Biochemical Analysis of Hypermutational Targeting by Wild Type and Mutant Activation-induced Cytidine Deaminase*

Received for publication, July 19, 2004, and in revised form, September 7, 2004
Published, JBC Papers in Press, September 14, 2004, DOI 10.1074/jbc.M408135200

Ronda Bransteitter‡§, Phuong Pham‡, Peter Calabrese‡¶, and Myron F. Goodman‡¶**

From the Departments of ‡Biological Sciences and ¶Chemistry, Hedco Molecular Biology Laboratories, University of Southern California, Los Angeles, California 90089-1340

© 2004 by The American Society for Biochemistry and Molecular Biology, Inc.

The synthesis of high affinity antibodies requires activation-induced cytidine deaminase (AID) to initiate somatic hypermutation and class-switch recombination. Here we investigate AID-catalyzed deamination of C → U on single-stranded DNA and on actively transcribed closed circular double-stranded DNA. Mutations are initially favored at canonical WRC (W = A or T, R = A or G) somatic hypermutation hot spot motifs, but over time mutations at neighboring non-hot spot sites increase creating random clusters of mutated regions in a seemingly processes manner. N-terminal AID mutants R35E and R35E/R36D appear less processive and have altered mutational specificity compared with wild type AID. In contrast, a C-terminal deletion mutant defective in CSR in vivo closely resembles wild type AID. A mutational spectrum generated during transcription of closed circular double-stranded DNA indicates that wild type AID retains its specificity for WRC hot spot motifs within the confines of a moving transcription bubble while introducing clusters of multiple deaminations predominantly on the nontranscribed strand.

AID* is required for the secondary Ig gene diversification processes, somatic hypermutation (SHM) and class-switch recognition (CSR) (1). SHM and CSR are abolished in mice and humans (HIGM-2 patients) deficient for AID (1, 2). SHM entails the generation of point mutations in the V region of Ig genes at about million-fold higher rate than the rest of the genome, whereas CSR performs recombinational events with downstream sequences that delete the intervening constant regions (3). The two processes help to generate high affinity antibodies with an optimized fit to an antigen (V gene SHM) along with specialized Ig isotypes (CSR) enabling a more efficient clearing of systemic infections, and both require active transcription (4–6). Although AID is synthesized in B cells selectively under tight regulation, the ectopic expression of AID can sufficiently induce SHM and CSR in non-B cell lines and can induce C-G → T-A mutations in uracil glycosylase-deficient Escherichia coli and Saccharomyces cerevisiae (7–11).

Based on its similarity in sequence to the apoB mRNA processing enzyme Apobec-1, AID was initially thought to use RNA as its substrate (3, 12). However, biochemical assays using partially purified AID (13–16) and in vivo studies with overexpressed AID (10, 17) offer convincing documentation that ssDNA is a substrate for AID. AID appears to be inactive on RNA, dsDNA, and RNA-DNA hybrid molecules in vitro (13), but it cannot be ruled out that AID might be active on RNA in vivo (18). AID simulates hallmark properties of SHM when acting alone on naked DNA in vitro (14, 19), especially by targeting canonical WRC (20) V gene mutational hot spots (W = A or T; R = A or G) preferentially, while avoiding SYC cold spots (S = G or C; Y = C or T) (14). AID-catalyzed C deaminations in vitro exhibit broad clonal mutagenic heterogeneity (14) reminiscent of Ig V gene mutational distributions (21, 22). Transcription could enable AID to act on ssDNA exposed within the moving transcription bubble. We (14) and others (15, 17, 23, 24) have shown that transcription-dependent AID-catalyzed deaminations in vitro are strongly favored on the exposed nontranscribed strand, but the specific nature of the sequence specificity and distribution of deaminations during transcription has not been addressed.

An in-depth in vitro analysis geared to reveal the biochemical basis for the distributions of mutations and the favored targeting of deaminations to hot spot motifs is the focus of this paper. We have taken a two-tiered approach to evaluate the biochemical properties of AID: when acting on naked ssDNA and when acting on closed circular dsDNA undergoing transcription. Mutational spectra are obtained using wt AID and mutant forms of AID. The AID mutants were designed to have reduced interactions with DNA and allow us to determine how binding and processivity perturb mutation frequency and specificity. An analysis of mutations performed on ensembles of individual DNA clones has enabled a determination to be made of the precise location, timing, and distribution of AID-initiated mutations.

EXPERIMENTAL PROCEDURES

Construction and Purification of AID Mutants and DNA Substrates—Mutant AID proteins (R35E, R24E, R35E/R36D), and a 10-amino acid C-terminal deletion were constructed by site-directed mutagenesis (QuikChange site-directed mutagenesis kit, Stratagene) using the pAcG2T-AID vector (13, 14) as the template. The following primers and their complementary strands were used: 5′-ttg aag ggt gag cgt gga acc tac-3′ (for R24E mutant), 5′-tac gta gtt aag cgt gag agg act-3′ (for R35E mutant), 5′-ctg ttc tae gta gaa gag ggt ggc agg act-3′ (for R35E/R36D mutant), and 5′-ctg tac gta ggt aag gct gag agg act aca ttc ttt-3′ (for R35E/R36D double mutant), and 5′-ctg tac gta ggt aag gct gag agg act aca ttc ttt-3′ (for R35E/R36D double mutant), and 5′-ctg tac gta ggt aag gct gag agg act aca ttc ttt-3′ (for R35E/R36D double mutant). The entire coding region of AID in the mutant constructs was verified by DNA sequencing. Following the recommended protocol, the mutant plasmids were then cotransfected with linearized baculovirus DNA (BD Biosciences) to generate recombinant baculovirus encoding for the mutant.
AID proteins. To express the mutant proteins, Sf9 insect cells (Invitrogen) were infected with the corresponding viruses at a multiplicity of infection of 3 for 72 h. Wild type and mutant GST-AID proteins were purified as described previously (13, 14), with an additional purification step. All AID proteins were fractionated on Superdex 75 (Amersham Biosciences) and eluted in a buffer containing 50 mM Tris-HCl, 200 mM NaCl, 5 mM dithiothreitol, and 10% glycerol. Fractions were stored at −80 °C. Our current level of purity for wt AID is about 85% based on analyzing integrated resolved band intensities from a Coomassie-stained denaturing polyacrylamide gel (Fig. 1). The deamination activity was tested for all preparations of AID and mutant AID using the primer extension assay described previously (13). The deamination activities (pmol of DNA deaminated/µg of enzyme/min) of the AID preparations on the one specific DNA oligomer, 5′-aaa ggg gaa agC aaa gag gaa ggg tga gga ggt-3′, were as follows: 2 pmol/µg/min (R53E), 1.4 pmol/µg/min (R53E/R53D), 1.8 pmol/µg/min (C-terminal), and 0.6 pmol/µg/min (wt AID). There was no activity for the R24E AID mutant above the background for our assay (~6 fmol/µg·min). T7 RNA polymerase was purchased from Promega and ultrapure rNTP from Amersham Biosciences. M13mp2 gapped substrate DNA was prepared as described previously (14). M13mp2T7 covalently closed circular dsDNA substrate was constructed by site-directed mutagenesis of M13mp2 phage DNA, in which lacZa nucleotides 279–299 were replaced with a T7 promoter sequence 5′-CTATACGAGCTCAGTTACC-3′.

**Fig. 1.** Purified recombinant human AID from baculovirus-infected insect cells. GST-tagged AID was purified by glutathione-Sepharose affinity chromatography followed by Superdex 75 gel filtration. Lane 1, protein molecular mass standards. Lane 2, 20 µl of a Superdex 75 peak fraction (~0.5 µg).

**AID-catalyzed SHM Mutation spectra**

### RESULTS

The specificity of AID-catalyzed deamination of C → U on ssDNA recapitulates keynote features of SHM in humans, perhaps most importantly the ability to distinguish between WRC hot spots and SYC cold spots (14). To probe the biochemical basis of AID deamination specificity, we have analyzed spectra from large numbers (~50) of individual clones generated with wt AID and AID mutants.

**Spectral and Clonal Analysis of Wild Type AID-catalyzed C Deamination Specificity on ssDNA—**The WRC hot spot and SYC cold spot distinctions are not absolute because some WRC sites are more reactive than others, even those having identical WR bases preceding a target C residue, and non-hot spot C residues are occasionally deaminated in preference to WRC sites (14). Here we generate and analyze wt mutational spectra derived from individual ssDNA phage clones to probe the mechanisms of AID deamination specificity.

Human wt AID was incubated for 2.5 and 10 min in the presence of gapped closed circular M13 phage DNA containing a lacZa reporter gene within the single-strand gap (Fig. 2a). C → U deaminations on gapped ssDNA were detected as C → T mutations in M13 phage (white and light blue plaques) (Table I) after transfection into uracil glycosylase-deficient E. coli (14) (see sketch, Fig. 2). The fraction of mutated clones is small, 3.7, 5.1, and 5.2% at 2.5, 5, and 10 min, respectively (Table I), and increases approximately linearly with increased levels of AID for incubations of at least 10 min (data not shown). The average number of deaminations/clone is 13, 17, 23 at 2.5, 5, and 10 min, respectively (Table I). Half of the mutated clones have large numbers of mutations (20–70 mutations), yet ~95% of the clones reacted with wt AID are not mutated, even after a 10-min incubation (Table I; see also Ref 14).

The observation that the fraction of mutated clones saturates after 5 min whereas the average number of mutations/clone increases for at least 10 min suggests that wt AID may be acting processively; the enzyme appears to remain bound and active on one ssDNA substrate for at least 10 min (Table I). As mentioned previously (14), our experiments using gapped M13 DNA substrate contain roughly equal numbers of unannealed 7.2-kb ssDNA and gapped circular dsDNA substrate. Therefore, it seems unlikely that AID cycles on the same gapped DNA molecule because one might expect that excess ssDNA would act to trap AID that had dissociated from the gapped substrate. Cycling between DNA substrates could, of course, be occurring on a longer time scale, where partitioning between processivity and cycling could depend on such factors as ssDNA sequence and length.

**Mutational spectra** (Fig. 2a, bar graphs) were obtained from sequencing approximately 50 mutant clones at each incubation time. Deamination patterns for individual DNA clones depict the deamination specificity and spatial distribution of mutations at high resolution (Fig. 2a). The 10 clones shown below each spectrum portray the spatial mutational patterns for the most weakly and strongly deaminated DNA (Fig. 2a, sparse and dense clones, respectively). The colored circular symbols show the locations of WRC hot spot sites (red circles), NNC neutral sites (green circles), and SYC cold spot sites (blue circles) in the target region. The colored T symbols denote hot spot, neutral, and cold spot sites where an AID-catalyzed deamination has occurred resulting in a C → T mutation in lacZa.

The spectrum for the 2.5-min incubation with wt AID shows that C → T mutations occur prominently in WRC hot spots.
Fig. 2. Analysis of C deamination specificity of wt AID acting on ssDNA. a, C → U deamination spectra and the identification of C → T mutations in DNA clones after incubation with wt AID for 2.5 and 10 min were obtained by sequencing about 50 individual mutant clones. Deaminations resulting in C → T mutations in the lacZα target sequence on M13 bacteriophage (sketch at top) are identified as clear or light blue plaques, whereas nonmutated phage appear as dark blue plaques (see “Experimental Procedures”). Each colored bar represents a percentage of mutated phage clones with a C → T mutation at the indicated position on the lacZα target sequence (−217 to +149). Red bars identify C deaminations occurring in 5'-WRC hot spot motifs, blue bars represent 5'-SYC cold spot motifs, and green bars represent neutral motifs (neither WRC nor SYC). Subsets of individual mutant phage DNA clones, five with lowest and highest numbers of C → T mutations, sparse clones, and dense clones, respectively, are shown below each spectrum. The colored dots show the locations of WRC hot spot sites (red), NNC neutral sites (green),
(Fig. 2a), where 40–60% of the clones have mutations at 11 different WRCl sites distributed more or less uniformly over the 365-nucleotide target sequence. Mutations at SYC cold spots occur infrequently; two SYC sites (−202, +109) are mutated in 20% of the clones and five occur in 5–10% of the clones (−204, −171, +56, +69, +98) (Fig. 2a). Neutral C sites having neither WR nor SY 5′-motifs are mutated at frequencies between those for hot and cold spots (Fig. 2a).

Deaminations tend to congregate in regions in which WRCl hot spots appear to act as nucleation sites for multiple deaminations at nearby intermediate and cold spot motifs (Fig. 2a, dense clones). The increased numbers of C → T mutations in the 10-min clones occur primarily in neutral and cold spots, with smaller numbers of new hot spot sites undergoing deamination at the later time point. In other words, hot spot deaminations occur first and become saturated in time so that they must decrease at longer incubation times relative to neutral and cold spot mutations, a result that is consistent with processive action of AID. Scatter diagrams show that as the number of mutations/clone increases, the mutant fraction occurring in hot spot mutations decreases, but the fractions in cold spot and neutral mutations increase (Fig. 2a).

The strongest hot spot, CAC, site 130, in the R35E AID deamination spectrum is mutated in ~50% of the clones but does not have a WRCl signature motif (Fig. 2b). Note that ~40% of the clones have a proximal WRCl site (site 128) mutated. Four of 50 clones contained just one mutation, with three of the four mutated at CAC130 (Fig. 3a, sparse clones 1–3, 5). The remainder of the spectrum generally resembles wt AID; the mutations are clustered near WRCl hot spots and are accompanied by sizable gaps between clusters within which hot spots are deaminated in some clones but not in others (Fig. 3a, dense clones). Introducing a second mutation in an adjacent amino acid to form R35E/R36D accentuates the specificity difference with wt AID (Fig. 3c). Although the C → T mutation at CAC130 is mutated in about 50% of the clones, as observed with R35E, deamination of the nearby WRCl hot spot is attenuated 4-fold, from 40 to 10% for clones with both mutations. The attenuated hot spot makes mutations in the neutral motif CAC130 stand out even more and provides another example of altered mutational specificity (compare C sites 128 (red) and 130 (green) in Fig. 3, c and b).

We suggest that the double mutant, as a consequence of its reduced N-terminal charge (+7) compared with wt AID (+11), appears to be less processive and cycles much more readily between DNA substrates. Reduced processivity is implied by a 40% reduction in the average number of deaminations/clone for R35E/R36D compared with wt AID (Table I, 2.5-min incubation). An increase in cycling for the double mutant is indicated by a significant increase in the ratio of mutated to nonmutated clones with increasing length of incubation. At 2.5 min, the

and SYC cold spot sites (blue) in the target region. Similarly, the colored T symbols indicate C → T mutations at hot, neutral, and cold spots. b, relationship between the numbers of mutations/clone and fractions of C → T mutations at hot, neutral, and cold spots. Each dot in the graphs represents a single clone with the indicated number of mutations and the fraction of C → T mutations at hot, neutral, and cold spots. Trend lines were obtained by fitting the data to the linear regression analysis using Sigma Plots (version 7.0) software. p values are for the null hypothesis that the slope of the trend lines equals 0.
The fraction of mutated clones is twice as large for wt compared with R35E/R36D, 3.7% versus 1.5%, respectively (Table I). However, the opposite is true at 10 min, where the mutated clone fraction is 5.2% for wt AID compared with 19.7% for the R35E/R36D double mutant (Table I). Thus, a reduction in N-terminal positive charge has enabled the AID double mutant to access different DNA substrates, presumably by cycling from one to another substrate, with reduced processivity on each.

The mutational spectrum for the C-terminal deletion mutant ΔC AID (Fig. 3d) is similar to wt AID (Fig. 2a). An examination
of individual clones for wt and ΔC shows no significant differences in the spatial distribution of mutations (data not shown). The average number of deaminations/clone and the fraction of mutated clones are similar for wt and ΔC (Table I). Thus, ΔC, which is proficient in SHM but deficient in CSR (25, 26), appears to mimic wt AID in carrying out SHM-specific processes, i.e. hotspot and cold spot mutational specificities and processivity.

**Transcription-dependent Deamination by AID, Spectral and Clonal Analysis**—We have determined the spectrum of transcription-dependent AID-catalyzed C deamination by integrating the lacZa reporter gene along with a bacteriophage T7 promoter into a covalently closed circular M13 phage dsDNA and transcribing the gene with T7 RNA polymerase in the presence of wt AID (Fig. 4). C → U deaminations were detected as C → T and G → A mutations in M13 mutant phage (white and light blue plaques) after transfection into uracil glycosylase-deficient E. coli (14) (see sketch, Fig. 4). The background mutant frequency of dsDNA substrate alone is 0.5 × 10⁻³. A slight increase in mutation frequency (~2-fold) is observed in the presence of either T7 RNA polymerase or AID alone (Table II). Both AID and T7 RNA polymerase are present, a 4-fold increase in mutation frequencies (3.7 × 10⁻³) is observed as a result of transcription-dependent AID-catalyzed deamination (Table II).

We selected 86 mutant phage clones from the transcription-dependent AID-catalyzed deamination reactions and performed a sequencing analysis. Among these clones, 68 contain only G → A or C → T mutations, characteristic of AID deamination on the nontranscribed and transcribed strands. The other 18 clones have single background mutations other than G → A or C → T that may have been generated during transcription because there was an increase in mutation frequency when only T7 RNA polymerase was added to the reaction. To determine the background frequency of G → A or C → T mutations generated by T7 transcription alone, we sequenced a few randomly selected mutant phage clones reacted with T7 RNA polymerase in the absence of AID. The data show that all clones contain single mutations, but only a small portion of clones have G → A (1 of 18) or C → T (4 of 18). Thus, most all of the G → A or C → T mutations observed in the 68 clones from reactions including both T7 RNA polymerase and AID represent deamination action of AID on actively transcribed dsDNA, especially in the clones where there are more than one mutations on either strand.

Half of the 68 analyzed clones have only a single mutation, whereas 20% have two, 23% have 3–6, and 9% have between 10 and 35 mutations (Table III). Most of the mutations are on the nontranscribed strand, with a 13:1 bias over the transcribed strand (219 mutations on the nontranscribed strand and 17 mutations on the transcribed strand). Deaminations occur on the nontranscribed strand in 60 of the clones and on the transcribed strand in 5 clones. In 3 clones, deaminations occur on both strands indicating that AID may be able to perform strand switching during transcription (Table III). The far greater numbers of mutations on the nontranscribed strand is consistent with AID acting on ssDNA (14, 15, 17, 23), i.e. on the single-stranded portion of the transcription bubble. Note that deaminations on the nontranscribed strand are detected as G → A mutations, whereas mutations occurring on the transcribed strand are shown as C → T (Fig. 4a).

We also observed that the largest numbers of mutations occur close to the promoter region and decline sharply as the distance increases farther downstream (Fig. 4a). Although the lacZa target can detect mutations from nucleotide positions −70 to 250 (see Fig. 4a) (27), most of mutations (206 of 236) are observed within 200 nucleotides from the T7 promoter (from positions 80 to 280). Few mutations (19 of 236) occur in the detectable portion of the lacZa target positioned more than 200 nucleotides from the promoter (positions −70 to 80).

An examination of the mutation spectrum of AID-catalyzed C → U deamination on an actively transcribed dsDNA template reveals that the strong specificity of AID for deaminating WRC hot spot motifs and the avoidance of SYC cold spots are maintained during transcription (Fig. 4a). Among 219 mutations scored on the nontranscribed strand, 124 (57%) occur in the WRC hot spots, and only 13 (6%) occur in the SYC cold spots. Adjusting for the base composition, the average number of mutations/site is 3.5 for WRC hot spots (124 mutations/35 sites), 1.2 for neutral spots (82 mutations/69 sites), and only 0.5 for SYC cold spots (13 mutations/26 sites). Thus, the average number of mutations at WRC hot spots appears to be about 7-fold higher than those at SYC cold spots.

A more detailed examination of the multiple mutations within individual mutant clones suggests that AID may be tracking along within the moving transcription bubble while deaminating cytosines in a processive manner. As noted above, 32% of the mutant clones have more than three mutations.

Clone 14 contains five small mutation clusters having deaminated C residues from 2 to 5 bases apart. The clusters are separated from each other by ~20–50 bases, with an occasional C deamination occurring between them (Fig. 4b). Clone 73 contains a minimum cluster composed of two nearby Cs but mainly has single C deaminations with wide spaces in between (Fig. 4b). Clones 87 and 66 show similar patterns of clustered deaminations (Fig. 4b). In these various clones, it seems unlikely that more than one AID molecule could deaminate a tightly spaced cluster of mutations considering the physical constraints of a transcription bubble and the proximity of these mutations, especially when mutations are only 1 or 3 nucleotides apart. These deamination patterns in many mutated clones are consistent with what we observe in AID-deaminated ssDNA clones (Fig. 2a).

Clone 36 contains mutational clusters on both the nontranscribed and transcribed strands (Fig. 4b). The mutations begin on the nontranscribed strand from just inside the promoter region and continue for about 30 nucleotides into the lacZa structural gene. AID appears to switch strands and catalyze deamination of C over a 40-nucleotide region on the transcribed strand and then switch back again, causing individual and clustered deaminations for an additional 50 nucleotides on the nontranscribed strand (Fig. 4b). Although we cannot rule out the possibility that more than one transcription bubble is occurring on one molecule of DNA or that more than one AID is acting on a substrate undergoing transcription, the evidence within the individual clones suggests that AID may be acting processively within the confines of a transcription bubble.

**DISCUSSION**

An investigation into the biochemical basis underlying the stochastic behavior of AID is the principal theme of this paper. Our approach is to probe broad spectral properties of wt AID-catalyzed C deamination and then focus on the identities and distributions of mutations by examining the individual DNA clones that contribute to the overall mutational spectra. One analysis is performed with AID and AID mutants acting on ssDNA. A related analysis examines the action of AID during the active transcription of a covalently closed circular DNA by T7 RNA polymerase. SHM and CSR require active transcription (4–6).

SHM patterns in B cells (20, 28) and E. coli (29) and in vitro (14) show that hot spots, although clearly favored, are not overtly dominant because C → T mutations in WRC occur only
FIG. 4. Analysis of C deamination specificity of wt AID acting on dsDNA undergoing transcription by T7 RNA polymerase. a, C → U deaminations on the nontranscribed strand give rise to G → A mutations (shown above the DNA sequence), and C → U deaminations on the transcribed strand give rise to C → T mutations (shown below the DNA sequence). Note that although C deaminations result in C → T mutations on both strands, the C → T mutations on the nontranscribed strand are detected as G → A when the DNA clone is sequenced. Arrows in the M13mp2T7 plasmid sketch on the top (see “Experimental Procedures”) show the transcription directions of the lacZα gene and the T7 promoter. Red A or T refers to mutations occurring at C residues located within WRC hot spot motifs; mutations at C located in SYC cold spots are shown as blue A or T; mutations at C sites having neither hot spot nor cold spot motifs are shown as green A or T. b, five
about ~2-fold over randomly mutated C sites (14, 28). Hot spot C residues are occasionally not mutated, whereas intermediate and cold spots Cs are (14, 28). A compilation of mutated V gene DNA sequences from humans, mice, and cultured B cell lines shows broad heterogeneity in the number and distribution of C → T mutations (21, 22). Our data show that wt AID working on naked ssDNA behaves in a similar manner (Fig. 2).

Beyond comparing in vitro with in vivo mutational data, a biochemical analysis allows us to delve more deeply into SHM mechanisms. An examination of individual DNA clones reveals the presence of densely clustered mutations separated by lightly mutated regions (Fig. 2a, dense clones). The wide variability in the location and number of mutations for different clones suggests that AID initially binds at a random position on ssDNA and deaminates C residues processively over a short (~10-base) region but then appears to translocate randomly to a different region on the same DNA substrate (Fig. 2a). Longer exposures to AID result in an increase in the fraction of deaminations occurring at intermediate and cold spots relative to hot spots, as would be expected for a processive scanning mechanism (Fig. 2b). However, our evidence to date suggests a processive scanning mechanism is indirect. Other mechanisms could account in principal for similar types of randomly clustered mutational patterns, including aggregation of AID or binding by different AID molecules. A direct measurement of AID scanning on a single ssDNA substrate will be necessary to resolve the matter of processivity.

A simple model to account for the seemingly disparate processive and translocation properties of AID, as revealed by the spectral and clonal data, envisions AID as a dimer composed of two monomer subunits containing independent active sites (30), bound at two distant locations on the same DNA substrate strand (Fig. 5). Each subunit scans a limited span of DNA processively resulting in mutational clusters with far fewer mutations in the intervening regions. The processive scanning action of each monomer of AID could be limited to short spans of DNA if the monomers bound at two distant locations were also anchoring each other to those locations via a pulling action by each monomer as it attempts to scan in opposite directions independently. If one monomer subunit was released from the DNA, it would be able to translocate to a distant position within the same DNA molecule, as the other monomer may still be bound and anchoring the AID dimer to that same DNA molecule. We have observed that AID binds with similar affinities to ssDNA (≥30 nucleotides) whether or not C residues are present,2 which might account for the random and heterogeneous distribution of mutations for different DNA clones.

The N-terminal α helical region of AID carries a large positive charge (+11) (Fig. 3a) that should facilitate binding to the negatively charged ssDNA backbone. AID mutants with N-terminal amino acid substitutions that reduce the charge would thereby be expected to cycle between different ssDNA substrates and be less processive compared with wt AID. Increased cycling and decreased processivity are observed for the R35E/R36D AID double mutant (+7 charge); the AID double mutant catalyzes deamination on 20% of the DNA substrates compared with just 5% for wt AID (Table I, 10-min incubation) and exhibits a concomitant 3-fold reduction in the average number of deaminations/clone (Table I, 2.5-min incubation).

Unexpectedly, however, the R35E/R36D mutant has also altered the mutational specificity so that the most prominent C deamination site, found in 50% of the clones, occurs in a non-hot spot motif (Fig. 3c). This aberrant selectivity is attributable to the R35E substitution (Fig. 3b). The additional R36D substitution further alters the specificity by reducing WRC hot spot mutations located adjacent to the favored non-hot spot motif by 4-fold, from 40 to 10% for clones having both mutations. Two recent studies document the biological importance of N-terminal mutants. First, the mutation R24W (Fig. 3a) has been identified in immunodeficient patients (25), and we have determined that a mutation at this residue (R24E) inactivates

---

2 R. Bransteitter, P. Pham, and M. F. Goodman, manuscript in preparation.
AID deaminase activity \textit{in vitro} (data not shown). Second, Honjo and co-workers (31) screened AID mutants for defects in SHM by measuring the reversion of G to C at a single hot spot and verified the mutations by sequencing individual clones. SHM-defective AID mutants were found to map at several locations in the positively charged N-terminal α helix. Our biochemical data predict that a change in mutational specificity coupled with a reduction in processivity would result in a general reduction in SHM in accord with the \textit{in vivo} data (31).

In contrast to the N-terminal mutations, a 10-amino acid C-terminal deletion (ΔC) appears to have essentially no effect on either the activity or specificity of AID \textit{in vitro} (Fig. 3d). The ΔC mutant causes specific deficiency in CSR \textit{in vivo} (25), but it still induces mutations in the Sμ region, suggesting that the C terminus of AID is required for the interaction with a CSR-specific factor (26).

SHM requires the active transcription of Ig V regions (4–6). In the model transcription-dependent AID-catalyzed deamination system (Fig. 4), we observe heterogeneous mutational clustering favoring WRC hot spot deaminations, consistent with the stochastic action of AID on ssDNA in the absence of transcription (Fig. 2). There is a stronger preference for deaminations at WRC hot spots in the transcription-dependent reaction (7-fold) than on naked ssDNA (4.6-fold) (Table II) (14). Unlike gapped ssDNA, where AID has access to all possible cytosines at any given time, within a moving transcription bubble, AID must select among a limited number of C residues present in a short span of ssDNA that is only briefly exposed. These results are consistent with the analysis of deamination by AID on ssDNA substrate at short incubation times. The hot spot preference is much more prominent in the AID mutation spectrum at 2.5 min than in the spectrum at 10 min, demonstrating that first AID targets C in WRC hot spots, then later on targets those in neutral and cold spots, provided that AID still has access to them (Fig. 2a).

It has been suggested that AID is unlikely to act processively on transcribed dsDNA because of physical constraints imposed by a moving transcription bubble (32); however, the observation that ~50% of mutant clones contain multiple mutations, many clusters with tightly spaced mutations only 1–3 nucleotides apart, suggests that AID may be able to keep up with a moving bubble (4). We also observed that mutations decrease as the distance increases away from the promoter (Fig. 4a). Although the observed decline of mutations in our transcription assay occurs only over a range of 200 nucleotides downstream of the promoter, nevertheless, it is reminiscent of V gene mutational polarity \textit{in vivo} where a significant drop in mutations is observed ~500 nucleotides from the promoter (33–35).

However, it is important to emphasize that the simplified \textit{in vitro} T7 transcription assay used here does not reveal the complex \textit{in vivo} events in which mutations occur on nontranscribed and transcribed strands in roughly equal numbers (36) and where V genes but not C genes are targets for mutation. Although C deaminations occur preferentially on the nontranscribed single-stranded portion of the bubble in agreement with earlier studies (14, 17, 23, 24), and clones exist for which it appears that AID is switching from the nontranscribed to the transcribed and then back to the nontranscribed strand (Fig. 4b), it would obviously be naive to suggest that strand switching in this simple model transcription system can reflect the complex \textit{in vivo} situation in which roughly equal numbers of C → T mutations occur on both strands (36). Nevertheless, the clonal data demonstrate the physical possibility that AID could access a strand undergoing transcription and switch back and forth deaminating Cs on both strands (Fig. 4b). \textit{In vivo}, within a human RNA polymerase II-generated transcription bubble, AID could interact with transcription factors that allow it to deaminate both strands. Alternatively, transient negatively supercoiled regions in DNA sequences undergoing transcription upstream of human RNA polymerase II might allow AID to access both DNA strands (37). It is perhaps even more likely that mutations on the transcribed strand could be introduced by error-prone DNA polymerases filling in repair patches or AID-deaminating C residues within those ssDNA gaps, created by the excision of AID generated uracils on the nontranscribed strand. What our data show is that once having gained access to a transcription bubble, AID is then able to target WRC hot spots for deamination, mainly on the nontranscribed strand (Fig. 4). A similar approach using eukaryotic RNA polymerase II acting in conjunction with specific transcription factors and human single-strand binding protein RPA (24) will help to elucidate how AID gains access to a particular region undergoing transcription.

The initiation step of SHM in actively transcribed Ig V genes appears likely to be AID-catalyzed C → U deaminitations. Therefore, the stochastic behavior and processive action of AID on ssDNA and on transcribed dsDNA substrates might serve as a biochemical basis for understanding the broad and heterogeneous SHM patterns seen \textit{in vivo}. Copying U by a high fidelity replicative DNA polymerase would generate C → T transitions at SHM WRC motifs. Uracil glycosylase-catalyzed conversion of U to an abasic site, with subsequent copying by a low fidelity enzyme such as polymerase η (38, 39), could generate general C → N transversions and transitions in the same motif. And, as mentioned above, the triggering of base excision or mismatch repair enzymes (40, 41) to excise an abasic moiety can lead to a diversification of SHM to facilitate mutations at additional hot spot motifs, e.g. TA on both DNA strands.

Acknowledgments—We thank Matthew D. Scharff for ongoing generous tutelage and for a critical reading of this manuscript.

REFERENCES
1. Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000) Cell \textbf{102}, 553–563
2. Revy, P., Muto, T., Levy, Y., Greissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufuorqur-labelouse, R., Gennery, A., Tezcan, I., Ersy, F., Kaysecin, H., Ugazio, A. G., Brouesse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A., and Durandy, A. (2000) Cell \textbf{102}, 565–575
3. Honjo, T., Kinoshita, K., and Muramatsu, M. (2002) Annu. Rev. Immunol. \textbf{20}, 165–196
4. Peters, A., and Storb, U. (1996) Immunity \textbf{4}, 57–65
5. Maimils, N. (1995) Cell \textbf{83}, 9–12
6. Lee, C., Kinoshita, K., Arockiasamy, A., Cerrietteli, S., Crouch, R., and Honjo, T. (2001) J. Exp. Med. \textbf{194}, 365–374
7. Martin, A., and Scharff, M. D. (2002) Proc. Natl. Acad. Sci. U. S. A. \textbf{99}, 12304–12308
8. Okazaki, I. M., Kinoshita, K., Muramatsu, M., Yoshikawa, K., and Honjo, T. (2002) Nature \textbf{416}, 340–345
9. Yoshikawa, K., Okazaki, I. M., Etu, T., Kinoshita, K., Muramatsu, M., Nagoaka, H., and Honjo, T. (2002) Science \textbf{296}, 2033–2036
10. Petersen-Mahrt, S. K., Harris, R. S., and Neuberger, M. S. (2002) Nature \textbf{418}, 99–103
11. Polotskaya, V. P., Wilson, S. H., Kunkel, T. A., and Pavlov, Y. I. (2004) J. Immunol. \textbf{172}, 4308–4313
12. Muramatsu, M., Sankaranandan, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O., and Honjo, T. (1999) J. Biol. Chem. \textbf{274}, 18470–18476
13. Brunetsteiner, B., Pham, P., Scharff, M. D., and Goodman, M. F. (2003) Proc. Natl. Acad. Sci. U. S. A. \textbf{100}, 4102–4107
14. Pham, P., Brunetsteiner, R., Petruesca, J., and Goodman, M. F. (2003) Nature \textbf{423}, 103–107
15. Chaudhuri, J., Tian, M., Khong, C., Chua, K., Pinaud, E., and Alt, F. W. (2003) Nature \textbf{421}, 726–730
16. Dickerson, S. K., Market, E., Besmer, E., and Papavasiliou, F. N. (2003) J. Exp. Med. \textbf{197}, 1291–1296
17. Ramiro, A. R., Stavropoulos, P., Jankovic, M., and Nussenzweig, M. C. (2003) Nat. Immunol. \textbf{4}, 452–456
18. Honjo, T., Muramatsu, M., and Fagarasan, S. (2004) Immunity \textbf{20}, 659–668
19. Yu, K., Huang, F. T., and Lieber, M. R. (2004) J. Biol. Chem. \textbf{279}, 6496–6500
20. Rogozin, I. B., and Kolchanov, N. A. (1992) Biochim. Biophys. Acta \textbf{51620}, 165–196
21. Honjo, T., Kinoshita, K., and Muramatsu, M. (2002) Annu. Rev. Immunol. \textbf{20}, 165–196
22. Rada, C., Williams, G. T., Nilsen, H., Barnes, D. E., Lindahl, T., and Neuberger, M. S. (2002) Curr. Biol. \textbf{12}, 1748–1755
23. Sohail, A., Klapacz, J., Samarantarayake, M., Ullah, A., and Bhagwhat, A. S.
AID-catalyzed SHM Mutation spectra

24. Chaudhuri, J., Khuong, C., and Alt, F. W. (2004) Nature 430, 992–998
25. Ta, V.-T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., Nonoyama, S., Tashiro, J., Ikegawa, M., Ito, S., Kinoshita, K., Muramatsu, M., and Honjo, T. (2003) Nat. Immunol. 4, 843–848
26. Barreto, V., Reina-San-Martin, B., Ramiro, A. R., McBride, K. M., and Nussenzweig, M. C. (2003) Mol. Cell 12, 501–508
27. Bebenek, K., and Kunkel, T. A. (1995) Methods Enzymol. 262, 217–232
28. Shapiro, G. S., Aviszus, K., Murphy, J., and Wysocki, L. J. (2002) J. Immunol. 168, 2302–2306
29. 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
30. Xie, K., Sowden, M. P., Dance, G. S., Torelli, A. T., Smith, H. C., and Wedekind, J. E. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 8114–8119
31. Shinkura, R., Ito, S., Begum, N. A., Nagaoka, H., Muramatsu, M., Kinoshita, K., Sakakibara, Y., Hijikata, H., and Honjo, T. (2004) Nat. Immunol. 5, 707–712
32. Bhagwat, A. S. (2004) DNA Repair (Amst.) 3, 85–89
33. Lebecque, S. G., and Gearhart, P. J. (1990) J. Exp. Med. 172, 1717–1727
34. Rada, C., Gonzalez-Fernandez, A., Jarvis, J. M., and Milstein, C. (1994) Eur. J. Immunol. 24, 1453–1457
35. Rada, C., and Milstein, C. (2001) EMBO J. 20, 4570–4576
36. Milstein, C., Neuberger, M. S., and Staden, R. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 8791–8794
37. Shen, H. M., and Storb, U. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 12997–13002
38. Zeng, X., Winter, D. B., Kasmer, C., Kraemer, K. H., Lehmann, A. R., and Gearhart, P. J. (2001) Nat. Immunol. 2, 537–541
39. Rogozin, I. B., Pavlov, V. I., Bebenek, K., Matsuda, T., and Kunkel, T. A. (2001) Nat. Immunol. 2, 530–536
40. Li, Z., Scherer, S. J., Ronai, D., Iglesias-Ussel, M. D., Peled, J. U., Bardwell, P. D., Zhuang, M., Lee, K., Martin, A., Edelmann, W., and Scharff, M. D. (2004) J. Exp. Med. 200, 47–59
41. Martomo, S. A., Yang, W. W., and Gearhart, P. J. (2004) J. Exp. Med. 200, 61–68