; Ins200^a and Ins200^b indicate different single-nucleotide insertion mutations at codon 200 that cause the C-terminal region to be read out of frame. Numbers below give the mean frequency of mutation to rifampicin resistance (Rif^r) relative to the empty vector. (b) The frequency of mutation to Rif^r relative to vector-only transformants at either 18 °C or 37 °C is shown for *E. coli* K16 transformed with plasmids encoding wild-type (WT) or mutated fugu AID, as indicated. Derivatives of Mut4.3 and Mut4.10 were constructed in which the nonsense mutation at 190 had been reverted, thereby yielding a wild-type C terminus (+C-term).



GST-AID fusion protein for the oligonucleotide substrate, as well as that of the upmutants Mut1.1 and Mut7.3, to be in the range 80–100 nM (values some five-fold greater than that determined for wild-type GST-AID on a hotspot bubble substrate¹⁵); however, the V_{max} of Mut1.1 and Mut7.3 were calculated to be 0.4 fmol min⁻¹ ng⁻¹ and 0.3 fmol min⁻¹ ng⁻¹, respectively, as compared to 0.07 fmol min⁻¹ ng⁻¹ for the wild-type fusion protein (Supplementary Fig. 1). Although appreciable, the increase in V_{max} of the mutant proteins (four- to six-fold) is notably less than the increase in the frequency of mutation to rifampicin resistance that they trigger in bacteria (20- to 30-fold). It is possible that this apparent discrepancy could reflect a high frequency of DNA deamination, catalyzed by the AID upmutants, that might overwhelm the normal bacterial DNA repair pathways.

This increased specific activity does not seem to have been accompanied by any gross change in the target specificity, as analysis of the *rpoB* mutations obtained using several human AID mutants did not reveal any major difference in their mutation spectrum (Fig. 3d).

Upmutations in fugu AID are in analogous positions

We were interested in extending this study to an evolutionarily distant AID to see whether analogous mutations would underpin increased mutator activity. We have shown previously that AID from pufferfish (which live at around 26 °C) shows little bacterial mutator activity when assayed at 37 °C, whereas mutator activity can be detected at 18 °C¹⁶. We therefore asked whether the papillation assay could be used to select mutants of fugu AID that gave robust papillation at 37 °C. We were able to isolate such mutants, but all the first-generation mutants we isolated harbored C-terminal truncation mutations, with the six mutants we obtained harboring five distinct truncation mutations (Fig. 4). Various amino acid substitutions, however, could then lead to enhanced papillation in second-generation mutants (Fig. 4a), with several of these occurring at positions analogous to the upmutations identified in human AID (Figs. 3a and 4a and Supplementary Fig. 2). Thus, the mutation (C88L) responsible for the increased activity of fugu AID Mut1.3 occurs at the equivalent position to the T82I mutation in human AID (Supplementary Fig. 2). Similarly, residues Phe121, Leu124 and Leu128 in fugu AID (each of which is a target for mutation in either two or three fugu upmutants) are all located in a stretch of fugu AID corresponding to residues 115–121 in human AID, a region from which we also obtained upmutants.

Although we detected C-terminal truncations among the panel of human AID upmutants, and such truncations have been shown to

give higher mutator activity in *E. coli* (refs. 17,18 and our unpublished data), it was nevertheless notable that all the first-generation mutants of fugu AID selected at 37 °C carried truncations at the C terminus. This led us to ask whether these C-terminal mutations underpinned increased thermal stability, and whether the amino acid substitutions giving rise to increased papillation in the second-generation fugu upmutants might have been indiscernible at 37 °C in the absence of a C-terminal truncation mutation. This does indeed seem likely. The C88L and L128P substitutions both give increased frequency of mutation to rifampicin resistance as assayed at 18 °C in the presence or absence of a C-terminal truncation. However, when assayed at 37 °C, these amino acid substitutions did not give any discernible increase in mutation frequency in the absence of the C-terminal truncation (Fig. 4b).

AID upmutation increases antibody diversification

Although several studies have revealed that the efficacy of class switching is limited by the abundance of AID polypeptide^{19–25}, it is unknown whether the efficacy of antibody gene diversification is also limited by the specific activity of AID. We were therefore interested in ascertaining whether AID upmutants showing increased specific activity would also yield increased antibody gene diversification. To this end, we expressed Mut3 (T82I), Mut8 (K34E K160E) and Mut7.3 (K10E E156G T82I) in an AID-deficient chicken DT40 B cell line in which somatic mutation of the IgV can be inferred from the frequency of generation of sIgM-loss variants²⁶. Both Mut3 and Mut7.3 seems to give substantially enhanced somatic mutation, as judged by this sIgM-loss assay. Furthermore, sequence analysis revealed that, after 1 month of clonal expansion, cells expressing these mutant AIDs indeed carried a higher mutation load in the IgV λ gene than did control cells expressing the wild-type enzyme (Fig. 5). Not only did a higher proportion of sequences carry mutations, but those that did carry mutations also carried a higher mutation load. This effect is particularly marked when account is taken of the fact that the mutant AID is expressed at lower abundance than its wild-type counterpart in these transfectants. In contrast, Mut8 did not give enhanced somatic mutation, indicating that the K34E and/or K160E substitutions are likely to diminish aspects of AID's function in B cells. Notably, Mut8 polypeptide was found at much higher abundance in the DT40 transfectants than were the Mut3 or 7.3 polypeptides. This is consistent with observations that we have made in other work (ref. 27 and our unpublished data) that AID mutants showing compromised activity in antibody

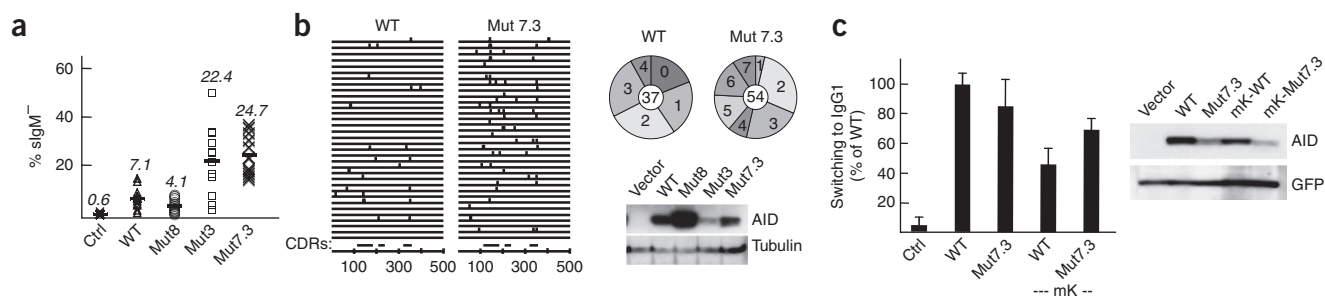


Figure 5 Enhanced antibody diversification by AID upmutants. **(a)** Somatic mutation of the IgV gene was assayed by monitoring surface IgM loss (% sIgM⁻) in AID^{-/-} ϕ V^{-/-} sIgM⁺ DT40 cells that had been stably transfected with constructs coexpressing the indicated AID mutants together with GFP (mean of 12 independent clonal transfectants). **(b)** The distribution of mutations in 34 IgV λ sequences obtained by PCR amplification from unsorted wild-type (WT)-transfected or Mut7.3-transfected DT40 cells 1 month after transfection is shown on the left. The corresponding positions of the complementarity-determining regions (CDRs) are shown below. Right, pie charts depict the number of mutations in PCR-amplified IgV λ sequences that had been obtained from the same transfectants but after sorting for sIgM loss; the number in the middle of the pie chart indicates the total number of sequences analyzed. Below the pie charts, western blots show AID abundance in the DT40 cell extracts with tubulin as loading control. The IgV λ sequences are provided in **Supplementary Figure 3**. **(c)** Histogram showing switching to IgG1 in LPS+IL4 cultures of AID-deficient B cells that have been transduced with vectors co-expressing GFP and WT or Mut7.3 (mean and s.d. of four experiments). 'mK' indicates that transduction was performed using vectors with a mutated Kozak sequence, so as to reduce the extent of AID overexpression. A western blot shows AID abundance in B cells extracts 3 d after retroviral transduction; the blot was reprobbed with anti-GFP antibodies as a control. Representative flow cytometry plots for **a** and **c** are shown in **Supplementary Figure 4**.

diversification and genomic mutation in DT40 cells tend to be expressed at higher abundance, without any evident alteration in intracellular localization. We suspect that the explanation for these differences in expression levels is that, in cell transfectants, there is selection against cells expressing high levels of AID proteins causing genomic instability.

To assay the activity of the mutant AID in class-switch recombination, we used an assay based on retroviral transduction of the mutant enzymes into AID-deficient mouse B cells. We were concerned that this assay involves considerable overexpression of AID, such that the amount of protein might saturate the switching assay. We therefore performed the assay using both the conventional pMX-Ig virus and a variant in which the transduced AID is expressed at lower levels through mutation of the Kozak sequence²¹. AID Mut7.3 was more effective in promoting CSR than the wild-type counterpart, even though Mut7.3 was expressed at lower levels (**Fig. 5**). However, upmutation of AID did not always yield increased class switching. Thus, for example, Mut7.3.2, although a potent mutator, gave no detectable switching in transduced B cells—presumably owing to its C-terminal truncation (data not shown).

AID upmutation increases chromosomal translocations

To ascertain whether the AID upmutants also gave increased chromosomal translocations, we extracted B cells from AID-deficient mice, retrovirally transduced the cells for AID expression and cultured them *in vitro* for 1–2 d⁷. We then subjected the genomic DNA from transduced cells to a PCR-based assay²⁸ that can detect translocation between the *c-myc* and IgH loci. Wild-type AID gave rise to an estimated frequency of 1 *c-myc*–IgH translocation per 2×10^5 cells. In contrast,

Mut7.3 gave a substantially higher frequency of 1 translocation per 5×10^4 cells (**Fig. 6**), with the increase in translocation frequency being similar to the increase in Mut7.3's specific deaminase activity.

Upmutations bring AID sequence closer to APOBEC3s

The analysis of the human AID upmutants identifies a collection of single-amino-acid substitutions, each of which can be demonstrated to be individually sufficient to increase the enzyme's mutator activity (**Fig. 7a**). On the basis of the sequence alignment of AID and APOBEC3s, we thought it should now be possible to design specific amino acid substitutions in individual APOBEC3 family members

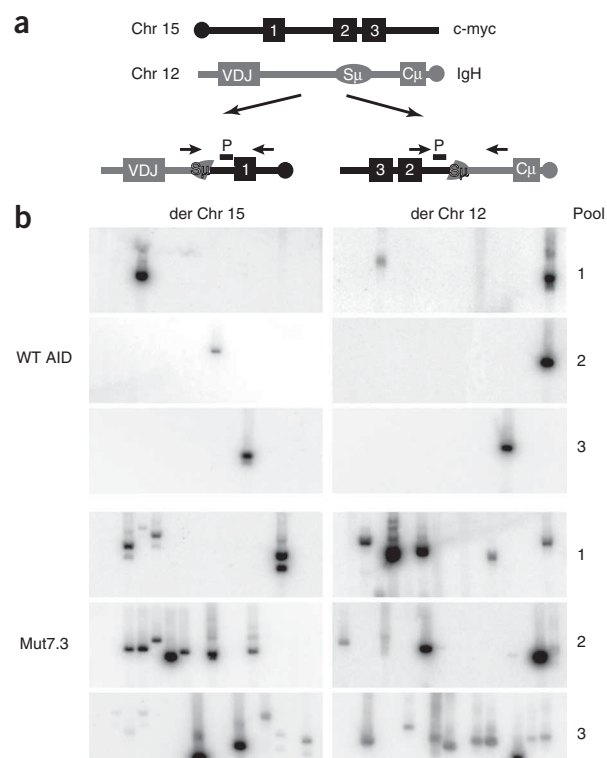
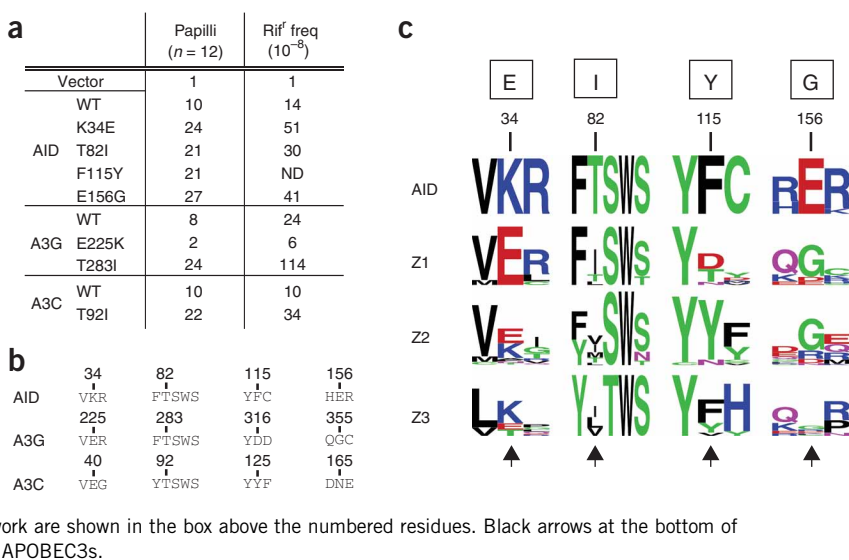


Figure 6 Increased chromosomal translocations by AID upmutants.

(a) Scheme of reciprocal translocation between the *c-myc* and IgH locus indicating the primers used for PCR amplification (arrows) and the probes (P) used for Southern blot hybridization. **(b)** Translocations derived from chromosomes (Chr) 15 and 12 (der Chr 15, der Chr 12) were amplified by PCR on three independent pools of genomic DNA from 2×10^5 cell-equivalents of AID-deficient B cells retrovirally transduced with pMX-GFP constructs directing the expression of either wild-type (WT) or Mut7.3 AID.

Figure 7 Comparison of AID with APOBEC3s.

(a) Mutator activity of mutants of AID, APOBEC3G and APOBEC3C carrying the indicated single-amino-acid substitutions, as monitored in the papillation and rifampicin-resistance (Rif^r) assays. (b) Sequence alignments of selected regions of AID, APOBEC3G and 3C. Numbers above the sequences indicate the positions of the residues in the parental sequence. (c) Web LOGO (<http://weblogo.berkeley.edu/>) alignment depicting amino acid conservation surrounding the major sites of upmutation of AID and the homologous regions in the Z1, Z2 and Z3 domains of mammalian APOBEC3s (cow, sheep, pig, dog, peccary, horse, cat, dog, mouse, rat, human and macaque; sequence accession numbers are provided in **Supplementary Table 1**). Any sequence with more than 90% amino acid identity to any other sequence was discarded from generation of the LOGO profiles. The AID upmutations identified in this work are shown in the box above the numbered residues. Black arrows at the bottom of the alignment highlight the homologous residues in the APOBEC3s.



that could be predicted to similarly up- or downregulate their activity. Thus, by comparing the sequence of AID to that of APOBEC3G (Fig. 7b), we predicted that mutation of the APOBEC3G Glu225 residue 'back' to lysine should diminish its mutator activity. This was indeed the case (Fig. 7a). Similarly, we successfully predicted that, analogously to what we found with AID, the mutator activity of both APOBEC3C and APOBEC3G should also be increased by substituting the threonine residue preceding their SWS motif (Thr283 in APOBEC3G; Thr92 in APOBEC3C) by an isoleucine (Fig. 7a).

A notable feature of the AID upmutations emerges when the sequence of AID is aligned to those of the family of APOBEC3 deaminases (Fig. 7c). The APOBEC3 family is presumed to have been derived from a duplication of AID and subsequently underwent rapid evolution to yield multiple APOBEC3 copies in higher animals; their zinc-coordination domains can be classified by sequence homology into one of three subgroupings (Z1, Z2 and Z3)^{16,29}. Alignment of AID sequences with those of the APOBEC3s reveals that most of the frequently selected upmutations in human AID serve to bring the sequence of AID closer to that of its APOBEC3 relatives (Fig. 7c). In fact, whereas the AID upmutation at Phe115 substitutes the amino acid preferred at the corresponding position in APOBEC3 Z2 domains (tyrosine), the upmutations at Lys34, Thr82 and Glu156 all substitute to the preferred amino acid at the corresponding position in the APOBEC3 Z1 domains. Notably, it is these Z1 domains that were previously found to be the most catalytically active of APOBEC3 domains³⁰. Thus, it seems that, whereas the deamination activity of AID can be artificially increased by specific upmutations, such upmutations may have been counterselected during the evolution of AID but not during the evolution of most APOBEC3s.

DISCUSSION

In this work, we have shown how it is possible to use a bacterial papillation assay to screen for active mutators and also obtain AID mutants with increased catalytic activity, with some of these AID mutants giving more efficacious antibody diversification, albeit at the expense of increased chromosomal translocations.

Recent studies have revealed that reducing the abundance of AID polypeptide reduces the efficiency of class-switch recombination or that increasing its abundance increases the risk of chromosomal translocations^{19–25}. However, the finding that switching and translocations

can both be increased by increasing the specific activity of AID (while actually diminishing its abundance) was unexpected. The results strongly suggest that the targeting of AID to the immunoglobulin locus will not always lead to deamination events and raise the possibility that post-translational modification of AID (for example, phosphorylation) could be used to regulate antibody diversification by modulating AID's specific activity, not just by controlling its interactions with partners^{31,32}. This will be an important topic for further work.

It will be interesting to ascertain whether the higher frequencies of somatic mutation that can be achieved with the AID upmutants could also lead to more efficient antibody mutation *in vivo*. The antibody hypermutation rate *in vivo* has been estimated to be on the order of 10⁻³ mutations per base pair per generation, with theoretical studies suggesting that this frequency might be optimal: a lower frequency might generate beneficial mutations too rarely, whereas a higher mutation rate might cause an overload of deleterious mutations^{33–38}. The construction of transgenic mouse lines expressing some of the AID upmutants described in this work might now allow one to address the issue of whether antibody maturation *in vivo* could be accelerated by increasing the hypermutation rate, provided that such mice do not show too severe an increase in cancer predisposition. Furthermore, if such upmutants do allow more efficient antibody maturation *in vivo*, the mouse lines might be useful in monoclonal antibody generation. Alternatively, the upmutants might well be used to allow more efficient antibody maturation through *in vitro* approaches that exploit hypermutating cell-lines^{39–41}.

Although many independent AID upmutants were selected in this work, it is notable that increased AID activity was repeatedly obtained through a relatively small cohort of individual amino acid substitutions. Several of the upmutations were obtained in equivalent positions in the human and fugu AID sequences. Some of the favored sites (Val57 and Thr82 in human AID; Cys88 in fugu AID) lie within the zinc-coordination motif and are presumably located at or close to the active site. Several of the others (Phe115 and Lys120 in human AID; Phe121, Leu124 and Leu128 in fugu AID) lie within a short stretch of AID that is equivalent to a region of APOBEC3 that, it has been speculated, might be implicated in DNA binding^{12–14}. The AID upmutants give good support to the idea that this region is indeed implicated in AID's interaction with DNA. The remaining hotspots of human AID upmutation are largely clustered in the regions of residues

9–13, 34–35 and 156–160. We note that a previous study⁴² also observed that a double mutation at positions 35 and 36 affected AID's catalytic activity. The mechanism by which these changes affect AID's activity is, however, less readily explicable, although more insight will presumably arise once the three-dimensional structure of AID has been elucidated.

A wholly unexpected finding of this work was that many of the favored amino acid substitutions selected in the AID upmutants (for example, at positions 34, 82, 115 and 156) had, in a sense, already occurred during the natural evolution of APOBEC3 family members. So why had they occurred during the evolution of APOBEC3s but not in AID? In the case of the K34E substitution, it is likely that this mutation was not selected during AID evolution because, although it increases AID's deaminase activity, the mutation destroys AID's *in vivo* antibody-diversifying activity (possibly by precluding a necessary interaction with a partner). However, the same argument does not apply to the T82I and E156G mutations, because AID upmutants harboring these substitutions do increase antibody diversification in B cells.

Although it is possible that increased efficacy of antibody diversification might not benefit the host organism, we believe, as discussed above, that a more likely explanation is that the risk of genomic instability has limited the upward evolution of AID's specific activity. It is possible that such restrictions may have been less stringent in limiting the specific activity of most APOBEC3 deaminases. Thus, the best characterized of the APOBEC3 family (APOBEC3G) is largely restricted to, and indeed retained in, the cytoplasm, where it can become incorporated into viral capsids^{43–45}. In contrast, AID shuttles into the nucleus to carry out its physiological mutator function. So the risk of genomic stability may have placed different constraints on the mutator activity of AID and some of the APOBEC3s. Consistent with this speculation, whereas enforced expression of AID in transgenic mice has been shown to lead to cancer⁶, we have observed no increased tumorigenesis in transgenic mice expressing high levels of APOBEC3C or APOBEC3G (C. Thomas and M.S.N., unpublished data). Nevertheless, limiting deaminase-specific activity is unlikely to be the sole mechanism for minimizing the risk of ectopic mutation. APOBEC3A, despite being the most active deaminase of the APOBEC3 family, is nevertheless found in the nucleus⁴⁶; it will be interesting to find out whether APOBEC3A overexpression is tumorigenic or whether other mechanisms act to limit its off-target activity.

Finally, we note that, although we have shown in this work that a bacterial papillation assay can be used as a high-throughput screen for AID upmutants, we can readily envisage that the assay could be extended to screen for upmutants of other members of the AID and APOBEC family or for deaminase mutants with altered target specificity. Furthermore, it is also possible that a papillation screen could be used to identify as-yet-uncharacterized active mutators; one could envisage that such active mutators might have a role in, for example, the initiation of variant surface glycoprotein variation in trypanosomes or in repeat-induced point mutation in fungi.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/nsmb/>.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We are indebted to J. Miller (Molecular Biology Institute and Department of Biology, University of California, Los Angeles) for kindly providing *E. coli* strain CC102 and recommendations regarding plating, to J.-M. Buerstedde (Munich) for

providing the DT40 cell lines, to O. Perisic (Cambridge) for providing the pOPTG vector, to Silvestro G. Conticello (Istituto Toscano Tumori) for helpful suggestions and the James Baird and Frank Elmore funds for support to M.W.

AUTHOR CONTRIBUTIONS

M.W. and Z.Y. performed the papillation assay and selection of human AID upmutants; M.W. performed all other experiments; C.R. designed and assisted with the class-switching assay and translocation assay; M.S.N. designed the overall research and wrote the manuscript.

Published online at <http://www.nature.com/nsmb/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Alt, F.W. & Honjo, T. (eds.). AID for Immunoglobulin Diversity. *Advances in Immunology* Vol. 94 (Elsevier, Amsterdam, 2007).
- Di Noia, J.M. & Neuberger, M.S. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22 (2007).
- Chiu, Y.L. & Greene, W.C. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu. Rev. Immunol.* **26**, 317–353 (2008).
- Haché, G., Manky, L.M. & Harris, R.S. Human APOBEC3 proteins, retrovirus restriction, and HIV drug resistance. *AIDS Rev.* **8**, 148–157 (2006).
- de Yébenes, V.G. & Ramiro, A.R. Activation-induced deaminase: light and dark sides. *Trends Mol. Med.* **12**, 432–439 (2006).
- Okazaki, I.M. *et al.* Constitutive expression of AID leads to tumorigenesis. *J. Exp. Med.* **197**, 1173–1181 (2003).
- Ramiro, A.R. *et al.* Role of genomic instability and p53 in AID-induced c-myc-IgH translocations. *Nature* **440**, 105–109 (2006).
- Robbiani, D.F. *et al.* AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell* **135**, 1028–1038 (2008).
- Nghiem, Y., Cabrera, M., Cupples, C.G. & Miller, J.H. The mutY gene: a locus in *Escherichia coli* that generates G-C→T-A transversions. *Proc. Natl. Acad. Sci. USA* **85**, 2709–2713 (1988).
- Ruiz, S.M., Létourneau, S. & Cupples, C.G. Isolation and characterization of an *Escherichia coli* strain with a high frequency of C-to-T mutations at 5-methylcytosines. *J. Bacteriol.* **175**, 4985–4989 (1993).
- Cupples, C.G. & Miller, J.H. A set of lacZ mutations in *Escherichia coli* that allow rapid detection of each of the six base substitutions. *Proc. Natl. Acad. Sci. USA* **86**, 5345–5349 (1989).
- Conticello, S.G., Langlois, M.A. & Neuberger, M.S. Insights into DNA deaminases. *Nat. Struct. Mol. Biol.* **14**, 7–9 (2007).
- Chen, K.M. *et al.* Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116–119 (2008).
- Holden, L.G. *et al.* Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **456**, 121–124 (2008).
- Larijani, M. *et al.* AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. *Mol. Cell. Biol.* **27**, 20–30 (2007).
- Conticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K. & Neuberger, M.S. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol. Biol. Evol.* **22**, 367–377 (2005).
- Barreto, V., Reina-San-Martin, B., Ramiro, A.R., McBride, K.M. & Nussenzweig, M.C. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol. Cell* **12**, 501–508 (2003).
- Ta, V.T. *et al.* AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat. Immunol.* **4**, 843–848 (2003).
- de Yébenes, V.G. *et al.* miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *J. Exp. Med.* **205**, 2199–2206 (2008).
- Dorsett, Y. *et al.* MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-IgH translocation. *Immunity* **28**, 630–638 (2008).
- McBride, K.M. *et al.* Regulation of class switch recombination and somatic mutation by AID phosphorylation. *J. Exp. Med.* **205**, 2585–2594 (2008).
- Pauklin, S., Sernandez, I.V., Bachmann, G., Ramiro, A.R. & Petersen-Mahrt, S.K. Estrogen directly activates AID transcription and function. *J. Exp. Med.* **206**, 99–111 (2009).
- Sernández, I.V., de Yébenes, V.G., Dorsett, Y. & Ramiro, A.R. Haploinsufficiency of activation-induced deaminase for antibody diversification and chromosome translocations both *in vitro* and *in vivo*. *PLoS One* **3**, e3927 (2008).
- Takizawa, M. *et al.* AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. *J. Exp. Med.* **205**, 1949–1957 (2008).
- Teng, G. *et al.* MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* **28**, 621–629 (2008).
- Arakawa, H., Saribasak, H. & Buerstedde, J.M. Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. *PLoS Biol.* **2**, e179 (2004).
- Conticello, S.G. *et al.* Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNBL1. *Mol. Cell* **31**, 474–484 (2008).
- Janz, S., Müller, J., Shaughnessy, J. & Potter, M. Detection of recombinations between c-myc and immunoglobulin switch alpha in murine plasma cell tumors and

- preneoplastic lesions by polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **90**, 7361–7365 (1993).
29. LaRue, R.S. *et al.* Guidelines for naming nonprimate APOBEC3 genes and proteins. *J. Virol.* **83**, 494–497 (2009).
 30. LaRue, R.S. *et al.* The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC Mol. Biol.* **9**, 104 (2008).
 31. Chaudhuri, J., Khuong, C. & Alt, F.W. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* **430**, 992–998 (2004).
 32. Basu, U. *et al.* The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* **438**, 508–511 (2005).
 33. Clarke, S.H. *et al.* Inter- and intraclonal diversity in the antibody response to influenza hemagglutinin. *J. Exp. Med.* **161**, 687–704 (1985).
 34. Berek, C. & Milstein, C. Mutation drift and repertoire shift in the maturation of the immune response. *Immunol. Rev.* **96**, 23–41 (1987).
 35. Celada, F. & Seiden, P.E. Affinity maturation and hypermutation in a simulation of the humoral immune response. *Eur. J. Immunol.* **26**, 1350–1358 (1996).
 36. Kepler, T.B. & Perelson, A.S. Somatic hypermutation in B cells: an optimal control treatment. *J. Theor. Biol.* **164**, 37–64 (1993).
 37. Kepler, T.B. & Perelson, A.S. Modeling and optimization of populations subject to time-dependent mutation. *Proc. Natl. Acad. Sci. USA* **92**, 8219–8223 (1995).
 38. Kleinstein, S.H., Louzoun, Y. & Shlomchik, M.J. Estimating hypermutation rates from clonal tree data. *J. Immunol.* **171**, 4639–4649 (2003).
 39. Arakawa, H. *et al.* Protein evolution by hypermutation and selection in the B cell line DT40. *Nucleic Acids Res.* **36**, e1 (2008).
 40. Cumbers, S.J. *et al.* Generation and iterative affinity maturation of antibodies *in vitro* using hypermutating B-cell lines. *Nat. Biotechnol.* **20**, 1129–1134 (2002).
 41. Wang, L., Jackson, W.C., Steinbach, P.A. & Tsien, R.Y. Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl. Acad. Sci. USA* **101**, 16745–16749 (2004).
 42. Bransteitter, R., Pham, P., Calabrese, P. & Goodman, M.F. Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. *J. Biol. Chem.* **279**, 51612–51621 (2004).
 43. Bennett, R.P. *et al.* APOBEC-1 and AID are nucleo-cytoplasmic trafficking proteins but APOBEC3G cannot traffic. *Biochem. Biophys. Res. Commun.* **350**, 214–219 (2006).
 44. Bennett, R.P., Presnyak, V., Wedekind, J.E. & Smith, H.C. Nuclear exclusion of the HIV-1 host defense factor APOBEC3G requires a novel cytoplasmic retention signal and is not dependent on RNA binding. *J. Biol. Chem.* **283**, 7320–7327 (2008).
 45. Stenglein, M.D., Matsuo, H. & Harris, R.S. Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. *J. Virol.* **82**, 9591–9599 (2008).
 46. Chen, H. *et al.* APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.* **16**, 480–485 (2006).

ONLINE METHODS

Bacterial mutation. For papillation assays, we transformed AID or APOBEC cDNAs in plasmid pTrc99 (ref. 47) into *E. coli* K12 strain CC102 (*ara(lac proB)*_{XIII} carrying an F *lacI⁻Z⁺proAB⁺* episome in which the *lacZ* carries a GAG→GGG missense mutation at codon 461 (ref 11)) and plated the cells on MacConkey-lactose agar (BD Biosciences) supplemented with ampicillin (100 µg ml⁻¹) and IPTG (1 mM). Plates were incubated at 37 °C for 3–6 d with papilli becoming visible after 3 d and their numbers increasing up until day 7. For analysis of arabinose-inducible expression, AID was expressed in plasmid pBAD30 (ref. 48).

The frequency of reversion of CC102 (pTrc99-AID or pTrc99-APOBEC) transformants to Lac⁺ was determined by plating cultures grown overnight to saturation in LB medium supplemented with ampicillin (100 µg ml⁻¹) and IPTG (1 mM) on M9 + 0.2% (w/v) lactose agar, whereas mutation to rifampicin resistance (50 µg ml⁻¹) was assessed following transformation into *E. coli* strain KL16 (Hfr (PO-45) *relA1 spoT1 thi-1*). We measured mutation frequencies by determining the median number of colony-forming cells that survived selection per 10⁷ viable cells plated with each median determined from 12 independent cultures. We determined the identity of mutations by sequencing PCR-amplified relevant sections of *lacZ* and *rpoB* (*lacZ*: forward, 5'-AGAATTCCTGAAGTTCAGATGT-3', and reverse, 5'-GGAATTCGAAACCGC CAAGAC-3'; *rpoB*: forward, 5'-TTGGCGAAATGGCGGAAAACC-3', and reverse, 5'-CACCGACGGATACCAC-CTGCTG-3').

PCR mutagenesis. The first and second generation human AID mutant libraries were generated by error-prone PCR using Taq polymerase (2.5 U; Bioline) on 1 ng of template DNA with 1 µM primer (forward, 5'-ATGGAATT CATGG-ACAGCCTCTTG-3', and reverse, 5'-CTGAAGCTTTCAAAGTCC CAAAGTA-3'), 250 µM dNTPs and 10 mM MgCl₂ in Taq buffer at 94 °C (2 min), followed by 30 cycles of 94 °C (30 s), 65 °C (30 s) and 72 °C (1 min). We generated the fugu AID and third-generation human AID mutant libraries using the Genemorph II Random Mutagenesis Kit (Stratagene) on 0.1 ng DNA template according to the manufacturer's instructions.

Antibody diversification. For assaying class switching, we analyzed surface IgG1 expression by flow cytometry in B cells (purified from AID^{-/-} mice and cultured in the presence of LPS+IL4 (48 h) following a 24 h-infection with AID-encoding retroviruses) as described⁴⁹. To facilitate a diminution in the extent of AID overexpression in the transduced B cells, a retroviral vector with a mutated Kozak sequence was used as described²¹. AID abundance in extracts prepared by heating cells (10⁶) in 50 µl of reducing SDS sample buffer was monitored by SDS-PAGE followed by western blot analysis using rabbit anti-AID antiserum (Abcam). GFP was detected using HRP-conjugated goat anti-GFP antiserum (Abcam).

We monitored AID-induced somatic mutation by measuring the frequency of sIgM loss variants in AID^{-/-} φV^{-/-} sIgM⁺ DT40 cells²⁶ transfected with AID-encoding vectors based on pExpressPuro2 (gift from J.-M. Buerstedde). For each construct, the percentage of sIgM⁻ cells was monitored in 12–24 independent transfectants that had been expanded under selection (0.25 µg ml⁻¹ puromycin) for 3 weeks before flow cytometry. Mutations in the IgVλ region

were characterized by sequencing genomic DNA that was PCR-amplified from either 100,000 unsorted or from (GFP⁺, sIgM⁻)-sorted cell equivalents⁵⁰.

Chromosomal translocations. B cells from AID-deficient mice were transduced with AID-expressing retrovirus and cultured in medium containing LPS (20 µg ml⁻¹) and IL4 (50 ng ml⁻¹) as described for the class-switching assays, followed by seeding of 8 × 10⁵ cell per ml in six-well plates. We subjected genomic DNA from 2 × 10⁵ cells that had been prepared using DirectPCR (Viotech) from sorted GFP⁺ cells 36 h after transduction to two rounds of nested PCR with Expand Long Template PCR system (Roche), followed by Southern blotting to amplify and detect both der12 c-myc-Igμ and der15 c-myc-Igμ translocations and the specific products, as described⁷.

Assaying deaminase activity. We purified GST-AID fusion proteins from pOPTG-AID transformants of the *E. coli* strain Rosetta (DE3) pLysS (pOPTG vector a gift from O. Perisic). Cells were grown at 37 °C in 2 × TY containing 100 µg ml⁻¹ ampicillin and 100 µM ZnCl₂ until the culture reached an absorbance of 0.8 at 600 nm, when it was induced with 1 mM IPTG for 16 h at 18 °C. The cells were then pelleted and lysed by a 30-min incubation on ice in lysis buffer (20 mM-Tris, pH 7.4, 100 mM NaCl, 0.1% (w/v) Triton X-100, 5 mM DTT, 4 µg ml⁻¹ RNase A and Roche complete EDTA-free protease inhibitor cocktail) followed by sonication. Cell lysates were clarified by centrifugation (95,000g for 1 h), and GST-AID was purified from these lysates by absorption onto glutathione-Sepharose (Amersham Pharmacia) at 4 °C for 5 h and elution, following extensive washing with lysis buffer supplemented with 50 mM reduced glutathione lacking Triton X-100. Eluted samples were stored at 4 °C for up to 1 week.

Deaminase activity of semipurified GST-AID (100–400 ng) was assayed at 37 °C in 10 µl of reaction buffer (8 mM Tris, pH 8.0, 8 mM KCl, 10 mM NaCl, 2.5 mM EDTA, 0.2 mM DTT, 5 µg RNase A and 0.4 U uracil-DNA glycosylase (NEB)) with 0.5 pmol oligodeoxyribonucleotide (fluorescein-5'-ATATGAA TA-GAATAGAGGGGTGAGCTGGGGTGAGCTGGGGTGAG-3'-biotin). Reactions were terminated at indicated times by addition of an equal volume of loading dye (formamide, 0.5 mM EDTA) and heating at 98 °C for 3 min. The resultant cleaved oligonucleotides were subjected to electrophoresis in 10% PAGE-urea gels, and fluorescence was detected with a Typhoon PhosphorImager (Molecular Dynamics). The extent of deamination was determined from the scanned images, expressing the pixel volume of the cleaved product bands (following background subtraction) as a percentage of the combined pixel volume of product and residual substrate bands.

47. Petersen-Mahrt, S.K., Harris, R.S. & Neuberger, M.S. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99–104 (2002).
48. Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
49. Di Noia, J.M. *et al.* Dependence of antibody gene diversification on uracil excision. *J. Exp. Med.* **204**, 3209–3219 (2007).
50. Sale, J.E., Calandrini, D.M., Takata, M., Takeda, S. & Neuberger, M.S. Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation. *Nature* **412**, 921–926 (2001).