

The activation-induced cytidine deaminase (AID) efficiently targets DNA in nucleosomes but only during transcription

Hong Ming Shen,¹ Michael G. Poirier,^{3,4} Michael J. Allen,² Justin North,^{3,4} Ratnesh Lal,² Jonathan Widom,⁵ and Ursula Storb¹

¹Department of Molecular Genetics and Cell Biology and ²Department of Medicine, University of Chicago, Chicago, IL 60637

³Department of Physics and ⁴Department of Biochemistry, Ohio State University, Columbus, OH 43210

⁵Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston, IL 60208

The activation-induced cytidine deaminase (AID) initiates somatic hypermutation, class-switch recombination, and gene conversion of immunoglobulin genes. *In vitro*, AID has been shown to target single-stranded DNA, relaxed double-stranded DNA, when transcribed, or supercoiled DNA. To simulate the *in vivo* situation more closely, we have introduced two copies of a nucleosome positioning sequence, MP2, into a supercoiled AID target plasmid to determine where around the positioned nucleosomes (in the vicinity of an ampicillin resistance gene) cytidine deaminations occur in the absence or presence of transcription. We found that without transcription nucleosomes prevented cytidine deamination by AID. However, with transcription AID readily accessed DNA in nucleosomes on both DNA strands. The experiments also showed that AID targeting any DNA molecule was the limiting step, and they support the conclusion that once targeted to DNA, AID acts processively in naked DNA and DNA organized within transcribed nucleosomes.

CORRESPONDENCE

Ursula Storb:
stor@uchicago.edu

Abbreviations used: AFM, Atomic force microscopy; AID, activation-induced cytidine deaminase; cDNA, complementary DNA; SHM, somatic hypermutation.

The somatic hypermutation (SHM) of *Ig* genes requires that the *Ig* genes are transcribed (for review see reference 1). The track of mutations coincides roughly with that of the transcribed 1–2 kb after the start site of transcription (2). The SHM process is initiated by the activation-induced cytidine deaminase (AID), which targets deoxycytidines (Cs) in cell-free assays in single-stranded DNA (3–5) and mainly in the non-transcribed strand during transcription of linear double-stranded DNA (6–10) but targets Cs on both DNA strands when they are flipped out in supercoiled DNA (11, 12). *In vivo*, AID targets both DNA strands, as shown in germinal center B cells of mice with deficiencies in the uracil (U) glycosylase Ung (13) or double-deficient in Ung, and mismatch repair proteins Msh2 or Msh6 (14, 15). In these DNA repair-deficient mice, a large proportion or all of the Us created by AID are converted to thymines (Ts) as a result of direct replication of the U-containing DNA strand. Both DNA strands show very similar ratios of C to T, thus agreeing with the *in vitro* findings that AID can equally target both DNA strands in supercoiled DNA. This finding has implications for the mechanism by

which AID accesses double-stranded DNA *in vivo*. We have hypothesized that AID may target negative supercoils, as they arise in the wake of the transcription complex during transcript elongation (11, 12). Another possibility is that AID targets both DNA strands in the very short interval of single-stranded DNA at the upstream end of the transcription bubble, where nascent RNA has dissociated from the transcribed strand but the two DNA strands have not yet reannealed. A third possibility can be suggested from the finding that the single-strand binding protein RPA is associated with AID in extracts of mutating B cells (7). Perhaps the transcription bubble remains open for an extended length because the two DNA strands are held apart by RPA.

Considering the situation during SHM *in vivo*, AID has to negotiate not only supercoiled DNA but also DNA that is organized into chromatin. Recent whole genome analyses have

© 2009 Shen et al. This article is distributed under the terms of an Attribution-Noncommercial-Share Alike-No Mirror Sites license for the first six months after the publication date (see <http://www.jem.org/misc/terms.html>). After six months it is available under a Creative Commons License (Attribution-Noncommercial-Share Alike 3.0 Unported license, as described at <http://creativecommons.org/licenses/by-nc-sa/3.0/>).

shown that although transcriptional promoters generally have greatly reduced concentrations of nucleosomes, there is a stable nucleosome positioned at the start of transcription, and the genome, including transcribed genes, is overall constrained in nucleosomes (16–18). During transcript elongation, nucleosomes are partially or completely removed and reassembled with the DNA after passage of the RNA polymerase (for review see reference 19). Relating these findings to SHM, it was possible that AID would only target DNA sequences that are not engaged in nucleosomes, such as the spacers between nucleosomes or DNA sequences that are temporarily vacated by nucleosomes during transcription. Alternatively, the conformational changes DNA undergoes when wrapped around the histone core (20, 21) may actually flip out cytosines and make them accessible to AID even without transcription. The latter possibility was suggested by the finding of mutation clusters of approximately nucleosome size interspersed with unmutated regions of approximately spacer size in an Ig transgene (22).

A special case of nucleosome topology is the stable positioning of a nucleosome over the region where transcription starts (16–18). Perhaps a nucleosome at the start, in the context of SHM, may be responsible for the lack of mutations at the start of the transcribed region. In general, the first ~100 bp of Ig genes are not mutated, increasing mutation frequencies are seen in the next ~100 bp to a plateau of mutation frequencies that is reached ~200 bp from the start of transcription (15, 23–25). It is thus important to determine whether AID can deaminate cytosines in nucleosomal DNA. To answer this question, we have created a plasmid, *pKMP2*, with two 147-bp nucleosome positioning sequences, MP2 (26), placed between a T7 promoter and the start of an ampicillin resistance (*Amp*^r) gene (Fig. 1 A). Histone octamers preferentially bind these positioning sequences, leaving a nucleosome-free spacer of 76 bp between them. The *Amp*^r gene was inactivated by changing the translational initiation ATG codon to ACG, thus requiring cytosine deamination by AID for activation. The *pKMP2* plasmid also has a kanamycin resistance (*Kan*^r) gene so that plasmid-carrying *Escherichia coli* can be selected in kanamycin whether or not the *Amp*^r gene has been activated by AID. This allows assessing mutations in and near the *Amp*^r gene independent of *Amp*^r function.

RESULTS

Nucleosomes can be positioned on supercoiled plasmid DNA

The map of the 3,898-bp plasmid *pKMP2* with the two nucleosome-positioning MP2 sequences is shown in Fig. 1 A. The MP2 sequence causes precise binding of histone octamers as shown by exonuclease digestion (Fig. 2). This is not seen when naked DNA is treated with exonuclease, confirming that none of the positions observed with nucleosomal DNA are caused by exonuclease pause sites that occur on naked DNA (unpublished data). We created 247-bp DNA molecules with a central MP2 positioning sequence and 50-bp arms. These molecules were reconstituted with histone octamer and purified on a sucrose gradient. Exonuclease III mapping was performed

and the digested fragments compared with a sequencing ladder that was prepared with the same 247-bp DNA fragment. There is a single exo-resistant fragment within this DNA construct, which proves that the MP2 sequence positions nucleosomes extremely accurately. Nucleosome-free DNA digested separately showed that exonuclease III does not stall at the position observed for nucleosomes (unpublished data). Each MP2 sequence contains a BamHI site that is not accessible to

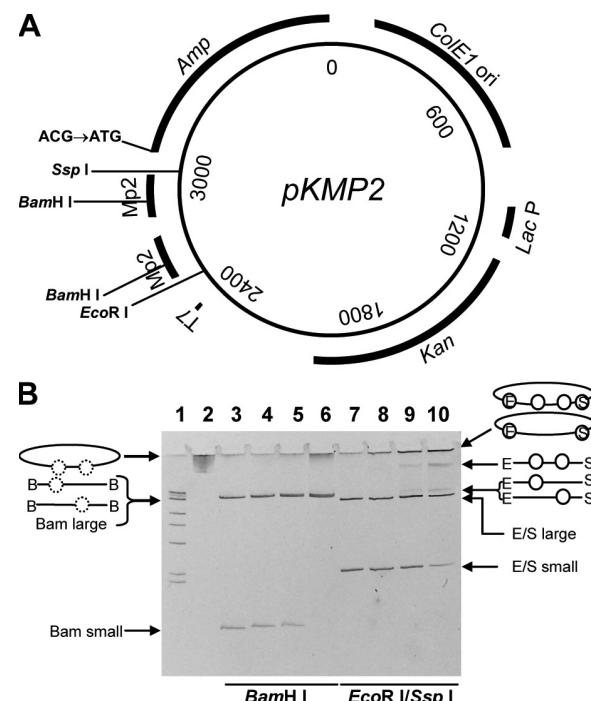


Figure 1. The *pKMP2* plasmid and its charging with nucleosomes. (A) Mutation substrate *pKMP2* was used to determine AID-induced cytosine deaminations in the *Amp*^r gene. T7, T7 RNA polymerase promoter for transcription in vitro; MP2, nucleosome positioning sequences; *Amp*, *Amp*^r gene coding region whose initiation codon ATG has been changed to ACG; ColE1 ori, LacP, and *Kan*, origin of replication, Lac promoter, and coding region of the *Kan*^r gene for expression of kanamycin resistance in *E. coli* transformed with *pKMP2*. Restriction sites used in this study are indicated. (B) Positioned nucleosomes formed on supercoiled *pKMP2* plasmid, naked plasmid DNA, and nucleosome containing plasmids with and without restriction enzyme digestion were electrophoresed on a 5% polyacrylamide gel. Lane 1 shows size markers of 1-kb DNA ladder with 500 bp, 517 bp, 1 kbp, 1.5 kbp, 2 kbp, 3 kbp, and 4 kb resolved on the gel. The final band contains four sizes: 5-, 6-, 8-, and 10-kbp DNA molecules. Lanes 2 and 3 show *pKMP2* naked DNA uncut or cut with BamHI. Lanes 4–6 show *pKMP2* nucleosomes cut with BamHI (0.8, 1.5, and 3.1 μ g of histone octamers per 10 μ g of DNA, respectively). Histones were removed after BamHI treatment. Lanes 7–10 show *pKMP2* naked DNA or nucleosomes (as in lanes 4–6, but histones were not removed after restriction enzyme digestion) cut with EcoRI and SspI. The conformation of the plasmids without and with nucleosomes (circles) are shown. On the left, images for lanes 4–6 are shown. Dotted circles indicate positioned nucleosomes that were removed by removal of histones after BamHI (B) digestion. On the right, images for lanes 8–10 are shown. Circles indicate nucleosomes in EcoRI (E)- and SspI (S)-digested plasmids. The data in B were repeated five times with similar results.

the restriction enzyme when the DNA is positioned on core histones (26).

When the supercoiled *pKMP2* plasmid was treated with increasing concentrations of histone octamers, the MP2 sequences became less and less accessible to BamHI (Fig. 1 B). In Fig. 1 B, lanes 2 and 3 are plasmid *pKMP2* DNA without histones and without (lane 2) or with (lane 3) BamHI digestion. BamHI creates two DNA fragments: one of 223 bp corresponding to the distance between the two BamHI sites in the positioning sequences and one of 3,675 bp corresponding to the rest of the plasmid. Lanes 4–6 of Fig. 1 B show BamHI-digested plasmids with increasing amounts of histone octamers (the plasmid DNAs were repurified after BamHI digestion). In lane 4 of Fig. 1 B, essentially all the DNA is cut with BamHI, resulting in a small and a large fragment as with naked DNA. In lane 5 of Fig. 1 B, completely undigested DNA becomes apparent, but plasmids without any positioned nucleosome are still present (small fragment), as are plasmids with a single positioned nucleosome (the large fragment is of greater intensity and presumably is a doublet of 3675 and 3898 bp). Finally, in lane 6 of Fig. 1 B, all plasmids have at least one positioned nucleosome, resulting in the absence of the small fragment. About 50% of the DNA is undigested, thus containing two positioned nucleosomes. However, at this concentration of histones, the plasmids are

not saturated with positioned nucleosomes, as shown by the presence of the large 3,898-bp fragment, representing DNA with one positioned nucleosome cut with BamHI in the other positioning sequence. Lanes 7–10 of Fig. 1 B confirm the increasing loading of nucleosomes. The DNAs were digested with EcoRI and SspI and were run directly on the gel without removing proteins. EcoRI and SspI digestion releases two fragments of 548 and 3,350 bp (Fig. 1 A). The histone-containing DNA samples in lanes 8–10 of Fig. 1 B show the small and large fragments as naked DNA (lane 7), indicating that they do not contain histones. At the highest concentration of histones (Fig. 1 B, lane 10), >50% of the molecules do not contain nucleosomes on the large fragment (E/S large). The E/S large band is only somewhat reduced in intensity. However, the small fragment in lane 10 of Fig. 1 B is reduced to ~30% of the amount in naked DNA (lane 7) and includes considerably less DNA relative to the large fragment in naked DNA, showing that the MP2 sequences are preferred binding sites for histone cores, compared with the rest of the plasmid. The small fragment is shifted to positions indicating loading with one or two nucleosomes, respectively, as indicated (Fig. 1 B). Most important are the E-o-o-S fragments showing the presence of two nucleosomes. Because they are flanked with both EcoRI and SspI sites, the short 548-bp DNA fragment clearly contains two nucleosomes

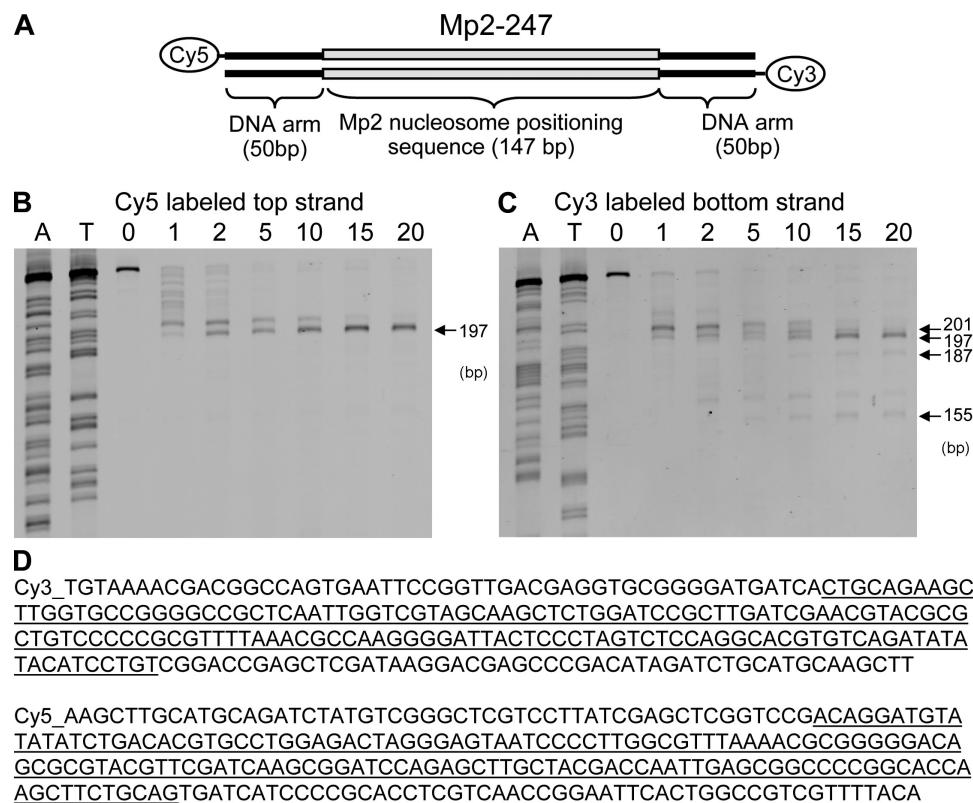


Figure 2. Exonuclease tests for nucleosome positioning on the MP2 sequence. (A) Map of the nucleosome and flanking DNA. (B and C) Exonuclease assays (details in Materials and methods). (D) Bottom and top strand sequences. The underlined sequence is MP2. The experiment was done twice with similar results.

and is flanked by naked DNA that permits digestion by EcoRI and SspI.

The DNA in the well (Fig. 1 B, lanes 8–10) is uncut circular plasmid that does not enter this polyacrylamide gel, presumably molecules with nucleosomes occupying the EcoRI and/or SspI sites. We did not attempt to inhibit histone octamer binding to the sequences outside of the MP2 sequences. However, they are expected to be random and rare compared with binding at the MP2 sequences (Fig. 1 B, lane 10). There are 129 and 50 bp between the MP2 positioning sequence and the EcoRI and SspI sites, respectively. Because spacers between nucleosomes can be as short as 30 bp (26), it is likely that in the DNA molecules trapped in the well the MP2 sequences are also positioned within nucleosomes. Thus, digestion with EcoRI and SspI supports the conclusion with BamHI digestion that ~50% of the DNA in lane 10 of Fig. 1 B contains two positioned nucleosomes.

To further confirm that dinucleosome DNA was formed, we used atomic force microscopy (AFM) to visualize nucleosome DNA. Fig. 3 A shows two adjacent nucleosomes in the supercoiled plasmid. By volume and surface topography (illustrated in this study with light reflection), two nucleosomes are distinguished from a single one (Fig. 3, B and C). The single-line traces confirm the nucleosome numbers (Fig. 3 D). These analyses show that supercoiled DNA can be charged with histone octamers. Large numbers of plasmids with two positioned

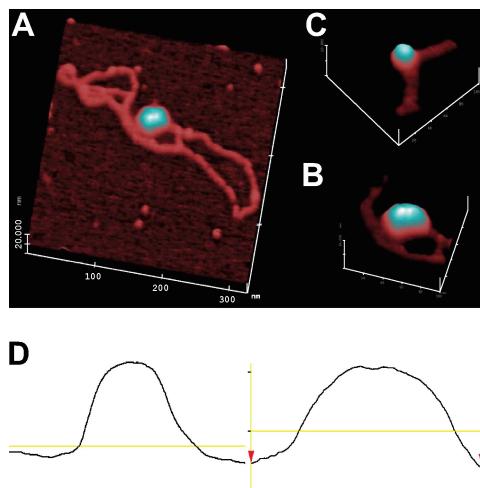


Figure 3. AFM of pKMP2 plasmids with nucleosomes. (A) Full-length plasmid (DNA contour length measured to be ~1.1 μ m or ~3.2 kB). The magnification is 200,000 \times , with a 20-nm full-scale z axis. Using the AFM offline software, the full contour length of the free DNA was segmented into ~50 linear segments, each ~20 nm in length, and the segment lengths were tallied. (B) Portion of another pKMP2 plasmid with a single nucleosome. The magnification is 300,000 \times , with a 10-nm full-scale z axis. (C) The magnification is 300,000 \times , with a 10-nm full-scale z axis of a portion of the plasmid shown in A. (D) Single-line topographic AFM traces of the mononucleosome (left) and dinucleosome (right) as shown in B and C, respectively. The trace of the dinucleosome shows a bipartite topography and is roughly twice the size of the mononucleosome. The experiment was done twice with similar results.

nucleosomes can be obtained as useful substrates for potential AID targeting.

AID efficiently targets DNA in nucleosomes only during transcription

The experimental procedures to test AID function are shown in Fig. 4 A. To analyze the interaction of AID with DNA in nucleosomes, we used the highest histone octamer to DNA ratio (Fig. 1 B, lanes 6 and 10). To exclude DNA molecules that did not have two positioned nucleosomes, the DNA was treated with BamHI for 1 h before addition of AID or of AID and T7 RNA polymerase. BamHI continued to be present throughout the subsequent incubation with AID. BamHI treatment of naked DNA completely eliminated all viable plasmids, as no Kan^r colonies were obtained after transformation

A Flowchart of Experimental Procedures

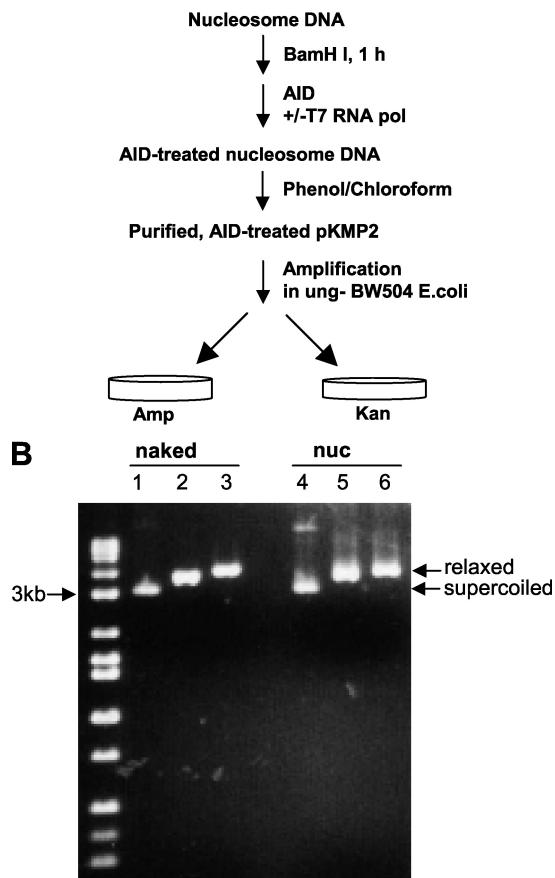


Figure 4. Experimental procedure to test AID activity in plasmid pKMP2. (A) Flowchart. (B) DNA with nucleosomes is supercoiled. Naked, naked DNA; Nuc, DNA with nucleosomes. Nucleosome-containing DNA was purified by phenol-chloroform extraction and electrophoresed directly (lane 4; the DNA is supercoiled; the high band is nicked DNA) or electrophoresed after restriction enzyme digestion (lanes 5 and 6; the DNA is relaxed). Lanes 1 and 4, no enzyme; 2 and 5, EcoRI; 3 and 6, Mscl. Size markers are 0.3, 0.4, 0.5, 0.75, 1, 1.4, 1.55, 2, 3, 4, and 6–10 kb. The experiments summarized in A and shown in B were done five times and once, respectively.

into U DNA glycosylase-deficient *E. coli* (unpublished data). When nucleosomal DNA was pretreated with BamHI and further treated continuously for 2 h together with AID, 10% viable plasmids were retained under kanamycin selection. This is lower than the 50% estimate from the restriction enzyme analyses shown in Fig. 1 B, presumably because in those tests 5× less BamHI and a 4.5× shorter time of incubation with the restriction enzymes were used. During prolonged incubation at 37°C, nucleosome “breathing” allows access to restriction enzymes (26). For mutation analysis of nucleo-

some-containing plasmids with AID in the presence of BamHI (Fig. 5), two positioned nucleosomes must have been constantly present. As shown by Poirier et al. (26), nucleosomes are resistant to cleavage with certain restriction enzymes contained within their sequences. Thus, plasmids cut by BamHI will not be sampled in our experiments because only intact circular plasmids are able to provide ampicillin or kanamycin resistance to *E. coli* (unpublished data).

When naked supercoiled *pKMP2* DNA was treated with AID, ampicillin resistance was readily obtained (Fig. 5 C and

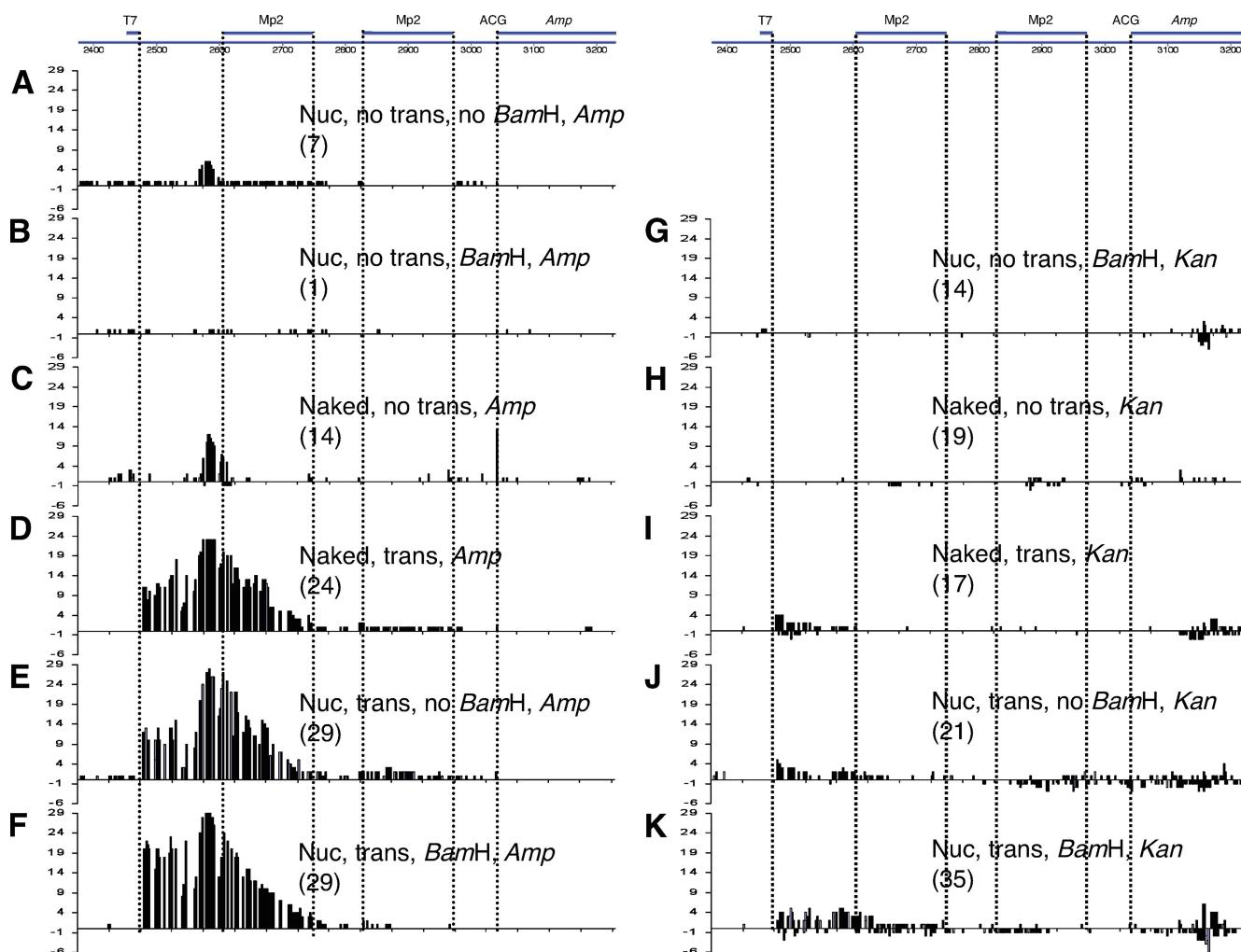


Figure 5. Mutation patterns in *pKMP2* treated with AID. Cytosine deaminations in the vicinity of the *Amp* gene. The sequenced region is shown at the top. Numbers in parentheses are the numbers of sequenced plasmids isolated from *E. coli* colonies that had any mutation. The *E. coli* were grown in ampicillin (A–F) or kanamycin (G–K). The treatments of *pKMP2* before *E. coli* transformation are indicated by the following: Nuc, nucleosomal DNA; trans, transcription; no trans, no transcription; BamH1, digestion with BamH1; no BamH1, no digestion with BamH1; naked, supercoiled DNA not incubated with histone octamers. Mutations on the top of the line are from the top (nontranscribed) strand; mutations on the bottom are from the bottom (transcribed strand). ACG, codon that is expected to be changed by AID to the initiation codon ATG. To obtain statistics, we tested the null hypothesis that the proportion of clones with mutations in the MP2 sequences in different conditions are equal, using a Fisher's exact test for numbers of clones with mutations in MP2 over total mutated clones in the same condition. 5C→5A, $P = 0.016$ (thus, 5C has significantly more MP2 mutations than 5A); 5H→5G, $P = 0.027$ (thus, 5H has significantly more MP2 mutations than 5G); 5K→5G, $P = 0.021$ (thus, 5K has significantly more MP2 mutations than 5G). The “spurious” clone 5B (see text) has the same proportion of MP2 mutations as 5H ($P = 1.0$). The numbers on the y axis indicate the numbers of sequences that had a mutation at the indicated position. The sequences of plasmids with mutations are shown in Figs. 6, 8, 9, and S1–S8. The experiments shown in Fig. 5 were repeated four times (B–E and G–K), three times (F), and two times (A) with similar results.

Nuc, transcription, BamH I, Amp

2481 AAGCTTCACGTACTGGGCCATCGCCCCCTGATAAGACGTTTTCCGGCCCTTCACGTTGGAGTCCACGTTCTTAATAGTGA
 1 ---t-t-t-----tt-t-ttt-----t-t-t-----t-t-----tt-t---t
 2 ---t-t-t-----t-t-t-----t-----t-----t-----t
 3 ---t-t-t-----t-t-t-----t-----t-----t-----t
 4 ---t-t-t-----t-t-t-----t-----t-----t-----t
 5 ---t-t-t-----t-----t-----t-----t-----t-----t
 7 ---t-t-----t-t-----t-----t-----t-----t-----t
 8 ---t-t-----t-----t-----ttt-----t-----t-----t
 9 ---t-t-----tt-t-t-----t-----t-ttt-----t-----t
 10 -----t-----t-----t-----t-----t-----t-----t
 11 ---t-t-----tt-t-t-----t-----t-ttt-----t-----t
 12 -----t-----t-----t-----t-----t-----t-----t
 13 ---t-t-----t-----t-----t-----t-----t-----t
 14 ---t-t-----t-----t-----t-----t-----t-----t
 16 -----t-----t-----t-----t-----t-----t-----t
 17 ---t-t-----t-----t-----t-----t-----t-----t
 18 -----t-----t-----t-----t-----t-----t-----t
 19 ---t-t-----tt-t-ttt-----t-----t-ttt-----tt-t---t
 20 ---t-t-----tt-t-ttt-----t-----t-ttt-----tt-----t
 21 ---t-t-----tt-t-ttt-----t-----t-ttt-----t-----t
 22 ---t-t-----tt-t-ttt-----t-----t-ttt-----tt-t---t
 23 ---t-t-----t-----t-----t-----t-----t-----t
 24 ---t-t-----t-----t-----t-----t-----t-----t
 25 -----t-----t-----t-----t-----t-----t-----t
 26 ---t-t-----tt-t-ttt-----t-----t-----t-----tt-t---t
 27 ---t-t-----t-----t-----t-----t-----t-----t
 28 -----t-----t-----t-----t-----t-----t-----t
 29 ---t-t-----tt-t-ttt-----t-----t-----t-----tt-t---t
 30 ---t-t-----t-----t-----t-----t-----t-----t
 2561 CTCTGTTCCAACTGGAAACACACTAACCCCTATAGATCTTCCTGCAGAAGCTGGTGGTGGCGGGCCGCTCAATTGGTCG
 1 t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 2 ---t-----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 3 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 4 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 5 ---t-----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 7 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 8 -----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 9 -----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 10 -----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 11 -----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 12 t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 13 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 14 -----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 15 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 16 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 17 t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 18 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 19 t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 20 -----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 21 -----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 22 t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 23 -----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 24 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 25 t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 26 t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 27 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 28 -----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 29 t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 30 t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 2641 TAGCAAGCTCTGGATCCGCTTGATCGAACGTACGCGCTGTCCCCGGTTTAAACGCCAAGGGGATTACTCCCTAGTCT
 1 ---t-----t-----tt-----t-----t-----t-----tttt-----t-----t-----t-----ttt-----t-----t-----t
 2 ---t-----t-----tt-----t-----t-----t-----tttt-----t-----t-----t-----ttt-----t-----t-----t
 3 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 5 -----t-----t-----tt-----t-----t-----t-----tttt-----t-----t-----t-----ttt-----t-----t-----t
 10 -----t-----t-----tt-----t-----t-----t-----tttt-----t-----t-----t-----ttt-----t-----t-----t
 11 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 12 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 13 -----t-----t-----tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 14 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 15 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 16 -----t-----t-----tt-----t-----t-----t-----tttt-----t-----t-----t-----ttt-----t-----t-----t
 17 -----t-----t-----tt-----t-----t-----t-----tttt-----t-----t-----t-----ttt-----t-----t-----t
 18 -----t-----t-----tt-----t-----t-----t-----tttt-----t-----t-----t-----ttt-----t-----t-----t
 22 -----t-----t-----tt-----t-----t-----t-----tttt-----t-----t-----t-----ttt-----t-----t-----t
 25 -----t-----t-----tt-----t-----t-----t-----tttt-----t-----t-----t-----ttt-----t-----t-----t
 26 -----t-----t-----tt-----t-----t-----t-----tttt-----t-----t-----t-----ttt-----t-----t-----t
 2721 CCAGGCACGTTGTCAAGATAATACATCCGTGAAGATCTCTCGGTCTATTCTTTGATTTATAAGGGATTTTGCCGATTTC
 2 tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 5 tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 10 tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 12 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 16 tt-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 2801 GGCCTATTGGTTAAAAAGGTACCCCTCGCAAGCTGGTGGCGGGGCGCTCAATTGGCTAGCAAGCTCTGGATCC
 1 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 5 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 16 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 2881 GCTTGATCGAACGTACGCGCTGTCCCCCGCGTTTAAACGCCAAGGGGATTACTCCCTAGTCTCCAGCGACGTGTCAAG
 2961 ATATACATCCGTGGGGTACCTGAGCTGATTAAACGGCAATTTACAAATTTGAAAAGGAAGGT
 9 -----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t
 3041 ACGTATATTCAACATTTCCGTTCGGCCCTATTCCCTTGGCGCATTTCCTGTTTGCTCACCCAGAAAC

Figure 6. Sequences of *pKMP2* plasmids treated with AID during transcription and amplified in *E. coli* in the presence of ampicillin. The WT sequence is shown on the top lines. The two MP2 nucleosome positioning sequences are in green (BamHI sites underlined). The in-frame ACG initiation

Fig. S3). This confirms, with a different plasmid than that previously used, that supercoiled DNA is a target for AID without being transcribed (11, 12). It also shows that the nucleosome positioning sequence in the absence of nucleosomes is targeted by AID in this cell-free assay. Efficient mutability may have been expected because the content of AID hotspots in the MP2 sequence (10 WRC and 11 GYW for \sim 25% of the total Cs and Gs in hotspots) was the same as in an Ig transgene (22) and slightly higher than the 22% in a related supercoiled AID test plasmid that did not contain nucleosome positioning sequences (12). The proportion of mutations in AID hotspots was also high, at 56% in the first nucleosome positioning sequence compared with 50% in the previous test gene (12). Thus, the nucleosome positioning sequence was a qualified AID target.

After digestion with BamHI, essentially no Amp^r colonies were obtained from nucleosome-containing plasmids without transcription. In multiple experiments of AID- and BamHI-treated nucleosome-containing plasmids without transcription, in which a total of 3.8×10^5 colonies were obtained in kanamycin, only a single Amp^r colony was seen that had some mutations in the nucleosome positioning sequences (Fig. 5 B). Possibly this one plasmid providing ampicillin resistance contained no nucleosomes over its MP2 sequences but was somehow sequestered from BamHI access. As shown in subsequent sections (Fig. 5 G), plasmids with two positioned nucleosomes selected in kanamycin have no mutations in the MP2 sequences, indicating that AID does not access DNA in nucleosomes. pKMP2 plasmids that had been incubated with histones and not digested with BamHI readily acquired large numbers of mutations induced by AID and conferred ampicillin resistance (Fig. 5 A). These clones present additional support that DNA in nucleosomes is not accessible to AID without transcription. The plasmids in Fig. 5 A were treated with histone octamers and AID without transcription but not treated with BamHI. All of them have mutations that are not seen in kanamycin selection when nucleosomal pKMP2 is treated with AID and BamHI without transcription (Fig. 5 G). The Fig. 5 A clones have either created a TATA box or ACG \rightarrow ATG (Fig. S1). Thus, all of the Amp^r clones in Fig. 5 A must have been from naked DNA or DNA with a single nucleosome. Plasmids that had been charged with two nucleosomes did not acquire ampicillin resistance because the mutations required for overcoming the inefficient ACG start codon did not occur in the presence of nucleosomes.

Thus, the pattern of mutations for Fig. 5 A (sequences in Fig. S1) is similar to clones in Fig. 5 C, which are naked DNA treated with AID without transcription (sequences in Fig. S3). However, most clones from naked DNA with no transcription (Fig. 5 C) had mutations in the ACG initiation

codon (Fig. S3), but only one from histone octamer-treated clones with no transcription (Fig. 5 A) had an ACG mutation (Fig. S1). Furthermore, within the MP2 regions, clones in Fig. 5 C had significantly more mutations than clones in Fig. 5 A ($P = 0.016$). This suggests that most of the 5A clones had only one nucleosome positioned on the second MP2 sequence, disallowing access of AID to the ACG but permitting creation of a novel TATA box upstream of the first naked MP2 sequence.

Incubation of BamHI-treated nucleosomal plasmid DNA with AID during transcription with T7 RNA polymerase restored ampicillin-resistant colonies to \sim 11% of the numbers seen in kanamycin. Cytosine deamination was seen in the DNA of both positioned nucleosomes (Fig. 5 F), indicating that DNA constrained in nucleosomes is readily accessible to AID when it is transcribed. Every C in nucleosome 1, the first of which is downstream of the T7 transcriptional promoter, was mutated in some of the 29 sequences obtained from Amp^r bacterial colonies, although no sequence had mutations in all 43 Cs in this nucleosome (Fig. 6). Because there was no C that was deaminated in nucleosome 1 in all 29 sequences and there was no C that was unmutated in all 29 sequences, it appears that within the nucleosome there were no strong topology-dependent hot or cold subregions. Overall, there was a steady decline of mutability from a peak just before the start of the first nucleosome throughout that nucleosome, the spacer between the nucleosomes, and nucleosome 2. None of the 29 sequences had mutations within the Amp^r gene (Fig. 5 F and Fig. 6). This diminishing mutation profile is apparently a result of the need for expression of ampicillin resistance, as it is not observed when the *E. coli* transformed with AID-treated plasmids was selected in kanamycin (see Nucleosomal DNA is accessible to AID...).

A similar profile of mutation frequencies is also seen in AID-treated transcribed naked DNA and in transcribed nucleosomal DNA not digested with BamHI (Fig. 5, D and E). The latter are a mixture of plasmids with two, one, or no positioned nucleosomes. Thus, AID-treated plasmids selected in ampicillin show that nucleosomal DNA can be readily accessed by AID but only when transcribed.

Expression of the Amp^r gene without the native ATG start codon after treatment with AID

Surprisingly, the majority of mutant plasmids that conferred ampicillin resistance were without the native ATG start codon in the Amp^r gene (Figs. 5 and 6). It was possible that mutations had created a new start codon of the Amp^r gene that allowed the gene product to form a fusion protein conferring Amp^r activity. Indeed, AID deamination activity created some new ATGs, which are potential start codons to form a fusion

codons are in italics and underlined. The in-frame stop codons are underlined and not italicized. The AID-created Ts resulting in a potential TATA box present in all 35 clones are in bold font. The Shine-Dalgarno sequence (AGGAAG) is in blue. The transcription start site of unmutated pKMP2 plasmids in *E. coli* is in purple and underlined. The 5' end of the Amp^r coding region is in red. The ACG start codon for Amp^r is at 3041–3043. Dashed lines represent agreement with the unmutated pKMP2 with mutations indicated by lowercase font. The sequences were obtained from one of three experiments.

protein with the *Amp*^r gene product (Fig. 6, four in-frame ACG triplets in italics and underlined were mutated to ATGs). However, downstream of these start codons are five stop

codons, including all three types of stops (Fig. 6, underlined triplets, not in italics). Thus, the new ATGs cannot initiate a protein to resist ampicillin.

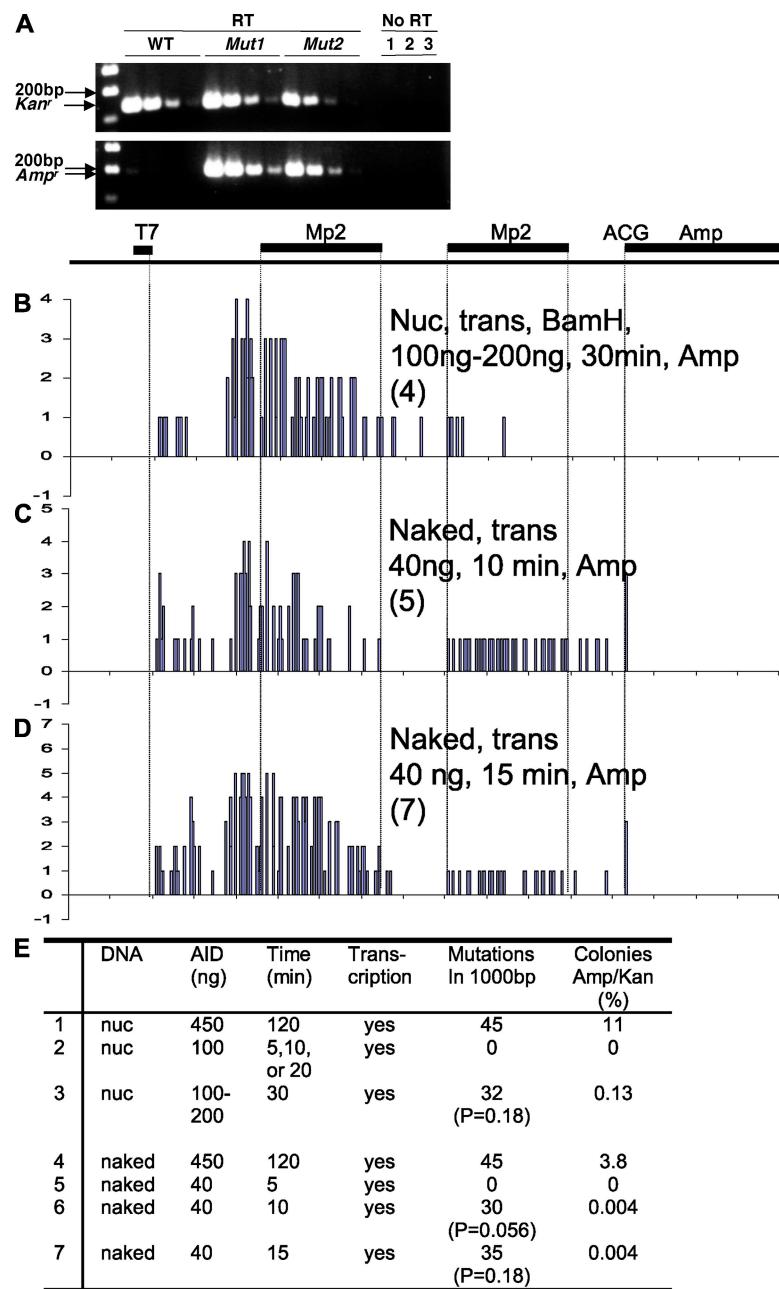


Figure 7. *pKMP2*-derived transcripts and mutations of the *Amp*^r gene in plasmids selected in ampicillin. (A) RT-PCR of RNA of *E. coli* transformed with unmutated (WT) or two mutated (Mut1 and Mut2) *pKMP2* plasmids that confer ampicillin resistance (Fig. 5 F and Fig. 6, clones 1 and 7). Kan^r and Amp^r are transcripts of the respective genes in fourfold sequential dilutions. Equal amounts of cDNA were used for the respective Kan^r and Amp^r reactions. RT-PCR products without reverse transcription with RNA samples corresponded to the highest concentrations of WT (1), Mut1 (2), and Mut2 (3) RNAs, respectively. (B-D) Mutation patterns in *pKMP2* after exposure to low amounts of AID for a short time. Treatments of the plasmids are indicated, with numbers of colonies obtained in *E. coli* selected in ampicillin shown in parenthesis. The y axes represent numbers of mutations. Relative to Fig. 5, the scales were expanded by factors of 7.25, 4.8, and 6.9 for B, C, and D, respectively, for easier comparison with the data in Fig. 5 F (for B) and Fig. 5 D (for C and D), where larger numbers of colonies were obtained. The sequences are shown in Figs. S9-S11. (E) AID acts processively/cooperatively. Selection of AID-treated plasmids in ampicillin. 2: no ampicillin resistance was seen with nucleosomal DNA and 5, 10, or 20 min of exposure to AID, T7 RNA polymerase, and BamHI. 5: no ampicillin resistance was seen with naked DNA and 5 min of exposure to AID and T7 RNA polymerase. The size markers are 100, 200, and 300 bp. The experiments were done four (E4), three (E1 and 2), two (A, C, D, and E5-7), and one (B and E3) times with similar results.

Another possibility was that the C→T mutations generated multiple A/T-rich promoters that significantly enhanced transcription. To investigate whether multiple promoters existed, we performed 5' rapid amplification of complementary DNA (cDNA) ends with RNA from *E. coli* transformed with unmutated and two mutated *pKMP2* plasmids that conferred ampicillin resistance (Fig. 6, numbers 1 and 7) to determine transcription start sites. Indeed a unique start was seen with the unmutated *pKMP2* transcripts, but the mutated plasmids had additional start sites (unpublished data). Sequences of the plasmid transcripts showed that in the nonmutant *pKMP2* plasmid the transcription start site was uniformly 45 bp upstream of the mutant Amp^r start codon (Fig. 6 A, underlined purple at position 2996). This apparently weak promoter does not have a recognizable TATA box at −10 bp upstream. In the two mutant *pKMP2* plasmids, a new TATA box (TAATAT) is created starting at position 2580 that is also present in every one of the other 27 sequences (Fig. 6). This, and presumably other newly created promoters, must have significantly enhanced transcription activity in the *Amp^r* region. Indeed, transcription activity in the *Amp^r* region in the two mutant plasmids was ~50× higher than that in the nonmutant *pKMP2* plasmid (Fig. 7 A). Thus, apparently, with high numbers of Amp^r transcripts, enough Amp^r protein was produced using the inefficient (27, 28) ACG start codon that

was preceded by an unmuted Shine-Dalgarno sequence (AGGAAG). The data show that transcription during AID contact is required for AID to access DNA in nucleosomes and, in these experiments, create an efficient promoter that acts in *E. coli*.

Nucleosomal DNA is accessible to AID on both DNA strands during transcription

When the plasmids were selected in ampicillin, all cytidine deaminations observed were on the coding strand (Fig. 5, A–F). Similar observations were previously made with related plasmids mutated by AID (11, 12). This finding is apparently the result of the need for expression of the *Amp^r* gene. Very few, if any, mutations were seen in the *Amp^r* gene coding region when the *E. coli* transformed with the AID-treated *pKMP2* plasmid were selected in ampicillin (Fig. 5, A–F). Such mutations were seen, however, when selection was in kanamycin (Fig. 5, G–K), suggesting that they would be detrimental to ampicillin function.

To investigate mutation patterns independent of plasmid selection, we sequenced *pKMP2* plasmids from bacteria grown in kanamycin (Fig. 5, G–I). Selection for kanamycin resistance does not depend on AID because the *Kan^r* gene has its own *E. coli* promoter and a functional translation initiation site. With kanamycin selection, there is no requirement for a functional

Nuc, no transcription, BamH I, Kan

2321	CGCCGGCTTCCCCGTCAGCTAAATCGGGGCTCCCTTAGGGTCCGATTTAGTGCTTACGGCACCTCGACCCCA
9C	-----aaa-----
2401	AAAAACTTGATTAGGGTATGGTCACGTAGTGCTCGAGCTGCAGACTAGT TAATACGACTCACTATAGGGAGA GAATTC
5G.1	-----a-----
6F	-----t-t-t-----
2481	AAGCTTCACGTAGTGGGCATGCCCTGATAGACGGTTTTCGCCCTTGACGTTGGAGTCCACGTTCTTAATAGTGGAA
1G	-----a-a-----
2561	CTCTTGTCCAACACTGGAACAACACTCAACCCCTATAGATCTTCCTGCAGAACGCTTGGTGCAGGGGGCGCTCAATTGGTCG
2641	TAGCAAGCTCT GATCC GCTTGATCGAACGTAACCGCTGTCCCCCGCTTAAACGCCAAGGGGATTACTCCCTAGCT
2721	CCAGGCACGTCTAGATATACAT CCGT GAAGATCTCGGCTATTCTTTGATTATAAGGGATTGGCATTTC
4D.1	-----a-----
2801	GGCCTATTGGTTAAAAAGGTACCCCTGCAGAACGCTTGGTGCAGGGGGCGCTCAATTGGTCGTAAGCCT TGGATCC
2881	GCTTGATCGAACGTAACGGCTGTCCCCCGCTTAAACGCCAAGGGGATTACTCCCTAGTCTCCAGGCACCTGTCAAGAT
2961	ATATACATCCCTGTTGGGTACCTGAGCTGATTAAACAAAATTAAACGGAATTAAACAAAATATTGAAAAGGAAGAGT
3041	ACG TATATTCAACATTCCTGTCGCCCTTATTCCCTTTTGCGGCATTGGCCTCCGTCTACCCAGAAAC
5E.1	-----t-----
6B	-----a-----
3121	GCTGGT AAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGGAGTGGTTACATCGAACATGGATCTCAACAGCGGTAAAGA
1G	-----a-----
1H	-----aa-a-----
3E.1	-----a-----a-----a-----
3D.1	-----t-----t-----t-----
5E.1	-----t-----t-----t-----t-----t-----t-----
5F.1	-----a-----a-----a-----
6C	-----a-a-aaa-----a-----
6A.1	-----t-----t-----
6E.1	-----a-a-aaa-----
9C	-----a-a-a-----
3201	TCC TTGAGAGTTTCGCCCGAAGAACGTTCCAATAATTAGCACTTTAAAGTTCTGCTATGTGGCGCGGTATTATCC
5E.1	-----tt-----t-t-----t-----tt-----a-----

Figure 8. Sequences of *pKMP2* plasmids treated with AID without transcription and amplified in *E. coli* in the presence of kanamycin. 14 sequences, in and near the *Amp^r* gene (Fig. 5 G) in *pKMP2*, with two positioned nucleosomes (BamHI digested) were treated with AID without transcription and selected in kanamycin. The WT sequence is shown on the top lines. The two MP2 nucleosome positioning sequences are in green (BamHI sites underlined). The T7 promoter is in orange. The 5' end of the *Amp^r* coding region is in red. The ACG start codon for *Amp^r* is at 3041–3043. Dashed lines represent agreement with the unmutated *pKMP2* with mutations indicated by lowercase font. The sequences were obtained from one of four experiments.

Amp' gene and the mutation pattern in the *Amp'* gene is therefore mainly a reflection of AID accessibility and action.

Naked *pKMP2* plasmid DNA treated with AID in the absence of transcription and selected in kanamycin shows efficient C deaminations in the nucleosome positioning sequences (Fig. 5 H). However, when nucleosomes were present, these sequences were not mutated despite mutations outside of the positioned nucleosomes (Fig. 5 G and Fig. 8). We showed previously that circular DNA that is supercoiled is an efficient target for AID (11, 12). When the circles are relaxed by topoisomerase I, the DNA is not accessible to AID (11, 12). In the experiments reported in this paper, the DNA remained supercoiled when charged with histones (Fig. 4 B). This explains why there were mutations outside of the nucleosome positions without transcription (Fig. 5 G). The supercoiling effect allows mutations within the nucleosome positions in naked DNA (Fig. 5 H). However, this effect is eliminated when the DNA interacts with histones (Fig. 5 G).

When the plasmids contained two nucleosomes (BamHI digested) and the *Amp'* gene was transcribed in vitro by T7 RNA polymerase, at least 10 \times more colonies were seen in kanamycin than in ampicillin. Sequencing DNA from random kanamycin-resistant colonies showed, on average, 11% of the DNA sequences with mutations in and 5' of the *Amp'* gene. Of 35 sequences, five appeared likely to confer resistance to ampicillin (Fig. 9). One had an ATG start codon (Fig. 9, number 3d) and four had potentially efficient promoters for transcription in *E. coli* and no mutations within the *Amp'* coding region (Fig. 9, numbers 7, 17, 7e, and 2a). AID-induced mutations were seen in the nucleosome positioning sequences as well as throughout the sequenced region in and upstream of the *Amp'* gene (Fig. 2 K and Fig. 6). Both DNA strands in nucleosomes were found to be targeted by AID when there was no selection for the function of the *Amp'* gene (Fig. 5 K and Fig. 9). Within the MP2 regions, of 35 sequences 6 had mutations at C (top strand) and 4 at G (bottom strand), totaling 62 C to T and 48 G to A transitions (mean C \rightarrow T in different experiments in the MP2 sequences, 15.5 ± 16.9 ; mean G \rightarrow A in MP2, 12 ± 13.3 ; Fig. 9, nt 2604–2750 and 2827–2973). Thus, single-stranded DNA allowing AID access is created in both DNA strands in nucleosomes when they are transcribed. Both strands are also accessible outside of the nucleosomes. In the sequenced regions outside of the MP2 sequences, of the 35 plasmids with mutations 18 had C to T, 10 had G to A, and 3 had both types of mutations (Fig. 9). It thus appears that AID rarely changes the strand of engagement. In one of the three sequences with both C and G mutations (Fig. 9, number 3), these were separated by 669 bp and, thus, presumably independent targetings. In the other two (Fig. 9, numbers 9 and 10), the C and G mutations were interspersed, perhaps as a result of the rare occasion of AID acting on both strands in the same event. There were no mutations at both C and G within nucleosomes, but the numbers are not high enough to determine whether AID can switch strands within a nucleosome. The *pKMP2* plasmids selected in kanamycin show that core histones prevent AID induction of C deamination in an otherwise highly susceptible

target and that both DNA strands become AID targets when the nucleosomes are transcribed.

The limiting step of AID targeting is accessing any DNA sequence

To determine how the concentration and time of interaction of AID with DNA affected the frequency and pattern of mutations, we treated transcribed *pKMP2* DNA containing two nucleosomes (BamHI digested) with two to four times less AID (100–200 ng equivalent to ~ 2.04 –4.08 pmol per reaction) for a quarter of the time (30 min). In all other experiments, the molar ratios were ~ 9.2 pmol AID to 100 fmol of plasmid DNA or approximately five AID molecules for 100 cytosines. With less AID and less time, 85 \times fewer *Amp'*/Kan^r colonies were obtained from nucleosomal DNA than with higher concentrations of AID for a longer time (Fig. 7 E). However, the pattern of mutations (compare Fig. 7 B with Fig. 5 F) and the frequency of mutations per mutated DNA sequences (Fig. 7 E, numbers 1 and 3) were very similar ($P = 0.18$).

Essentially the same findings were made when the concentration of AID and incubation times were further reduced in an experiment with naked DNA (Fig. 7 E, numbers 4–7). With $>10\times$ less AID than in the experiments shown in Fig. 5 and incubation times of only 10 and 15 min, 950 \times fewer *Amp'* colonies were obtained (Fig. 7 E, numbers 4, 6, and 7) but the pattern and frequencies of mutations were almost the same as with higher exposure to AID (Fig. 7, C, D, and E [numbers 4, 6, and 7]; $P = 0.056$ and 0.18). This suggests that the limiting step in the AID action was access to any DNA molecule and that, once engaged, AID acted either processively or cooperatively (see Discussion).

DISCUSSION

For this study, we used circular supercoiled plasmids that efficiently interacted with histone octamers and resulted in high numbers of kanamycin-resistant bacterial colonies. The *Kan'* gene was transcribed from a lac promoter requiring *E. coli* RNA polymerase for its function. Because this polymerase was not present in the AID assays and because mutations by AID in the *Kan'* gene in the absence of transcription were low (11, 12), almost all AID-treated plasmids were likely unmutated in the *Kan'* gene. In the current study, we have transformed *E. coli* with 17 of the ampicillin-resistant clones with the highest numbers of mutations in the *Amp'* gene (Fig. 5 F and Fig. 6) and found that all confer kanamycin resistance. Thus, the number of plasmids selected in ampicillin over those selected in kanamycin gave a reasonable estimate of beneficial mutations in and near the *Amp'* gene.

DNA sequences wrapped around nucleosomes without transcription are not targets for AID. This is especially apparent when the *E. coli* are grown without selection in ampicillin. Comparing naked DNA (Fig. 5 H and Fig. S6) with nucleosomal DNA (Fig. 5 G and Fig. 8) shows mutations in the nucleosome positioning sequences only when no nucleosomes are present. A putative TATA box at position 2580 seen in transcribed nucleosomal DNA is not mutated in any of the

Downloaded from http://rupress.org/jem/article-pdf/206/5/1909113/jem_20082678.pdf by guest on 10 February 2026

Figure 9. Sequences of *pKMP2* plasmids treated with AID during transcription and amplified in *E. coli* in the presence of kanamycin. 35 sequences (Fig. 5 K) are shown. The legend for Fig. 8 contains other details. The sequences were obtained from four experiments.

14 clones of nontranscribed nucleosomal DNA selected in kanamycin (Fig. 8). Thus, the nucleosome appears to prevent access to AID not only in the wrapped DNA but also in the nucleosome's vicinity to at least 19 nt, 5' from where the TATA box would be created when the nucleosome was transcribed (compare Figs. 8 and 9). Perhaps the nucleosome prevents flipping out of cytosines that negative supercoils induce near the nucleosome positioning sequence in naked DNA (12, 29).

The inefficient access by AID to DNA in nucleosomes parallels the lack of access by the Rag1 and Rag 2 proteins in V(D)J recombination (30). Similarly, restriction enzymes are sterically occluded from their target sites when wrapped into nucleosomes (31, 32), reducing the rate of DNA digestion by 10²–10⁶-fold as compared with naked double-stranded DNA. It is for this reason that the MP2 sequence used in our experiments is refractory to BamHI when associated with a nucleosome. AID requires single-stranded C as a target, and the contortions of nucleosomal DNA apparently do not suffice for AID access. Despite targeting of outside regions, nucleosomes cannot be accessed by AID without transcription (Fig. 5 G).

Processivity of AID (discussed in the fourth subsequent paragraph) must include unraveling double-stranded DNA in our system. However, association with nucleosomes appears to stop that unwinding process. This would mean that when AID gains access to transcribed DNA *in vivo*, and if there is no RNA polymerase removing a nucleosome 5' or 3' of a catalytically active AID molecule, deamination must stop outside of that nucleosome. The histones used here came from chicken red blood cells, which are transcriptionally inactive and whose histones are minimally acetylated (33). Recent *in vitro* analysis of the accessibility of restriction enzymes to chromatin would suggest, however, that acetylation alone would have only a modest effect (26).

AID access was efficient when the DNA in nucleosomes was transcribed (Fig. 5, K vs. G). The requirement for transcription for SHM *in vivo* may therefore, at least in part, be a result of the need for accessing nucleosomal DNA. In the experiments in this paper, presumably the octamer would be knocked off or temporarily shifted by the passage of the T7 RNA polymerase. This is based on experiments with phage polymerase which displaced the complete octamer (34). It had

#	1 st MP2					2 nd MP2				
	1	2	3	4	5	6	7	8	9	10
17	1C	18C	10C							
3a	1C									1C
7e	1C	15C	16C							
2		4G								
3		1C							1G	
7		4C	22C		1C					
8		2G	5G	5G					9G	23G
14		1G								
15		2C								
18		3C	20C			1C	1C			9C
19		2C								1C
2a		5G								
3f		11C								
10f			1G							
12d			8C	24C	5C					
2a'			3C							
1h				3G	8G	19G	8G			
5f						1C				
10g					6G	1G				
10e'						1C				
5							1C	1C		6C
2h							1G			
4								1C	1C	
11								1G	5G	
20								1C		
3d								1C		
1									7C	
6									1G	
9									4C, 4G	
10									5G, 2C	
12									1C	
13									5G	
16									1C	
21									1C	
22										8G

Figure 10. Compilation of mutations in transcribed nucleosomal DNA selected in kanamycin. The 11 groups represent the 11 stretches of 80 nt of sequences per line, as shown in Fig. 9. The potential TATA box at position 2580 is in group 3. Sequences that include MP2 are indicated at the top. The numbers are the numbers of mutations at C or G in each 80-nt stretch. The data were compiled from four experiments.

been concluded that with eukaryotic pol II, only one H2A/H2B heterodimer would be removed (35). It appears now that this finding was caused by the very high salt concentrations in the experiment and that eukaryotic pol II also displaces the whole nucleosome (36; for review see reference 19). Thus, as concerns the displacement of nucleosomes, the cell-free experiments are related to the situation of SHM in vivo.

With low levels of AID and short incubation time, the numbers of Amp^r colonies were drastically reduced (Fig. 7 E). However, the mutation frequencies overall in plasmids with any mutations were very similar in all cases (Fig. 7 E). Thus, AID engagement with any DNA molecule is the limiting step. Once engaged, mutations apparently accumulate relatively independently of the AID concentration. Because Amp^r required significant mutations, we cloned low-AID/short-time plasmids in kanamycin to determine whether a higher proportion had small numbers of mutations. Confirming the findings in Amp^r clones, of 47 Kan^r clones only one was mutated but with large numbers (17) of mutations (unpublished data). Because the binding of AID to single-stranded DNA is apparently independent of the presence of cytosines and independent of the concentration of AID hotspots (37), the mode of catalysis by AID must be the reason for the rapid multiple cytosine deaminations once initiation has occurred.

The summary in Fig. 10 raises the question of how AID tracks its target. Apparently, AID creates mutations for relatively short stretches of DNA. The longest stretch is in Fig. 10 (number 1h), from position 2688 to 2911 (224 nt). The shortest comprises only a single C→T (Fig. 10, 5f) or G→A (Fig. 10, 2). Six sequences have more than one cluster of mutations (Fig. 10, numbers 3, 7, 8, 18, 10e', and 5). Possibly, these sequences were targeted by two independent encounters with AID. In a given cluster, not every C is mutated. This was also observed by Bransteitter et al. (6), who suggest that “AID can access distal target C residues...while remaining bound to a single ssDNA substrate.” The facts that with short exposure to AID the number of mutated clones is drastically reduced but that the mutation frequency in mutated clones is almost the same as with longer exposure and that jumps between stretches of mutations occur support a jumping tracking process.

Perhaps the U/G mismatch created by the first cytosine deamination aids cooperative access of independent AID molecules to other cytosines. However, one may expect that both DNA strands would benefit from the unannealed U/G gap or tilt in DNA topology if it resulted in the consecutive assembly of multiple AID molecules. In our experiments, very few individual plasmids had mutations on both DNA strands but most had multiple mutations, often in consecutive Cs, on the same strand (Figs. 8, 9, and S1–S11). Also, because the reactions with 40 ng AID had only approximately one AID molecule per 100 cytosines, a cooperative mode is difficult to envision (Figs. S9–S11). Furthermore, it was found by AFM that most AID/DNA interactions comprised only an AID monomer (38). Thus, our findings support previous suggestions (5, 12) and reasonably convincing evidence (39, 40) that AID may act processively.

How can the first ~100 nt of Ig genes disallow C deamination by AID (15, 25)? May a positioned nucleosome be responsible for the lack of mutations in the first 100 or so bp of Ig genes? As we have shown in this paper, nucleosomes can be targeted by AID in vitro when transcribed, and it is expected that this will be confirmed in vivo. The positioning sequence we used is not expected to result in a nucleosome structure that differs vastly from the average nucleosome, as it exhibits the same sequence features found in natural nucleosomes in vivo (16, 41), is close in nucleosome affinity to those of natural sequences in the mouse genome (41), and is not evidently deleterious when integrated into the yeast genome in vivo (unpublished data). However, the first transcribed nucleosome may have an unusual structure; for example, genome wide, nucleosomes at the start of transcription are greatly enriched in H2A.Z at the expense of conventional H2A histones (16, 42). The role of H2A.Z is controversial and there may be other as yet unknown modifications of the first nucleosome. Indeed, the first nucleosome comprises the region where abortive transcripts are frequent (43). It remains, thus, to further investigate the intriguing lack of SHMs at the start of AID target genes.

MATERIALS AND METHODS

Plasmid construction. A BglII site and a KpnI site were inserted into plasmid *pKMT7* (11) between 2595 and 2596 and between 2655 and 2656, respectively, to produce plasmid *pBK*. A 147-bp positioning sequence MP2 (26) was inserted into *pBK* to produce *pKMP2* with two copies of MP2 at 291 and 67 bp upstream of the *Amp^r* gene, respectively, with a spacer of 77 bp. *pKMP2* also has a functional *Kan^r* gene and a disabled *Amp^r* gene caused by a mutation in the start codon ATG to ACG.

In vitro reconstitution of nucleosome DNA. Nucleosome DNA was reconstituted by double dialysis as previously described (44). The histone octamer was purified from chicken erythrocytes as previously described (45).

Exonuclease assay of nucleosomes. Reactions were performed with 20 nM of sucrose gradient-purified nucleosomes and 30 U exonuclease III (New England Biolabs, Inc.) in buffer 1 (New England Biolabs, Inc.). To test quenching efficiency, EDTA was added before addition of the nucleosomes to separately generate the zero time point. At each time point, a 10-μl aliquot of the reaction was quenched with EDTA to a final concentration of 20 mM EDTA. A final concentration of 1 mg/ml of proteinase K and 0.02% SDS was added to each time point to remove the histone octamer from the DNA. The samples were run on denaturing 8% polyacrylamide gels with 7 M urea in TBE. The sequencing ladders were prepared with a SequiTherm Excel II DNA sequencing kit (Epicentre Biotechnologies). The sequencing reactions were done with the cy3- or cy5-labeled primers used to prepare the mp2-50 DNA molecule described in the previous sections and the unlabeled mp2 DNA molecule itself. Ladders were prepared separately with ddATP and ddTTP. This allowed the direct comparison of the exonuclease III-digested nucleosomal DNA to the different lengths of the same MP2-247 DNA. After denaturing PAGE, the gel was imaged by a Typhoon 8600 variable mode imager (GE Healthcare), which detects cy3 and cy5 separately. This allowed us to detect the two DNA strands within the same gel and, therefore, did not require separate exonuclease III reactions for each of the DNA strands. In addition, the cy3 and cy5 ladders could be loaded in the same lanes to increase accuracy of the mapping gel readout.

AID purification. The baculovirus GST-AID plasmid vector was a gift from M. Goodman (University of Southern California, Los Angeles, CA). The

method of GST-AID protein expression and purification was as previously described (3, 12).

AID treatment. Plasmids with nucleosomes were predigested with BamHI for 1 h (Fig. 4 A). Naked DNA, nucleosome DNA, and the BamHI-digested nucleosome DNA were treated with AID with or without T7 RNA polymerase. The reaction mixtures were incubated at 37°C for 2 h (or less in Fig. 7 [B–D]), and the DNAs were purified and transformed into U DNA glycosylase-deficient *E. coli* BW504 (gift from A. Bhagwat, Wayne State University, Detroit, MI) (46). Bacteria with the AID-treated plasmids were grown in kanamycin or ampicillin overnight at 37°C. In the absence of AID, we never obtained any Amp^r colonies under any condition.

AFM. The AFM system used in this study consists of a MultiMode AFM/NanoScope IIIa controller/phase Extender module/“E” piezoelectric tube scanner (Veeco Instruments) operated in the tapping mode. We used NanoScope software version 5.31r1 to measure the DNA contour length (by tallying short linear segments) and to render the data images in three dimensions using a height-encoded color scale offline.

Statistics. For Fig. 5, using a Fisher exact test, we tested the null hypothesis that the proportion of clones with mutations in MP2 are equal. For Fig. 7, to compare the number of mutations in different experiments, we used the Wilcoxon nonparametric rank test.

E. coli RNA. *E. coli* disruption and RNA purification were according to the instructions of RiboPure bacteria (Applied Biosystems). After the remaining DNA was removed by DNase I, 1 µg RNA was reverse transcribed with the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen). The RT products were treated with RNase H. This cDNA was used both for determining the transcription start sites by 5' rapid amplification of cDNA ends and the relative quantities of Kan^r and Amp^r RNAs by semiquantitative RT-PCR.

To measure the relative quantities of Kan^r and Amp^r mRNAs, RT-PCR was performed on the primary cDNA using internal primers in the *Kan^r* and *Amp^r* genes. The PCR primers for Kan^r were 5'-GGATTGCAC-GCAGGTTCTCCG-3' and 5'-CGGTCTTGACAAAAAGAACCG-3', and for Amp^r were 5'-CCCTTTTTGCGGCATTTGC-3' and 5'-AAA-ACTCTCAAGGATCTTACC-3'. The PCR conditions for Kan^r were the following: 1 cycle at 94°C for 90 s, 57°C for 30 s, and 72°C for 50 s; 28 cycles at 94°C for 30 s, 57°C for 30 s, and 72°C for 50 s; and 1 cycle at 94°C for 30 s, 57°C for 30 s, and 72°C for 7 min. The PCR conditions for Amp^r were the same except that 26 cycles were used. The RT-PCR products are 193 bp in the 5' region of the *Amp^r* gene transcript and 138 bp in the 5' region of the *Kan^r* gene transcript.

Online supplemental material. 11 DNA sequences corresponding to the experiments in Figs. 5 and 7 are shown in the supplemental figures. Fig. S1 corresponds to Fig. 5 A, Fig. S2 corresponds to Fig. 5 B, Fig. S3 corresponds to Fig. 5 C, Fig. S4 corresponds to Fig. 5 D, Fig. S5 corresponds to Fig. 5 E, Fig. S6 corresponds to Fig. 5 H, Fig. S7 corresponds to Fig. 5 I, and Fig. S8 corresponds to Fig. 5 J. Fig. S9 corresponds to Fig. 7 B, Fig. S10 corresponds to Fig. 7 C, and Fig. S11 corresponds to Fig. 7 D. Online supplemental material is available at <http://www.jem.org/cgi/content/full/jem.20082678/DC1>.

We are very grateful to D. Nicolae (Genetic Medicine, University of Chicago) for performing the statistics tests. We thank M. Casadaban for suggesting that ampicillin resistance may have been achieved in *E. coli* by improved transcription, and we thank T. E. Martin for constructive discussions of these experiments and critical reading of the paper. We are grateful to M. Goodman for the AID plasmid, A. Bhagwat for *E. coli* strain BW504, and J.L. Whitlock for her fine technical support.

This work was supported by National Institutes of Health grants R01-AI047380 to U. Storb, R01-GM083055 to M.G. Poirier, and R01 GM054692 and R01 GM058617 to J. Widom, and Department of Medicine Development Fund to R. Lal. M.G. Poirier acknowledges support through a Career Award in the Basic Medical Sciences from the Burroughs Wellcome Fund.

The authors have no conflicting financial interests.

Submitted: 26 November 2008

Accepted: 26 March 2009

REFERENCES

1. Longerich, S., U. Basu, F. Alt, and U. Storb. 2006. AID in somatic hypermutation and class switch recombination. *Curr. Opin. Immunol.* 18:164–174.
2. Peters, A., and U. Storb. 1996. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity*. 4:57–65.
3. Bransteitter, R., P. Pham, M. Scharff, and M. Goodman. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. USA*. 100:4102–4107.
4. Dickerson, S.K., E. Market, E. Besmer, and F.N. Papavasiliou. 2003. AID mediates hypermutation by deaminating single stranded DNA. *J. Exp. Med.* 197:1291–1296.
5. Pham, P., R. Bransteitter, J. Petruska, and M. Goodman. 2003. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*. 424:103–107.
6. Bransteitter, R., P. Pham, P. Calabrese, and M.F. Goodman. 2004. Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. *J. Biol. Chem.* 279:51612–51621.
7. Chaudhuri, J., C. Khuong, and F.W. Alt. 2004. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature*. 430:992–998.
8. Chaudhuri, J., M. Tian, C. Khuong, K. Chua, E. Pinaud, and F. Alt. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature*. 422:726–730.
9. Ramiro, A.R., P. Stavropoulos, M. Jankovic, and M.C. Nussenzweig. 2003. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on nontemplate strand. *Nat. Immunol.* 4:452–456.
10. Sohail, A., J. Klapacz, M. Samaranayake, A. Ullah, and A. Bhagwat. 2003. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res.* 31:2990–2994.
11. Shen, H.M., S. Ratnam, and U. Storb. 2005. Targeting of the activation-induced cytosine deaminase is strongly influenced by the sequence and structure of the targeted DNA. *Mol. Cell. Biol.* 25:10815–10821.
12. Shen, H.M., and U. Storb. 2004. Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. *Proc. Natl. Acad. Sci. USA*. 101:12997–13002.
13. Rada, C., G. Williams, H. Nilsen, D. Barnes, T. Lindahl, and M. Neuberger. 2002. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.* 12:1748–1755.
14. Rada, C., J. Di Noia, and M. Neuberger. 2004. Mismatch recognition and uracil-excision provide complementary paths to both immunoglobulin switching and the second (dA:dT-focussed) phase of somatic mutation. *Mol. Cell.* 16:163–171.
15. Shen, H.M., A. Tanaka, G. Bozek, D. Nicolae, and U. Storb. 2006. Somatic hypermutation and class switch recombination in Msh6(−/−) Ung(−/−) double-knockout mice. *J. Immunol.* 177:5386–5392.
16. Segal, E., Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I.K. Moore, J.P. Wang, and J. Widom. 2006. A genomic code for nucleosome positioning. *Nature*. 442:772–778.
17. Ioshikhes, I.P., I. Albert, S.J. Zanton, and B.F. Pugh. 2006. Nucleosome positions predicted through comparative genomics. *Nat. Genet.* 38:1210–1215.
18. Schones, D.E., K. Cui, S. Cuddapah, T.Y. Roh, A. Barski, Z. Wang, G. Wei, and K. Zhao. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell*. 132:887–898.
19. Li, B., M. Carey, and J.L. Workman. 2007. The role of chromatin during transcription. *Cell*. 128:707–719.
20. Luger, K., A. Maeder, R. Richmond, D. Sargent, and T. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 389:251–260.

21. Richmond, T.J., and C.A. Davey. 2003. The structure of DNA in the nucleosome core. *Nature*. 423:145–150.
22. Michael, N., T.E. Martin, D. Nicolae, N. Kim, K. Padjen, P. Zhan, H. Nguyen, C. Pinkert, and U. Storb. 2002. Effects of sequence and structure on the hypermutability of immunoglobulin genes. *Immunity*. 16:123–134.
23. Rogerson, B.J. 1994. Mapping the upstream boundary of somatic mutations in rearranged immunoglobulin transgenes and endogenous genes. *Mol. Immunol.* 31:83–98.
24. Rada, C., A. Gonzalez-Fernandez, J.M. Jarvis, and C. Milstein. 1994. The 5' boundary of somatic hypermutation in a *Vκ* gene is in the leader intron. *Eur. J. Immunol.* 24:1453–1457.
25. Longerich, S., A. Tanaka, G. Bozek, and U. Storb. 2005. The very 5' end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions. *J. Exp. Med.* 202:1443–1454.
26. Poirier, M.G., M. Bussiek, J. Langowski, and J. Widom. 2008. Spontaneous access to DNA target sites in folded chromatin fibers. *J. Mol. Biol.* 379:772–786.
27. Sussman, J.K., E.L. Simons, and R.W. Simons. 1996. *Escherichia coli* translation initiation factor 3 discriminates the initiation codon in vivo. *Mol. Microbiol.* 21:347–360.
28. Raman, A., C. Guarraia, D. Taliaferro, G. Stahl, and P.J. Farabaugh. 2006. An mRNA sequence derived from a programmed frameshifting signal decreases codon discrimination during translation initiation. *RNA*. 12:1154–1160.
29. Liu, L.F., and J. Wang. 1987. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA*. 84:7024–7027.
30. Kwon, J., A. Imbalzano, A. Matthews, and M. Oettinger. 1998. Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1. *Mol. Cell.* 2:829–839.
31. Anderson, J.D., and J. Widom. 2000. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J. Mol. Biol.* 296:979–987.
32. Polach, K.J., and J. Widom. 1995. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J. Mol. Biol.* 254:130–149.
33. Mutskov, V., D. Gerber, D. Angelov, J. Ausio, J. Workman, and S. Dimitrov. 1998. Persistent interactions of core histone tails with nucleosomal DNA following acetylation and transcription factor binding. *Mol. Cell. Biol.* 18:6293–6304.
34. Studitsky, V.M., D.J. Clark, and G. Felsenfeld. 1994. A histone octamer can step around a transcribing polymerase without leaving the template. *Cell*. 76:371–382.
35. Kireeva, M.L., W. Walter, V. Tchernajenko, V. Bondarenko, M. Kashlev, and V.M. Studitsky. 2002. Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Mol. Cell.* 9:541–552.
36. Schwabish, M.A., and K. Struhl. 2006. Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. *Mol. Cell.* 22:415–422.
37. Larjani, M., A.P. Petrov, O. Kolenchenko, M. Berru, S.N. Krylov, and A. Martin. 2007. 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.
38. Brar, S.S., E.J. Sacho, I. Tessmer, D.L. Croteau, D.A. Erie, and M. Diaz. 2008. Activation-induced deaminase, AID, is catalytically active as a monomer on single-stranded DNA. *DNA Repair (Amst.)*. 7:77–87.
39. Pham, P., L. Chelico, and M.F. Goodman. 2007. DNA deaminases AID and APOBEC3G act processively on single-stranded DNA. *DNA Repair (Amst.)*. 6:689–692.
40. Pham, P., M.B. Smolka, P. Calabrese, A. Landolph, K. Zhang, H. Zhou, and M.F. Goodman. 2008. Impact of phosphorylation and phosphorylation-null mutants on the activity and deamination specificity of activation-induced cytidine deaminase. *J. Biol. Chem.* 283:17428–17439.
41. Thaström, A., P.T. Lowary, H.R. Widlund, H. Cao, M. Kubista, and J. Widom. 1999. Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.* 288:213–229.
42. Raisner, R.M., P.D. Hartley, M.D. Meneghini, M.Z. Bao, C.L. Liu, S.L. Schreiber, O.J. Rando, and H.D. Madhani. 2005. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell*. 123:233–248.
43. Dvir, A. 2002. Promoter escape by RNA polymerase II. *Biochim. Biophys. Acta*. 1577:208–223.
44. Thaström, A., P.T. Lowary, and J. Widom. 2004. Measurement of histone-DNA interaction free energy in nucleosomes. *Methods*. 33:33–44.
45. Feng, H.P., D.S. Scherl, and J. Widom. 1993. Lifetime of the histone octamer studied by continuous-flow quasielastic light scattering: test of a model for nucleosome transcription. *Biochemistry*. 32:7824–7831.
46. Klapacz, J., and A.S. Bhagwat. 2002. Transcription-dependent increase in multiple classes of base substitution mutations in *Escherichia coli*. *J. Bacteriol.* 184:6866–6872.