

Activation-induced Cytidine Deaminase Deaminates 5-Methylcytosine in DNA and Is Expressed in Pluripotent Tissues

IMPLICATIONS FOR EPIGENETIC REPROGRAMMING*

Received for publication, July 8, 2004, and in revised form, September 7, 2004
Published, JBC Papers in Press, September 24, 2004, DOI 10.1074/jbc.M407695200

Hugh D. Morgan^{‡§}, Wendy Dean[‡], Heather A. Coker[¶], Wolf Reik^{‡**}
and Svend K. Petersen-Mahrt^{¶‡‡}

From the [‡]Laboratory of Developmental Genetics and Imprinting, Developmental Genetics Programme, The Babraham Institute, Cambridge CB2 4AT, United Kingdom, the [¶]DNA Editing Laboratory, Cancer Research UK, Clare Hall Laboratories, South Mimms EN6 3LD, United Kingdom, and the [¶]Protein and Nucleic Acid Chemistry Division, Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, United Kingdom

DNA deaminases of the Aid/Apobec family convert cytosine into uracil and play key roles in acquired and innate immunity. The epigenetic modification by methylation of cytosine in CpG dinucleotides is also mutagenic, but this is thought to occur by spontaneous deamination. Here we show that Aid and Apobec1 are 5-methylcytosine deaminases resulting in a thymine base opposite a guanine. Their action can thus lead to C → T transition mutations in methylated DNA, or in conjunction with repair of the T:G mismatch, to demethylation. The Aid and Apobec1 genes are located in a cluster of pluripotency genes including Nanog and Stella and are co-expressed with these genes in oocytes, embryonic germ cells, and embryonic stem cells. These results suggest that Aid and perhaps some of its family members may have roles in epigenetic reprogramming and cell plasticity. Transition in CpG dinucleotides is the most frequent mutation in human genetic diseases, and sequence context analysis of CpG transitions in the APC tumor suppressor gene suggests that DNA deaminases may play a significant role in tumor etiology.

DNA in multicellular organisms is generally maintained intact during development and in various adult tissues. Notable exceptions to this rule are recombination in lymphoid tissues and immunoglobulin diversification during B cell development. These are initiated by the Rag recombinase, and by Activation induced cytidine deaminase (Aid or Aicda), respectively. Aid is required both for somatic hypermutation and for class switch recombination of immunoglobulin genes (1, 2). Aid is related to Apobec1, a known cytidine deaminase that acts on a specific RNA (for apolipoprotein B), converting a C residue to a U resulting in a premature stop codon (3). Although it was initially envisaged that Aid would also act on RNA (2), subsequent studies have shown that Aid can act as a cytosine deaminase on DNA in *Escherichia coli* (4), yeast (5), vertebrate cells (2, 6), and *in vitro* (7–9). Importantly, loss of uracil DNA glycosylase

(UDG)¹, which repairs U:G mismatches in DNA, affects both hypermutation and class switching (1, 10, 11). Thus the current more widely accepted model envisages that Aid deaminates C to U in the DNA of immunoglobulin genes, leading to somatic mutations or class switch recombination depending on the subsequent pathway of DNA repair (1).

In addition to Aid and Apobec1, there are other members of this deaminase family, including Apobec2 and Apobec3g. Some of these can mutate DNA in *E. coli* (12) and single-stranded DNA *in vitro* (13–15). Apobec3g is also known as the cellular Cem15 protein and is critical for innate immunity to retroviruses, partly by mutating viral single-stranded DNA (14). Thus Aid and Apobec3g (and perhaps other members of the family) have roles in acquired and innate immunity by mutating cytosine in DNA. An ancestral role of DNA deaminases in protecting the organism against invading DNA may hence have been adapted to mutating endogenous DNA to better combat pathogens (1).

Another mechanism for the protection of the genome against invading DNA such as transposable elements is by methylating their DNA at CpG dinucleotides (16). This leads to transcriptional silencing of transposons so that they cannot continue to spread. The methylation of cytosine also increases the rate of mutation of C (17) so that methylated transposons accumulate mutations. It is generally thought that this occurs by spontaneous deamination (17), because enzymes that can deaminate 5-methylcytosine (5meC) have not been identified. As a result the CpG dinucleotide has become depleted in species with CpG methylation. In addition to its role in genome defense, methylation is important for epigenetic gene regulation. This includes parental imprinting, X chromosome inactivation, and possibly more general gene regulatory mechanisms in development and differentiation (18, 19).

Recently it has been found that genomic patterns of DNA methylation are reprogrammed genome-wide in early embryos and in primordial germ cells (19–21). Reprogramming may be necessary for resetting the epigenetic state of the gametic genomes so that totipotency is restored to the embryonic genome. Epigenetic reprogramming involves demethylation, which can occur passively as a result of DNA replication in the absence of DNA methyltransferase 1 (16, 18–21). However, the dramatic demethylation of the paternal genome in the zygote before replication of DNA (22–24), and the rapid demethylation in

* This work was supported by the Medical Research Council, Biotechnology and Biological Sciences Research Council, and Cancer Research United Kingdom. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Funded in part by a C. J. Martin Postdoctoral Fellowship.

** To whom correspondence should be addressed. E-mail: wolf.reik@bbsrc.ac.uk.

‡‡ Supported in part by the Arthritis Research Campaign.

¹ The abbreviations used are: UDG, uracil DNA glycosylase; 5meC, 5-methylcytosine; PGCs, primordial germ cells; EG, embryonic germ; ES, embryonic stem; E, embryonic day; mTDD, *Methanobacterium thermoautotrophicum*; RT, reverse transcription.

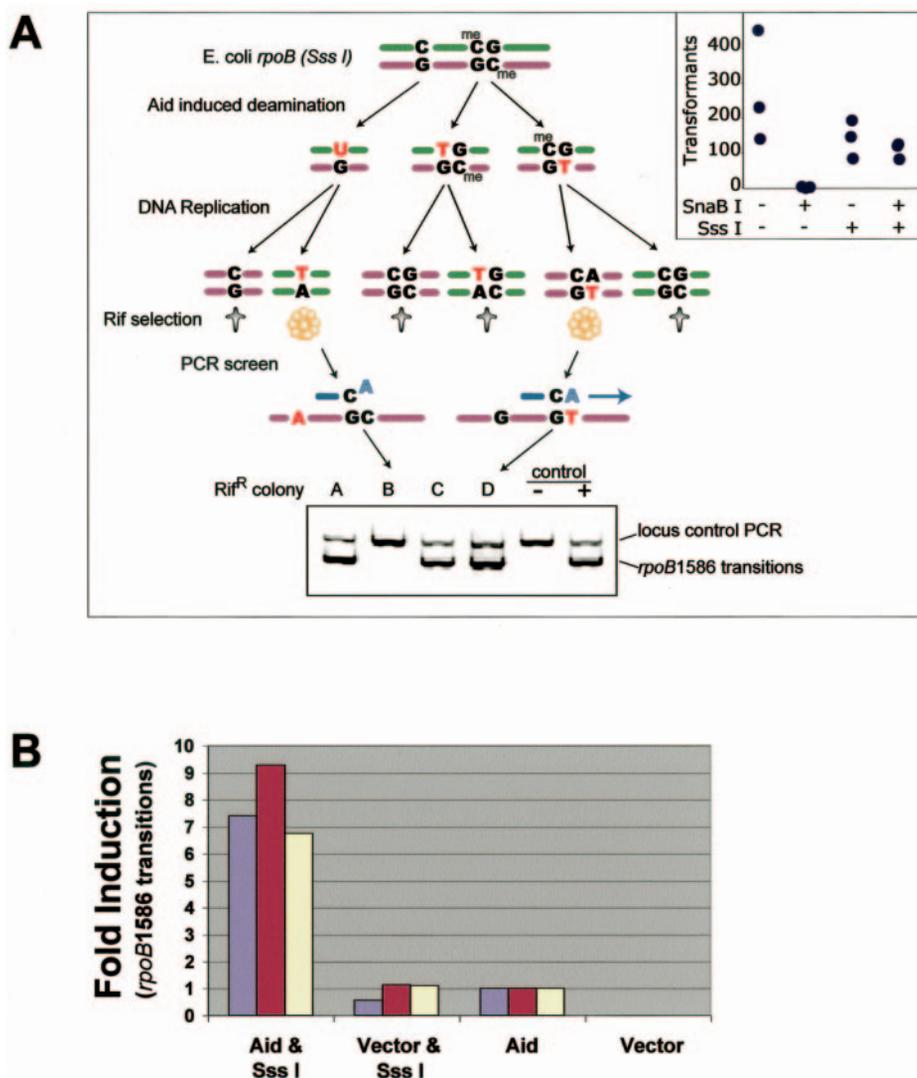


FIG. 1. Aid deaminates 5meC in *E. coli* DNA. A, design of mutation assay to detect 5meC deamination. Expression of Sss I methyltransferase in *E. coli* leads to methylation of chromosomal DNA at CpG dinucleotides, including in the selectable marker *rpoB* at positions 1585 and 1586. Deamination by Aid of C results in U, and deamination by Aid of 5meC results in T. Without repair and following DNA replication, mutations will appear in the *rpoB* gene some of which confer rifampicin resistance (*Rif^R*) (4), including transition mutations at 1586. Viable clones are screened for the transition mutation at 1586 using a single nucleotide detection PCR assay. Regardless of the mutation, the PCR produces a constant size band (*locus control PCR*), but if 1586 is mutated from 5meC to T an additional band is produced because of the annealing of the detection oligonucleotide (*rpoB1586 transitions*). For the minority of occasions where the CpG at 1586 is not methylated, deamination would lead to a uracil, which would also give a positive band in the PCR screening assay. A representative example of the PCR screen is shown. *Inset*, Sss I-induced *in vivo* CpG methylation protects restriction of a plasmid by methylation sensitive SnaBI. Plasmids with a single SnaBI restriction site were isolated from a culture that had been grown in the absence or presence of Sss I methyltransferase. The independent plasmids were digested with restriction endonuclease SnaBI. Digested or control undigested plasmids were used to transform *E. coli*. The number of transformants represents the number of uncut or protected plasmids. The average percent protection by Sss I is 82%, compared with less than 1% without Sss I. B, -fold increase in transitions at *rpoB1586* caused by expression of Aid and Sss I in three independent experiments. Values were calculated for each experiment from the fraction of median total number of transitions at *rpoB1586* to the viable cell number. The mutation rate by Aid alone was arbitrarily set to 1. The expression of Sss I (and vector) results in transitions at *rpoB1586* presumably because of spontaneous deamination of 5meC and lack of repair of the deaminated residue in *E. coli*. The expression of Aid in an Sss I background increases transitions at *rpoB1586* by ~8-fold.

primordial germ cells (25, 26) (PGCs) are likely to take place by a replication-independent mechanism. A number of mechanisms have been proposed for the loss of methylation (27), involving direct demethylation (28), oxidative demethylation (29), 5meC base excision repair (30), or nucleotide excision repair (27). However, no enzyme activity has been convincingly identified that can initiate these reactions (31–33).

We have asked whether cytosine deaminases can deaminate 5meC in DNA. We found that Aid and Apobec1 have 5meC deaminase activities, resulting in a thymine base opposite a guanine. If this mismatch is repaired, a methylated cytosine is replaced by an unmethylated one. If it is not repaired, it results in a C → T transition mutation. The expression of Aid has previously been thought to be limited to lymphoid tissues, but

embryonic tissues were not examined (34). Unexpectedly we found that Aid and Apobec1 are colocalized within a cluster of pluripotency genes and are expressed in oocytes and primordial germ cells, which undergo epigenetic reprogramming. This suggests new roles for DNA mutators in the physiology and pathology of DNA methylation.

EXPERIMENTAL PROCEDURES

Cloning and Expression of Sss I—The isopropyl 1-thio-β-D-galactopyranoside-inducible Sss I expression cassette from the pCAL7 (New England Biolabs) plasmid was transferred into pACYC177 (NEB) by standard cloning techniques (details on request). The resulting plasmid, p15ASss I Kan, was transformed into a methylation restriction-deficient *E. coli*, ER1821 (Elisabeth Raleigh, NEB) (*F-endA1 thi-1 supE44 spoT⁺ rfbD⁺ mcrA5 (mrr-hsdRMA-mcrB)1::IS10*) and checked

for isopropyl 1-thio- β -D-galactopyranoside-inducible methylation of DNA. We ensured the endogenous *rpoB* C1586 site could be methylated by bisulfite sequence analysis as described (23, 25) using *rpoB*-specific primers for bisulfite converted DNA (gtgtttttttgatagtagtagtagtg, ctcaattataatccaaacaacc). To quantitate the percentage methylation of a particular site, we isolated plasmid DNA from three independent cultures after induction of Sss I and digested them with CpG methylation-sensitive SnaBI. Digested or control undigested plasmids were transformed into a methylation non-restricting *E. coli* (ER1821). From this we were able to assess that *in vivo* methylation by Sss I caused protection of the restriction site on 82% of plasmids (unmethylated digested plasmid gave <1% transformation efficiency; see Fig. 1A). This was also consistent with the extent of methylation determined by agarose gel analysis of the digested plasmids (data not shown).

Screening for DNA Deamination at 5meCpG—Rifampicin screening was done as published (4) with modifications. Bacterial strains containing Aid with Sss I, Sss I without Aid (but with empty vector pTrc99A), Aid alone, and pTrc99A alone were induced with isopropyl 1-thio- β -D-galactopyranoside for 24 h in culture. Rifampicin-resistant colonies were selected on 100 μ g/ml rifampicin LB agar plates containing ampicillin (Aid plasmid) or kanamycin (Sss I plasmid) as appropriate. The median number of rifampicin resistant colonies/number of viable cells was determined for each experiment consisting of 20 cultures/category (Aid and Sss I, vector and Sss I, Aid, vector). The CpG to CpA transition at the target site in *rpoB* was identified by a PCR assay that differentiates between the single nucleotide G or A at this position (1586). The screening primer (*rpoB* G \rightarrow A: ctgagattacgcacaaactgca) with a 3'-A can only anneal when a G \rightarrow A transition has occurred at nucleotide position 1586 (1586G \rightarrow A), giving an expected product of 429 bp using *rpoB* reverse (caccgacggatcaccactgctg) as the second primer. For the locus control PCR, forward (*rpoB* forward: ttggcgaatggcggaaaacc) and reverse primers (*rpoB* reverse) always produce a band of 627 bp, irrespective of mutations, acting as an internal control for success of the PCR. The PCR screen was performed on two randomly selected rifampicin colonies from each of the cultures within a category. The frequency of the 5meC transition was calculated from the number of 5meC transitions (at *rpoB*1586) to the total number of mutations.

Preparation of Recombinant Aid and Apobec1—Human Aid was subcloned into a pET30 derived vector with a C-terminal His tag (plasmid sequences on request). Preparation of a rat Apobec1 expression vector was described previously (13). Production and purification of recombinant protein was as described (13).

UDG-based Deamination Assay—The assay was done as described previously (13) with minor modifications. Aid/Apobec1 samples (0.4–1 μ l) were incubated at 37 °C for 15 min in 10 μ l of buffer R (40 mM Tris, pH 8.0, 40 mM KCl, 50 mM NaCl, 5 mM EDTA, 1 mM dithiothreitol, 10% glycerol) with 2.5 pmol of 5'-biotinylated oligonucleotides, 3'-labeled with fluorescein. The oligonucleotides substrates SPM319-SPM330 are all based on 5'-biotin-ATAAGAATAGAATGAFFFFFA-ATGAATSSSS-SATGAATAGTA-fluorescein-3' where **F** and **S** are the first and second motifs indicated in Fig. 2B. Oligonucleotides were purified on streptavidin magnetic beads (Dynal) and washed at 70 °C. Deamination was monitored by incubating the bead-immobilized oligonucleotides at 37 °C for 1 h with excess uracil-DNA glycosylase (1 unit, New England Biolabs). Resultant cleaved oligonucleotides were subjected to 15–20% PAGE-urea gel electrophoresis, and the fluorescent signal was detected with a FLA-5000 scanner (Fuji). Percent conversions were calculated from the scanned images, percent conversion = pixel volume cleaved product (minus background)/pixel volume cleaved product (minus background) + pixel volume substrate (minus background). Data for rate calculations were based on numerous experiments on SPM320 and SPM348 for C deamination, and SPM347 and SPM359 for the 5meC deamination. Other oligonucleotides used: SPM348ATTATTGTTATTAGCTATTTGTTTATTGTTTATTATTT-fluor, SPM351ATTATTGTTATTAGCGATTGTTTATTGTTTATTATTT-fluor.

TDG-based Deamination Assay—Samples were treated as above, except that an excess (5 pmol) of the reverse complement oligonucleotide was added after the bead isolation reaction, heated to 90 °C, and allowed to anneal by cooling slowly to room temperature. Deamination of 5-methylcytosine was monitored by incubating the double-stranded, bead-immobilized oligonucleotides at 47 °C for 1 h, with excess thymine DNA-glycosylase (2 units, Trevigen). Products were resolved and analyzed as above. The 5meC containing oligonucleotides substrates SPM-356-SPM363 are all based on 5'-biotin-ATAAGAATAGAATGAFFFFG-AATGAATSSSSGATGAATAGTA-fluorescein-3' where **F** and **S** are the first and second motifs indicated in Fig. 2C, and each cytosine incorporated is methylated. Other oligonucleotides used were SPM279-ATTATTGTTATTAA^{me}CGATTTGTTTATTGTTTATTATTT-fluor,

TABLE I
Median mutation frequency ($\times 10^{-9}$), average -fold induction, and average percent transitions at *rpoB*1586 for three independent experiments

Each median mutation frequency was determined from >16 independent cultures.

	Experiment	Rif ^R	-Fold induction	Transitions at <i>rpoB</i> 1586
Vector	1	190		
	2	200	1	0
	3	93		
Aid	1	1600		
	2	1400	9	47
	3	1100		
Vector + Sss I	1	760		
	2	1300	7	55
	3	1000		
Aid + Sss I	1	7800		
	2	8400	45	72
	3	4800		

SPM281ATTATTGTTATTAG^{me}CGATTTGTTTATTGTTTATTATTT-fluor, and SPM347ATTATTGTTATTAG^{me}CTATTTGTTTATTGTTTATTATTT-fluor.

Semiquantitative RT-PCR—Total RNA was isolated from various mouse tissues using Qiagen miniRNAeasy reagents (Qiagen, Germany). Random hexamer and SuperscriptII (Invitrogen) reagents were used to make cDNA. Primers for Aid (forward caggcagcgatgagacct and reverse tcagctctggctgttcaca), Apobec1 (forward ctctgtcatgatctggatgtagtcacac and reverse catcgcagcaacataagctcc), and Hprt (forward ctctgtggattacattaagaact and reverse gtaaggcgatcatccaa) were used to amplify a PCR product from cDNA using Promega HotStarTaq. The amplification cycles were as follows: Aid and Apobec1, 95 °C 15 min, then 23–32 cycles of 94 °C 30 s, 65 °C 30 s, and 72 °C 30 s; Hprt, 95 °C 15 min, then 20–29 cycles of 94 °C 30 s, 55 °C 30 s, and 72 °C 30 s. For the amplification of Aid for 35–44 cycles, a nested approach was used with the first 20 cycles with outer primers (outer forward ttctttaccattcaaaaatgtccg and outer reverse aatagttttgaaggctgatgcc) at 95 °C 15 min, 20 cycles of 94 °C 30 s, 63 °C 30 s, and 72 °C 30 s, then 15–24 cycles as above for Aid. The products were resolved on a 2% agarose gel, stained, scanned on a Fuji FLA-3000, and analyzed using Aidi v3 two-dimensional densitometry software. For Fig. 3C the intensity of the Hprt signal at 29 cycles was arbitrarily set to 100 units, and the intensity of the Aid signal at 32 cycles is displayed for each tissue relative to Hprt.

RESULTS

Aid Mutates 5meC in *E. coli*—We were interested to see whether we could identify DNA deaminases that can deaminate 5meC. Using the deaminase motif we isolated cDNAs for candidate deaminases in the mouse genome, and screened these initially for expression in oocytes and other pluripotent tissues (data not shown). One of the cDNAs that was expressed in oocytes was Aid. We therefore tested Aid in a modified *E. coli* genetic system (4) which makes use of mutations of Cs in the *rpoB* gene leading to resistance to rifampicin. Aid mutates predominantly three Cs in the *rpoB* gene, one of which (position 1586) is in a CpG dinucleotide.

We expressed the CpG specific methyltransferase Sss I in *E. coli*; bisulfite sequencing confirmed that the selectable *rpoB* gene was methylated at CpG residues including C1586 (not shown), and *in vivo* CpG methylation by Sss I was able to protect 82% of sites from restriction (Fig. 1A). We determined the relative frequency of deamination at the methylated C1586 in *rpoB* using rifampicin selection and a single nucleotide discriminating PCR assay (Fig. 1A). If the methylated C were not a substrate for Aid, there would be no increase in its mutation rate over that with Sss I alone. However, the mutation rate at the C1586 position of *rpoB* in three independent experiments was ~8-fold higher in *E. coli* with both Sss I and Aid than in those having only Sss I (Fig. 1B, Table I). Thus this experimental system demonstrates Aid has a 5meC deaminase activity. It is not possible to directly compare the relative activities of Aid

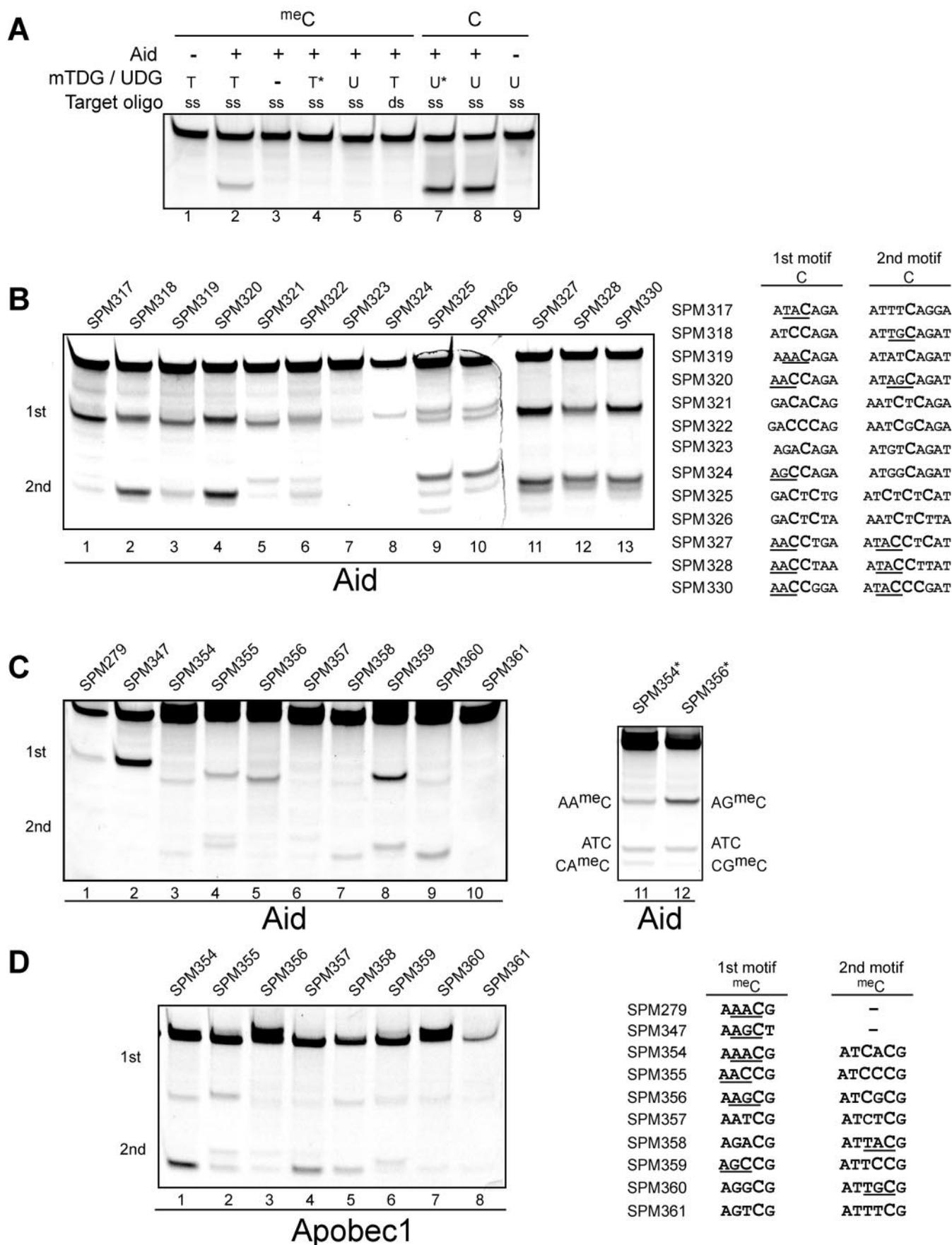


FIG. 2. Deamination of 5meC by DNA deaminases *in vitro*. Oligonucleotides fluorescently labeled at their 3'-end and containing 5-methylcytosine or cytosine were incubated with purified recombinant protein for 15 min. Deamination was monitored by cleavage with UDG (U) or mTDG (T). A, Aid deaminates 5meC. The oligonucleotide SPM347 containing 5meC (lanes 1-6) or oligonucleotide SPM348 containing C (lanes 7-9) were incubated with or without Aid protein. Deamination was monitored by incubating the product with mTDG or UDG and detecting cleavage

as a C deaminase (Aid *versus* vector background) to Aid as a 5meC deaminase (Aid and Sss I *versus* vector and Sss I), because the repair systems for U:G (without Sss I) and T:G (with Sss I) are different and their relative efficiencies are not known. The increase in frequency of mutation with expression of Sss I alone may be in part because of spontaneous deamination. Whether Apobec1 and Apobec3g can deaminate 5meC could not be tested in the *E. coli* assay, because they do not mutate the C1586 position.

Aid and Apobec1 Deaminate 5meC in Vitro—We sought to confirm the 5meC deaminase activity of Aid biochemically, using the purified protein in a previously established DNA deamination assay (13) (Fig. 2). Oligonucleotides were methylated at a single CpG, and upon incubation with Aid any T resulting from deamination of 5meC was detected using a thermostable thymine DNA glycosylase from *Methanobacterium thermoautotrophicum* (mTDG). Aid showed strong deamination activity (Fig. 2A), which was dependent on the presence of the protein (*lane 1 versus lane 2*) and the substrate being single-stranded (*lane 2 versus lane 6*). Specific detection of deaminated 5meC by mTDG requires double-stranded DNA (*lane 2 versus lane 4*). Because mTDG can act on either T:G or U:G mismatches (35) (data not shown), we needed to verify that no U was produced from C that had remained unmethylated. UDG (36) was unable to cleave the Aid-treated oligonucleotide containing 5meC, while being able to cleave a control unmethylated oligonucleotide treated with Aid (*lane 5 versus lane 8*). The methylated oligonucleotide reacted with Aid was also not cleaved by SMUG, an enzyme that recognizes a number of mismatches including U and 5-hydroxy meC, but not T (37) (data not shown). A catalytic inactive mutation of Aid (E58G) did not show any deamination activity toward either the unmethylated or methylated C (data not shown). These experiments clearly show that Aid produces T from 5meC in DNA by deamination.

DNA deaminases require particular sequence motifs as targets, on unmethylated DNA Aid prefers the sequence WRC (A/T, A/G, C) *in vivo* and *in vitro* (8, 15, 38), which we confirmed with a panel of oligonucleotides containing different target motifs (Fig. 2B). In a subset of these oligonucleotides all Cs were replaced by 5meCs, and we found that the WRC motif is still the preferred target (Fig. 2C), with a minor difference. Although in the unmethylated oligonucleotides all four possible WRC targets were deaminated to the same extent, once methylated, there does appear to be a preference for AGC (Fig. 2C, *lane 1 versus lane 2, lane 4 first motif versus lane 8 first motif*). From Fig. 2A we estimate that the activity of Aid on 5meC is ~3-fold lower than on C in the context AGC (SPM347 *lane 2, SPM348 lane 8*). Oligonucleotides containing both methylated and unmethylated targets demonstrate that the sequence context of each cytosine can render the deamination reaction of 5meC equivalent to C (Fig. 2C, *lane 11, AA5meC versus ATC*), or 5meC preferred to C (Fig. 2C, *lane 12, AGmeC versus ATC*). For the analysis in Fig. 2C we used three times as much enzyme than in Fig. 2B. From Coomassie and Western analysis (data not shown) we estimate a turnover of 60–70 fmol/min/ μ g for unmethylated (SPM320 and SPM348) and 20–25 fmol/

min/ μ g methylated C deamination (SPM347 and SPM359) based on a 15-min incubation of 2.5 pmol of substrate and an ~40% conversion rate using 1 and 3 μ g of Aid, respectively. Deaminase activity on 5meC by purified Aid has been observed previously (7).

Apobec1 has also been shown to have C deaminase activity but could not be tested for 5meC deaminase activity in the *E. coli* assay. We previously showed that *in vitro* Apobec1 has a preference for NTC (13, 15). Apobec1 was also able to deaminate 5meC, but the target preference was considerably altered to AC (Fig. 2D). Thus both Aid and Apobec1 have 5meC deaminase activity *in vitro* that is affected by the sequence context.

Aid and Apobec1 Genes Are Colocalized with Nanog and Stella and Are Expressed in Pluripotent Tissues—Aid is thought to be exclusively expressed in B cells undergoing class switch recombination or somatic hypermutation (34) and thus finding it expressed in oocytes in our initial screen was unexpected. However, bioinformatic analysis revealed a striking arrangement of the *Aid* and *Apobec1* genes within a 200-kb cluster of pluripotency genes, including *Nanog*, *Stella* (*Dppa3*), and *Gdf3* on mouse chromosome 6 (Fig. 3A). *Nanog* and *Stella* have important roles in embryonic stem (ES) cell identity (39, 40) and preimplantation development (41), respectively, and all three genes are expressed in pluripotent tissues. We therefore carried out semiquantitative RT-PCR analysis of Aid mRNA and found high levels of expression in oocytes and ovaries, and moderate levels in embryonic germ (EG) cells, ES cells, E12.5 PGCs (isolated from genital ridges using a green fluorescent protein-expressing transgene) that undergo methylation reprogramming (25, 26), and the genital ridges of E11.5–12.5 embryos containing PGCs (Fig. 3, B and C). Apobec1 mRNA was also found in ovaries, oocytes (albeit at lower levels than Aid), and ES cells (Fig. 3D). The expression profiles of the five genes in the cluster are summarized in Fig. 3E; all genes are transcribed in pluripotent tissues that can undergo epigenetic reprogramming (oocytes, EG/PGC, ES cells). The clustered organization of the five genes is also found in the human genome (data not shown). To what extent regulatory elements for expression in pluripotent tissues are shared between the genes needs to be determined. It is interesting to note that *Stella* and *Gdf3* are also expressed in lymphoid tissues (42).

DISCUSSION

We have identified Aid and Apobec1 as 5meC deaminases in DNA, which are co-organized and co-expressed with a cluster of pluripotency genes. These observations raise the hypotheses that Aid and other members of this family may have physiological roles in epigenetic reprogramming and perhaps pathological roles in contributing to transition mutations at CpGs in human genetic disease and cancer.

It is now firmly established that Aid, Apobec1, and Apobec3g can mutate cytosine in DNA in a variety of different systems, including mammalian cells (4–6, 12, 14). Because a considerable proportion of Cs in mammalian genomes is methylated in the CpG dinucleotide, it is of great importance to ask whether C deaminases can act on 5meC. Although *E. coli* cytosine deaminase can act on both free bases cytosine and 5-methylcy-

product by electrophoresis. The mTDG reaction requires a double-stranded mismatch; complementary oligonucleotides were annealed before the reaction (*lane 4*, no complementary oligonucleotide). UDG does not recognize the deamination product of 5meC (*lane 5*). Aid cannot deaminate 5meC in a double-stranded oligonucleotide (*lane 6*). An unmethylated single-stranded oligonucleotide reacted with Aid is recognized by UDG in single- or double-stranded form (*lanes 7 and 8*). B, a panel of unmethylated oligonucleotides (see “Experimental Procedures”) were incubated with recombinant Aid for 15 min, and products were analyzed with UDG as above. The preferred target sequence for deamination is WRC. C, a panel of oligonucleotides with all cytosines methylated (except *, see “Experimental Procedures”) were incubated with 3 \times as much recombinant Aid for 15 min, and products were analyzed with mTDG as above. *Lanes 11 and 12* were as *lanes 1–10* but with the oligonucleotides SPM354* and SPM356*, which contained a single unmethylated cytosine in the context ATC of the second target motif. D, the same panel of oligonucleotides with all cytosines methylated were incubated with purified Apobec1 for 15 min, and products were analyzed with mTDG as above.

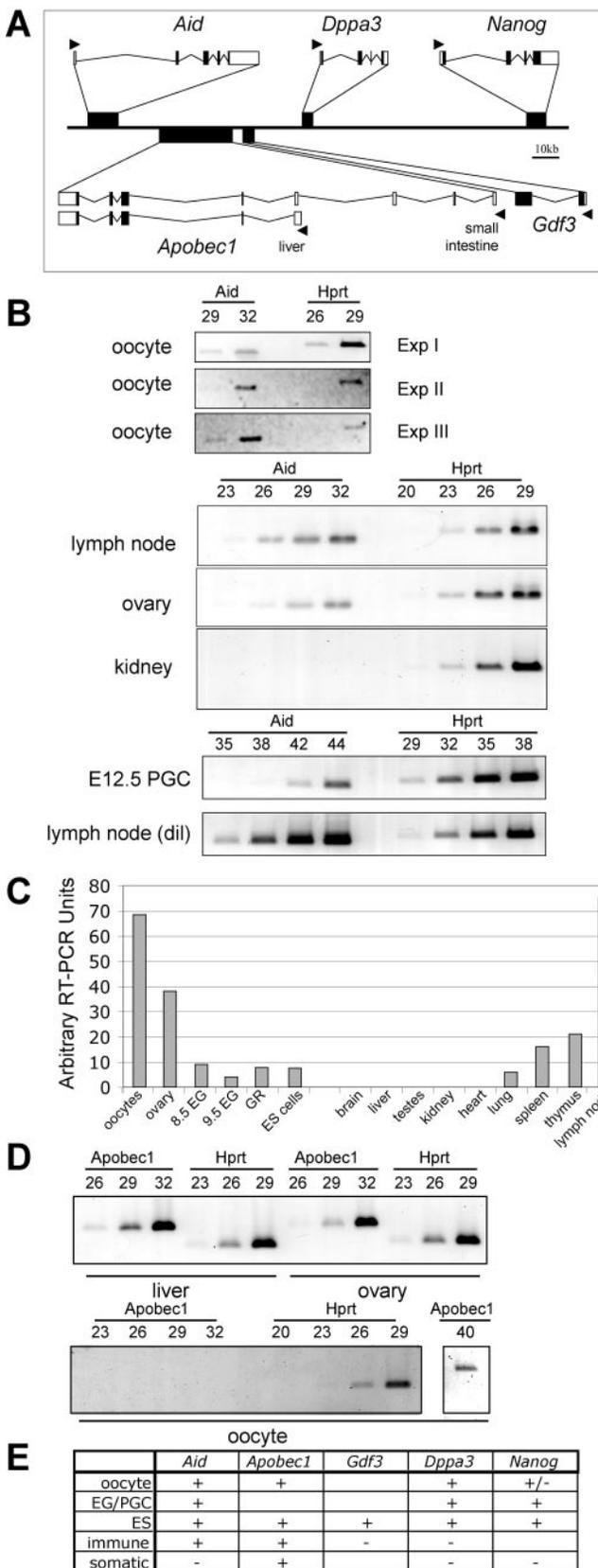


FIG. 3. Organization and expression of Aid and Apobec1 genes. A, region of mouse chromosome 6 (Build 32 from Ensembl mouse genome browser) containing *Aid* (also known as *Aicda*), *Apobec1* (with a small intestine and a liver and other tissues specific promoter), *Gdf3*, *Dppa3* (also known as *Stella*), and *Nanog* genes. Coding exons are filled. B, detection of Aid transcript in unfertilized oocytes, ovary, E12.5 PGCs, and a comparable amount of mesenteric lymph node RNA as judged by Hprt product amplification. Random-primed RT-PCR was performed on total RNA from mouse tissues with Aid- and Hprt-specific

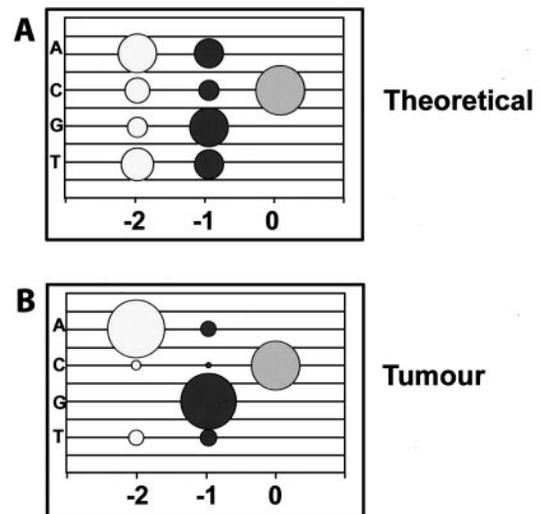


FIG. 4. Sequence context of CpG mutations in the APC gene. The APC gene has 1264 possible codons that can be mutated to a stop codon by a single base change. Of these at least 500 are near the N terminus of APC, giving rise to a premature stop codon with an established phenotypic loss of function (47), of these, 27 are in CpGs. A, the sequence context (-1 and -2 of mutated C) of all 27 targets. B, the sequence context of C transition mutations in sporadic tumors ($n = 73$). (Data derived from Thierry Soussi APC data base).

tosine (43), no prediction is possible for the distantly related Aid/Apobec deaminases, which work exclusively in the context of DNA/RNA. A methyl group in the 5-position of cytosine can certainly lead to steric hindrance for enzymes that act on DNA (for example, restriction enzymes). On the other hand, the deamination reaction by Aid is thought to occur by nucleophilic attack on position 4 of the pyrimidine ring of cytosine (1), and this is not necessarily hindered by the methyl group in position 5. Our results clearly established that Aid and Apobec1 deaminate 5meC in single-stranded DNA. It is difficult to compare directly their relative activities on C and 5meC, but our findings suggest they are in the same order of magnitude with other factors likely to modulate activities *in vivo*. The difference between the rate of spontaneous deamination (44) of 5meC and the enzymatically catalyzed one is thus 5–6 orders of magnitude (calculated from Refs. 7 and 13).

The DNA deaminase activity of Aid poses a potential danger to the genome in cells in which the enzyme is expressed. Deaminated 5meC is particularly vulnerable to mutation, because the resulting T can escape correction if not immediately repaired while mispaired with G; U can always be recognized as an error in DNA. Indeed, the ectopic expression of Aid and of Apobec1 in mice causes cancer (45, 46), and the preferred target sequences of the C deaminases are found around many C transitions in oncogenes and tumor suppressor genes in human tumors (15). Our work reveals the preferred sequence

primers. Total PCR cycles are shown; nested primers were used for Aid amplification of 35–44 cycles. C, the quantity of Aid product at 32 cycles (relative to Hprt product at 29 cycles set to 100) is shown for unfertilized oocytes, ovary, 8.5 and 9.5 days *post coitum* (dpc)-derived EG cell lines, 11.5 dpc whole genital ridge (GR, containing primordial germ cells), ES cells, adult lung, kidney, heart, 6-week male spleen, 6-week male thymus, mesenteric lymph node, brain, liver, and testes. D, detection of Apobec1 transcript in oocytes, ovary, and liver. RT-PCR was performed on total RNA using Apobec1- and Hprt-specific primers. E, summary of transcription profiles of genes in the Aid/Nanog cluster in pluripotent tissues (oocyte, EG/PGC, ES), immune system tissues, and other somatic tissues. Data are based on this work, EST databases, and published Northern, RT-PCR, *in situ* RNA hybridization, and knock-in reporter constructs.

context for Aid as a 5meC deaminase to be AG^{5meC}CG. We analyzed the sequence context of all CpG mutations leading to a premature stop codon of the APC tumor suppressor gene in colorectal cancer (47). The sequence context of the 27 possible CpGs is shown in Fig. 4A; yet in tumors, of 73 CpG mutations analyzed, 55 (75%) occurred in the preferred target for Aid, AG(C)G (Fig. 4B). This is an enrichment of over 25-fold over the expected frequency of random CpG mutations giving rise to a stop codon, and is consistent with the possibility that misregulated or mistargeted activity of Aid can cause CpG mutations in methylated DNA.

The expression of Aid (and Apobec1) in pluripotent tissues, particularly in ovulated oocytes and in primordial germ cells, was unexpected. The transcripts of Aid in some of these tissues reach the same level as in lymphoid tissues. Aid expression in non-lymphoid tissues is thus particularly targeted to the tissues that undergo large scale epigenetic reprogramming, including demethylation, during development. There is no confirmed enzymatic pathway for demethylation of DNA (31–33). Our observations suggest that Aid, Apobec1, and perhaps other members of this protein family play a role in epigenetic reprogramming. Targeted deamination of 5meC can be repaired by mismatch glycosylases such as TDG (35) or Mbd4 (ref. 48), whose action could also initiate regional nucleotide excision repair. Although Aid has been knocked out in mice, there is no published analysis of any developmental phenotypes of these mice (49). Of particular importance are the reproductive performance and offspring phenotypes when breeding from homozygous parents. Alternatively, there could be redundancy with other 5meC deaminases (including Apobec1); thus a careful genetic analysis *in vivo* of all of the components of the pathway discovered here is needed to unravel the precise role in development of DNA deaminases.

In vivo the activity and targeting of Aid is likely to be tightly regulated, probably involving control of nuclear trafficking (50, 51), sequence preference (15, 38), the requirement for single-stranded DNA, and chromatin modifications (52). Both transcription and chromatin remodeling (53) can lead to the exposure of single-stranded DNA. Thus the large scale chromatin remodeling at fertilization that removes protamines from the sperm genome and replaces them with histones is likely to lead to a widespread exposure of single-stranded DNA. Single-stranded DNA could also be exposed by transcription or chromatin remodeling occurring during reprogramming in PGCs.

It is possible that similar pathways play similar roles in organisms other than mammals. In the seed plant *Arabidopsis*, the DNA glycosylase Demeter is required for the expression of the imprinted genes *Medea* and *FWA* and acts antagonistically to DNA methylation (54), suggesting that a similar system (perhaps involving yet unidentified 5meC deaminases) plays a role in epigenetic reprogramming in plants. In *Neurospora*, duplicated DNA becomes methylated and mutated at cytosines, which is likely to involve deamination (55). It is intriguing that the only enzymes thus far identified as being able to modify bases in DNA as part of a physiological process are DNA methyltransferases and DNA deaminases. Thus an ancestral system using methylases and deaminases to protect the organism from invading DNA may have evolved to take on additional roles in acquired immunity and the regulation of epigenetic information.

Acknowledgments—We thank C. Rada, S. Conticello, T. Lindahl, and M. Neuberger for advice and help, S. Andrews for help with bioinformatics, G. Durcova-Hills and A. McLaren for EG cells, G. Echenery for technical assistance, I. Watt for APOBEC1, E. Raleigh for *E. coli* strains, the oligosynthesis group at Cancer Research UK, and F. Santos for discussion.

REFERENCES

- Neuberger, M. S., Harris, R. S., Di Noia, J., and Petersen-Mahrt, S. K. (2003) *Trends Biochem. Sci.* **28**, 305–312
- Honjo, T., Kinoshita, K., and Muramatsu, M. (2002) *Annu. Rev. Immunol.* **20**, 165–196
- Scott, J. (1995) *Cell* **81**, 833–836
- Petersen-Mahrt, S. K., Harris, R. S., and Neuberger, M. S. (2002) *Nature* **418**, 99–103
- Poltoratsky, V. P., Wilson, S. H., Kunkel, T. A., and Pavlov, Y. I. (2004) *J. Immunol.* **172**, 4308–4313
- Martin, A., Bardwell, P. D., Woo, C. J., Fan, M., Shulman, M. J., and Scharff, M. D. (2002) *Nature* **415**, 802–806
- Bransteitter, R., Pham, P., Scharff, M. D., and Goodman, M. F. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4102–4107
- Pham, P., Bransteitter, R., Petruska, J., and Goodman, M. F. (2003) *Nature* **424**, 103–107
- Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F. W. (2003) *Nature* **422**, 726–730
- Rada, C., Williams, G. T., Nilsen, H., Barnes, D. E., Lindahl, T., and Neuberger, M. S. (2002) *Curr. Biol.* **12**, 1748–1755
- Imai, K., Slupphaug, G., Lee, W. I., Revy, P., Nonoyama, S., Catalan, N., Yel, L., Forveille, M., Kavli, B., Krokan, H. E., Ochs, H. D., Fischer, A., and Durand, A. (2003) *Nat. Immunol.* **4**, 1023–1028
- Harris, R. S., Petersen-Mahrt, S. K., and Neuberger, M. S. (2002) *Mol. Cell.* **10**, 1247–1253
- Petersen-Mahrt, S. K., and Neuberger, M. S. (2003) *J. Biol. Chem.* **278**, 19583–19586
- Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., Neuberger, M. S., and Malim, M. H. (2003) *Cell* **113**, 803–809
- Beale, R. C., Petersen-Mahrt, S. K., Watt, I. N., Harris, R. S., Rada, C., and Neuberger, M. S. (2004) *J. Mol. Biol.* **337**, 585–596
- Bestor, T. H. (2000) *Hum. Mol. Genet.* **9**, 2395–2402
- Jones, P. A., and Baylin, S. B. (2002) *Nat. Rev. Genet.* **3**, 415–428
- Jaenisch, R., and Bird, A. (2003) *Nat. Genet.* **33**, 245–254
- Li, E. (2002) *Nat. Rev. Genet.* **3**, 662–673
- Reik, W., Dean, W., and Walter, J. (2001) *Science* **293**, 1089–1093
- Surani, M. A. (2001) *Nature* **414**, 122–128
- Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000) *Nature* **403**, 501–502
- Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000) *Curr. Biol.* **10**, 475–478
- Dean, W., Santos, F., Stojkovic, M., Zakhartchenko, V., Walter, J., Wolf, E., and Reik, W. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 13734–13738
- Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M. A. (2002) *Mech. Dev.* **117**, 15–23
- Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A., and Ishino, F. (2002) *Development* **129**, 1807–1817
- Cedar, H., and Verdine, G. L. (1999) *Nature* **379**, 568–569
- Bhattacharya, S. K., Ramchandani, S., Cervoni, N., and Szyf, M. (1999) *Nature* **397**, 579–583
- Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., and Sedgwick, B. (2002) *Nature* **419**, 174–178
- Zhu, B., Zheng, Y., Hess, D., Anglikler, H., Schwarz, S., Siegmund, M., Thiry, S., and Jost, J. P. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5135–5139
- Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999) *Nat. Genet.* **23**, 58–61
- Wade, P. A., Geggion, A., Jones, P. L., Ballestar, E., Aubry, F., and Wolffe, A. P. (1999) *Nat. Genet.* **23**, 62–66
- Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002) *Dev. Biol.* **241**, 172–182
- Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O., and Honjo, T. (1999) *J. Biol. Chem.* **274**, 18470–18476
- Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2003) *Nucleic Acids Res.* **31**, 2261–2271
- Lindahl, T. (1976) *Nature* **259**, 64–66
- Boorstein, R. J., Cummings, A., Jr., Marenstein, D. R., Chan, M. K., Ma, Y., Neubert, T. A., Brown, S. M., and Teebor, G. W. (2001) *J. Biol. Chem.* **276**, 41991–41997
- Yu, K., Huang, F. T., and Lieber, M. R. (2004) *J. Biol. Chem.* **279**, 6496–6500
- Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003) *Cell* **113**, 631–642
- Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003) *Cell* **113**, 643–655
- Payer, B., Saitou, M., Barton, S. C., Thresher, R., Dixon, J. P., Zahn, D., Colledge, W. H., Carlton, M. B., Nakano, T., and Surani, M. A. (2003) *Curr. Biol.* **13**, 2110–2117
- Clark, A. T., Rodriguez, R. T., Bodnar, M. S., Abeyta, M. J., Cedars, M. I., Turek, P. J., Firpo, M. T., and Reijo Pera, R. A. (2004) *Stem Cells* **22**, 169–179
- Cohen, S. S., and Barner, H. D. (1957) *J. Biol. Chem.* **226**, 631–642
- Ehrlich, M., Norris, K. F., Wang, R. Y., Kuo, K. C., and Gehrke, C. W. (1987) *Biosci. Rep.* **6**, 387–393
- Okazaki, I. M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003) *J. Exp. Med.* **197**, 1173–1181
- Yamanaka, S., Balestra, M. E., Ferrell, L. D., Fan, J., Arnold, K. S., Taylor, S., Taylor, J. M., and Innerarity, T. L. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 8483–8487
- Beroud, C., and Soussi, T. (1996) *Nucleic Acids Res.* **24**, 121–124
- Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P. D., Bishop, S. M., Clarke, A. R., and Bird, A. (2002) *Science*

- 297, 403–405
49. Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000) *Cell* **102**, 553–563
50. Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M., and Honjo, T. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1975–1980
51. McBride, K. M., Barreto, V., Ramiro, A. R., Stavropoulos, P., and Nussenzweig, M. C. (2004) *J. Exp. Med.* **199**, 1235–1244
52. Nambu, Y., Sugai, M., Gonda, H., Lee, C. G., Katakai, T., Agata, Y., Yokota, Y., and Shimizu, A. (2003) *Science* **302**, 2137–2140
53. Havas, K., Flaus, A., Phelan, M., Kingston, R., Wade, P. A., Lilley, D. M., and Owen-Hughes, T. (2000) *Cell* **103**, 1133–1142
54. Xiao, W., Gehring, M., Choi, Y., Margossian, L., Pu, H., Harada, J. J., Goldberg, R. B., Pennell, R. I., and Fischer, R. L. (2003) *Dev. Cell* **5**, 891–901
55. Freitag, M., Williams, R. L., Kothe, G. O., and Selker, E. U. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8802–8807

**Molecular Basis of Cell and
Developmental Biology:
Activation-induced Cytidine Deaminase
Deaminates 5-Methylcytosine in DNA and
Is Expressed in Pluripotent Tissues:
IMPLICATIONS FOR EPIGENETIC
REPROGRAMMING**

Hugh D. Morgan, Wendy Dean, Heather A.
Coker, Wolf Reik and Svend K.
Petersen-Mahrt

J. Biol. Chem. 2004, 279:52353-52360.

doi: 10.1074/jbc.M407695200 originally published online September 24, 2004

Access the most updated version of this article at doi: [10.1074/jbc.M407695200](https://doi.org/10.1074/jbc.M407695200)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](https://www.jbc.org/).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 53 references, 21 of which can be accessed free at
<http://www.jbc.org/content/279/50/52353.full.html#ref-list-1>