Intricate targeting of immunoglobulin somatic hypermutation maximizes the efficiency of affinity maturation

Nai-Ying Zheng, Kenneth Wilson, Matthew Jared, and Patrick C. Wilson

Molecular Immunogenetics Program, The Oklahoma Medical Research Foundation, Oklahoma City, OK 73104

It is believed that immunoglobulin–variable region gene (IgV) somatic hypermutation (SHM) is initiated by activation–induced cytidine deaminase (AID) upon deamination of cytidine to deoxyuracil. Patch–excision repair of these lesions involving error prone DNA polymerases such as polη causes mutations at all base positions. If not repaired, the deaminated nucleotides on the coding and noncoding strands result in C-to-T and G-to-A exchanges, respectively. Herein it is reported that IgV gene evolution has been considerably influenced by the need to accommodate extensive C deaminations and the resulting accumulation of C-to-T and G-to-A exchanges. Although seemingly counterintuitive, the precise placement of C and G nucleotides causes most C-to-T and G-to-A mutations to be silent or conservative. We hypothesize that without intricate positioning of C and G nucleotides the efficiency of affinity maturation would be significantly reduced due to a dominance of replacements caused by C and G transition mutations. The complexity of these evolved biases in codon use are compounded by the precise concomitant hotspot/coldspot targeting of AID activity and Polη errors to maximize SHM in the CDRs and minimize mutations in the FWRs.

CORRESPONDENCE
Patrick C. Wilson:
wilsonp@omrf.ouhsc.edu

Abbreviations used: AID, activation-induced cytidine deaminase; CSR, class switch recombination; IgV, immunoglobulin-variable region gene; IgVH, IgV heavy chain gene; SHM, somatic hypermutation.

Individual B cells activated by antigen binding and T cell help will undergo a rapid proliferation to form many identical clones. During this process the Ig of each antigen-reactive clone is altered by somatic hypermutation (SHM; 1), which causes a million-fold increase in the frequency of point mutations of the Ig-variable region genes (2) and rare insertions and deletions (3). Current paradigms teach that individual clones then compete for antigen in a Darwinian fashion, resulting in the clonal selection of B cells harboring receptors of the highest specificity and affinity for the antigen. Recent progress has demonstrated that the molecular processes of both SHM and immunoglobulin class switch recombination (CSR) require the activity of activation-induced cytidine deaminase (AID; 4, 5).

AID is a cytidine deaminase but there is controversy as to if its substrate in vivo is a specific mRNA (RNA editing model) or if AID acts directly on the DNA (DNA deaminase model). According to the RNA-editing model, AID will deaminate a C to form a U on an unknown mRNA molecule that will re-

sult in the change of a codon and expression of a new amino acid sequence. This change in the peptide sequence will result in an activated protein that mediates SHM and CSR. This model was conceived because the protein most homologous to AID is APOBEC-1 that has an RNA-editing function (6). However, most investigators work under the assumption that AID directly deaminates Cs of immunoglobulin variable region (IgV) gene DNA or class switch regions to initiate SHM or CSR (7). The deamination of C to U on the coding strand would result in a U/G base mismatch (phase I of SHM; 8). Phase II of SHM involves various DNA repair processes. Pairing of the U with an adenosine during DNA replication would result in one of the daughter cell lineages having a C-to-T transition mutation. Similarly, deamination of noncoding strand Cs would result in G-to-A transitions if not repaired. Alternatively, the U/G mismatch can be repaired by various mechanisms including short-patch excision repair with error-prone polymerases that would introduce additional mutations at all base positions in the surrounding sequence. There is now ample evidence that the error-prone DNA polymerases η (9–13),

The online version of this article contains supplemental material.

ζ (14, 15), and possibly ι (16,17) have a direct role in SHM (for review see reference 18). Evidence that AID directly deaminates DNA has been presented in a number of recent reports (7, 19–27), and AID has now been shown to associate with a synthetic DNA substrate of SHM in conjunction with replication protein A (28). Direct evidence for the entire model of SHM involving initiation by DNA deamination and then various pathways of error-prone repair was recently reported by Rada and colleagues (29). In this report CSR was ablated and SHM was limited only to mutations that would result from C deamination (C-to-T and G-to-A) in mice deficient for both uracil nucleotide glycosylase and the mismatch repair factor MSH2. In fact, although C and G mutations in these mice were all transitions, these mutations still occurred frequently, suggesting that many C and G mutations occur due to the direct activity of AID.

Immunoglobulin SHM is transcription dependent (30), displays nucleotide exchange bias for transitions and sequence motifs referred to as hotspots accumulate mutations at increased frequencies and coldspots at reduced frequencies (31–34). Hotspots of somatic mutation have classically been defined as "RGYW" motifs representing the nucleotide sequence AG/G/CT/AT with targeting of the underlined G, or its reverse complement (WRCY or AT/GA/C/GA). The hotspot motif has now been further refined to the motif DGYW (AGT/G/CT/AT) and complement WRCH (AT/GA/C/TAC; 35). This hotspot motif likely represents the composite of AID preferred sites for deamination and the most error prone sites of the polymerases involved in SHM. Through in vitro analyses of AID activity it is now known that a large part of mutational hotspot/coldspot targeting is likely due to the preferred (hotspot) nucleotide substrate of AID, which is the C nucleotide of WRC (AT/AG/C) motifs, and the repression (coldspot) of AID activity at SYC (GC/CT/C) or the third Cs in the trinucleotides: TTC, CAC, GGC and GAC (23,27). The reverse complement of these hotspots and coldspots modulate AID activity on the noncoding strand. In addition, it is known that Polη is critical for AT mutations during SHM and is particularly error prone for coding strand adenosines preceded by A or T (WA) that are preferentially mutated to G (9–13, 18). Loss of surface Ig leads to B cell death (36), thus replacement mutations of structurally important amino acids and nonsense mutations will be avoided. Independent of these selective influences, it is evident that the antibody V gene sequences have evolved targeting with mutational hotspots to maximize mutations of the diverse CDRs while minimizing mutations to the structurally important FWRs (33, 37, 38).

With recent progress in understanding the molecular mechanism of SHM and the role of AID and C deamination, it is important to better define the role of the V genes themselves in directing this process. In recent years our laboratory has sequenced many somatically mutated human IgV heavy chain genes (IgVH). For this study, a total of 28,307 somatic mutations were analyzed for nucleotide exchange frequency, codon bias, AID hotspot and coldspot targeting, and silent versus replacement mutation tendencies. In addition, the entire set of human IgVH genes was similarly characterized for mutational targeting. From these analyses we conclude that targeting of SHM through evolution of the IgV gene sequences goes well beyond preferential positioning of hotspot motifs in the CDRs.

Evidence is presented that IgVH genes have evolved to support the initiation of SHM by AID, but to minimize the occurrence of C-to-T–induced amino acid replacements through intricate positioning of coding strand Cs. Targeting of AID is also present but not as dramatic for the noncoding strand as evident by placement of guanosines such that as for C-to-T exchanges, mutations to A (noncoding strand dC-to-dU deaminations) result in predominantly silent mutations in the FWRs, and the G-to-A exchange is biased to produce conservative amino acid replacements in the CDRs. C and G transitions are also targeted to the CDRs and avoided in the FWRs by positioning in AID hotpots and coldspots. The result is a normalization of the frequency of amino acid replacements that arise during patch excision–repair of the nucleotides proximal to the AID-induced dU/G mismatches. The evolution of the IgV genes to support C-to-T and G-to-A transition mutations provides compelling evidence that AID acts as a direct DNA C deaminase. Finally, juxtaposed on the precise placement of Cs and Gs is preferential placement of As and Ts in hotspots of Polη mutation in the CDRs and exclusion from these sites in the FWRs. Thus, targeting of amino acid replacement mutations to the CDRs rather than the FWRs, which has traditionally been attributed to selective processes, is also directed by the codons used to encode V gene sequences. This analysis demonstrates that evolution of the V genes as a substrate of the SHM process is significantly more intricate than previously appreciated. A model is presented where in the absence of this targeting the efficiency of affinity maturation would be greatly reduced.

RESULTS
Analysis of somatically mutated variable region gene sequences
B cells from various populations that harbor somatic mutations (germinal center, memory, plasma cells, and the Cδ class-switched B cells Cδ-CS or IgD-only B cells) were isolated from the tonsils and blood of 27 donors by flow cytometry and were used to clone and sequence 1,919 IgVH gene cDNAs. Many (662) of the sequences were generated from B cells of the extensively mutated Cδ-CS lineage (39, 40) thus providing a large data set of mutations for analysis. A total of 28,307 somatic mutations in 555,729 IgVH gene nucleotides sequenced were analyzed for nucleotide exchange frequency, codon bias, hotspot and coldspot targeting, and silent versus replacement mutation tendencies. Table S1 (available at http://www.jem.org/cgi/content/full/jem.20042483/DC1) provides a list of all mutations analyzed in addition to NCBI accession numbers for the reference V

genes, and indicates the B cell populations from which each V gene was sequenced. Clonal sequences (derived from the same founder), which are common in certain B cell populations, were scored both for unique mutations occurring for each clone only and with all mutations included. The results were virtually identical and thus all mutations that occurred are included in Table S1. Background mutations due to PCR or reverse transcription errors were determined to occur at a frequency of 0.0025 (0.25%) based on analyses of a portion of the Ig constant region sequenced with many of the IgVH gene sequences. Thus, 71 of the 28,307 mutations reported may have been due to experimental error, which is well below a threshold that could confound these analyses. In addition, VH gene sequences were compared with the V gene sequence set from the Immunogenetics database (41) as this set includes most known V gene polymorphic variants, and thus nucleotides differing from the germline genes due to polymorphisms would not be identified as mutations. In certain instances nucleotides altered for all instances of a V gene sequenced from a particular donor were considered to be previously unreported polymorphisms, and thus were not counted as mutations in the analyses herein. Estimation of the frequency of mutations that occurred as secondary muta-

tions (for instance C-to-T-to-A) by analysis of clonally related sequences demonstrated that these mutations were unsubstantial to the whole and so were not accounted for.

Somatically mutated human IgV$_H$ genes accumulate C-to-T base exchanges that are predominantly silent

As illustrated in Fig. 1 A, of the 28,307 mutations analyzed G-to-A (4,730, 17%) and C-to-T (4,098, 14%) base exchanges were the most common. Either of these mutations could have occurred through fixation of the AID-targeted deamination of deoxycytidine to deoxyuracil. The C-to-T exchanges would result if Cs were deaminated on the coding strand and G-to-A exchanges would result from C deamination on the template strand. As shown in Fig. 1 B, most C mutations occurred at AID hotspot motifs. Similarly, preferential targeting of Cs within reverse complement hotspots (resulting in G mutations) is indicative of AID activity on the noncoding strand (Fig. 1 B). Based both on the predominant hotspot targeting and on previous analyses attributing many C and G transitions to AID activity (Introduction), the high frequency of C and G mutations are likely due to AID-induced C deaminations that are not repaired or are improperly repaired. There is also a clear preference for A-to-G and
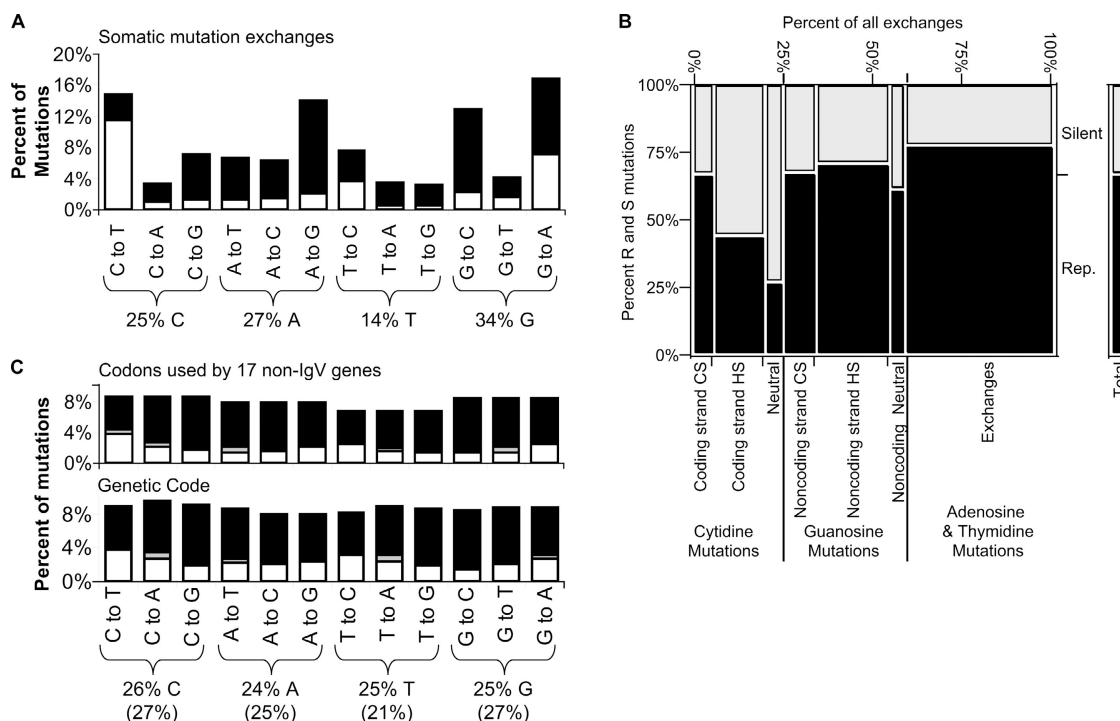

Figure 1. SHM is biased for the introduction of excessive silent C–to–T and G–to–A exchanges. (A) The silent (white areas of bars) and replacement (black areas of bars) base exchange tendencies of 28,307 somatic mutations from 1,919 human IgVH genes. (B) Distribution of C-to-T and G-to-A silent (gray) and replacement (black) mutations to hotspots, coldspots, and neutral positions (x-axis), and the proportion of the various exchanges in AID hotspots or coldspots (y-axis) compared with A and T exchanges. AID hotspots and coldspot definitions are within the text

(see introduction). (C) Analysis of the genetic code and non-IgV genes. Predicted occurrence of silent (white), replacement (black), or nonsense (gray) mutations for the 12 nucleotide exchanges. This analysis represents the sum result if every nucleotide was mutated to each of the three other nucleotides, and thus the sum of all possible mutations that could occur based on codon composition. The non-IgVH genes analyzed are listed within the text. CS, cold spot; HS, hot spot; Rep., replacement.

less for T-to-C transitions (Fig. 1 A), corresponding with many previous analyses demonstrating overall increased transitions by SHM, and consistent with the mutation spectrum introduced by Polη that is critical for SHM, particularly at AT nucleotides (9–13, 18). As previously described transversion mutations were generally reduced, however G-to-C transversions were unexpectedly common. It is notable that 70% of these atypical G-to-C transversions were silent or caused highly conservative S-to-T or V-to-L amino acid exchanges at particular positions in CDR1 and FWR3 (see Fig. 2, many at positions 35 and 83), suggesting a role for selection in the accumulation of these mutations.

Strikingly, as shown in Figs. 1 A, the C-to-T mutations that occurred were 77% (3,157/4,098) silent compared with the expected frequency based on the genetic code or analysis of non-IgV genes (42 and 48%, respectively) if all codons were used equivalently (χ^2, P < 0.0001; Fig. 1 C). Overall, despite 12 possible exchanges, one third of all silent mutations (3,157/9,394 or 34%) and eleven percent of all mutations (both silent and replacement) were silent C-to-T transitions. This is nearly threefold greater than the expected frequency of 3.8% for silent C-to-T mutations based on the

genetic code or analysis of non-IgV genes (Fig. 1 C, χ^2, P < 0.001). Thus the ratio of replacement (R) to silent (S) mutations for C-to-T exchanges (1/3 R to S) was opposite to that of all mutations that occurred as 3/1 R to S (9,394 silent in 28,307 total mutations, or 33%). It is notable that G-to-A exchanges (noncoding strand C-to-T) were also skewed to cause an increased frequency of silent mutations relative to that predicted by the genetic code, but not as dramatically as the skewing of C-to-T mutations (Fig. 1, A and B). In conclusion, C-to-T exchanges occurred as predominantly silent mutations, suggesting that the IgV genes have evolved to avoid replacements due to the initiating AID-mediated deamination of Cs.

C-to-T transitions in the CDRs are predominantly silent, whereas all other exchanges in the CDRs are predominantly replacements

Regional accumulation of mutations is a hallmark SHM of functional antibody genes. This tendency is believed to be due in large part to selection against mutations in the structurally important FWRs, which would be predicted to disrupt the structure of the Ig, causing the BCR–deficient cells
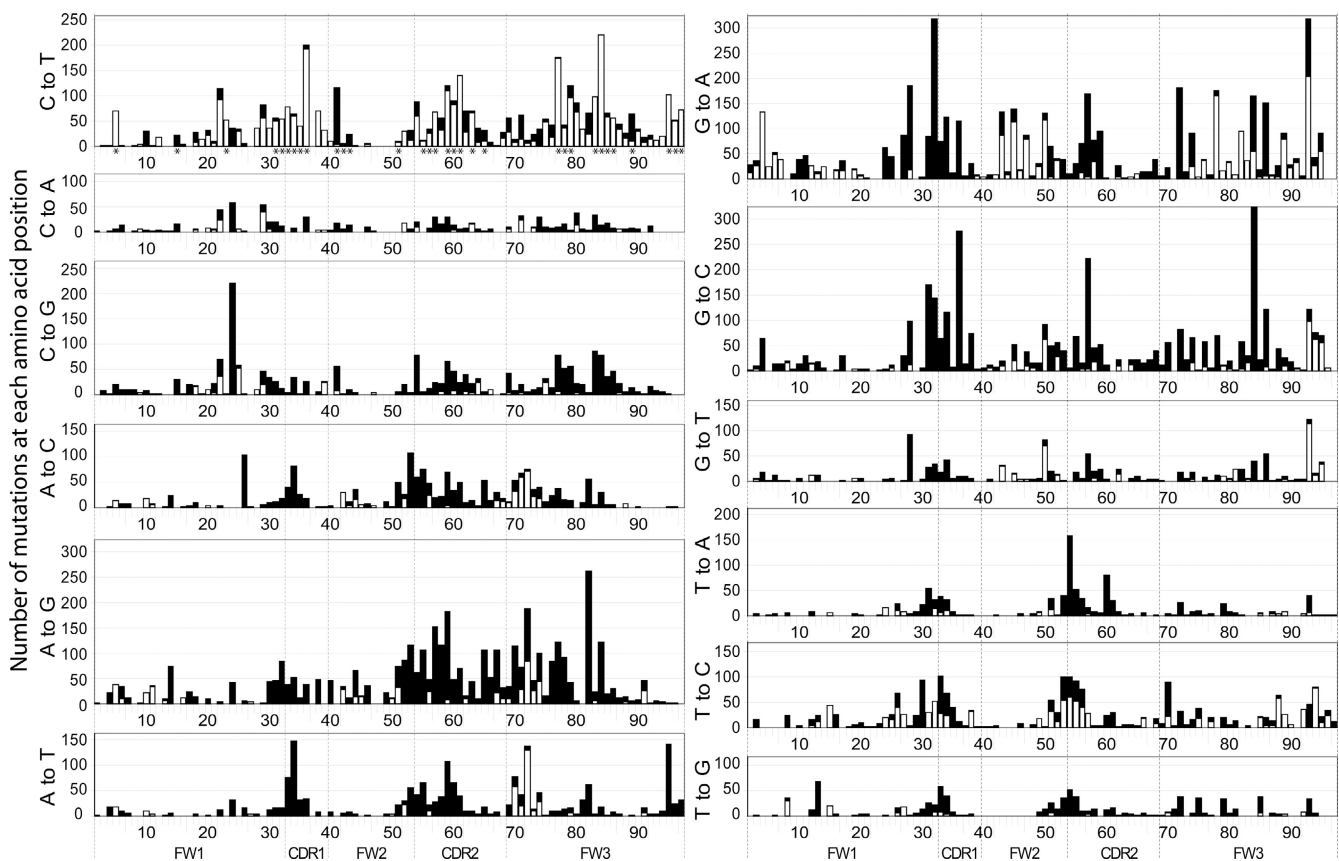

Figure 2. **A high frequency of silent C-to-T exchanges in the CDRs corresponds to mostly replacement mutations by all other exchanges.** The distribution of silent (white portion of bars) and replacement (black portion of bars) mutations. Asterisks indicate amino acid positions where most of the C-to-T exchanges occurred at AID hotspot motifs (WRC). Whereas C-to-T exchanges are nearly all silent, G-to-A exchanges are predominantly silent in the FWRs only.

to die (36). Both silent and replacement mutations accumulated in the CDRs whereas mutations that occur in the FWRs are preferentially silent (Fig. 2). However, unlike other exchanges that are predominantly replacement mutations in the CDRs, C-to-T exchanges in the CDRs are nearly all silent (Fig. 2, white portions of bars). In addition, most silent C-to-T mutations in the FWRs are found within the last several codons proximal to the CDRs (Fig. 2, codons 27–30 of FWR1, codon 51 of FWR2, codons 94–96 of FWR3) or within the hypervariable portion of FW3 between codons 76–85. The accumulation of mutations and apparent AID targeting at residues 76–85 of FWR3 is not surprising as this region is known to be highly variable and mutable in humans and has been demonstrated to contribute to the antigen combining surface for certain antigens, and therefore has even been suggested by some to be "CDR4" (42). Strikingly, coincident with the bias for silent C-to-T exchanges in the CDRs is a significant bias for replacements caused by all other exchanges. Bias for silent mutations in the FWRs is likely due to selection against B cells with loss of function mutations in the FWRs. However, the silent mutation frequency for G-to-A exchanges (indicative of AID-induced C deaminations on the noncoding strand) is particularly high, suggesting more than just selection. Indeed, as described below, Gs are preferentially placed in the FWRs so that mutation to A is more likely to be silent. In conclusion, although the SHM mechanism may be initiated by frequent AID-induced deaminations of C nucleotides, amino acid replacements due to C deamination are avoided in the CDR and FWRs of the coding strand and in the FWRs of the noncoding strand. Although a portion of these mutations may have resulted from selection against replacement mutations overall, the analyses presented below demonstrate that in fact targeting of AID activity by the V gene sequences themselves have precisely evolved to direct much of these mutational biases.

The genetic code favors silent C–to–T exchanges
Targeting of the SHM process such that most C-to-T mutations are silent may be intrinsic to the genetic code or to normal biases in codons used to encode most genes. To address this possibility, the results of all possible mutations to the 61 codons that encode amino acids (excluding the three stop codons) were scored. Thus, each nucleotide of each codon was scored for the result of exchanges to the other three nucleotides and the frequency that each of the 12 exchanges could cause silent or replacement mutations were tallied (Fig. 1 C). Of the 48 Cs used in various codons of the genetic code, C-to-T exchanges at 20 of these Cs will result in silent mutations. Thus if all codons are used equivalently to encode peptides, 56% of Cs if mutated to thymidine would cause amino acid substitutions, 2% nonsense or stop codons, and 42% would cause silent mutations (Fig. 1 C). This is despite a ratio of ∼3/1 for replacement versus silent mutations for all other exchanges (24.8% silent, 71.6%

amino acid replacement, and 3.6% nonsense mutations). If all codons are used stochastically to encode genes then this bias in the genetic code for silent mutations following C-to-T exchange should be reflected in all genes. To address this possibility, 17 genes that were not IgVH genes were randomly chosen and it was similarly determined what the sum result of all possible exchanges would be (Fig. 1 C). Surprisingly the codons used by these genes have an even greater tendency of C-to-T mutations to be silent than that predicted by the genetic code in that a 1/1 ratio of silent to replacement mutations following C-to-T exchange is expected (3,092 of 6,227 Cs). This analysis was based on the compiled codons used by the following genes: RNase A, myoglobin, β-actin, MAP4 kinase, p53, NFκB, cyclin D2, Raf, GAPDH, collagen, DNA polymerase β, CD40 ligand (CD154), Fas ligand, CD38, the immunoglobulin Fcγ receptor 1, IgG constant region, and AID itself. In conclusion, the genetic code is biased so that C-to-T mutations are more frequently silent relative to all other base exchanges. Thus the SHM mechanism may have evolved to use AID, which is a C-deaminating enzyme (a C-to-T mutator) to generate the initiating lesion in order to allow mutational targeting, but with a minimum of amino acid replacements. However, bias in the genetic code for C-to-T transitions to be silent accounts for only a portion of the 77% of C-to-T exchanges observed from SHM that are silent, and provides no insight as to why G-to-A exchanges would be preferentially silent.

Biases in codon usage target C-to-T and G-to-A exchanges
Natural selection may have evolved the IgVH genes to have a biased distribution of C nucleotides so that when mutated to thymidine, silent mutations preferentially result. To address this possibility the V genes themselves were analyzed to determine what mutations could occur. The entire set of IgVH molecules in this study was separated into individual codons to determine if indeed V genes have a biased distribution of C and G nucleotides (unmutated, germline counterparts of the IgVH genes were analyzed). Tables S2 through S6 present the counts and analyses of each codon used by variable gene region (available at http://www.jem.org/cgi/content/full/jem.20042483/DC1). The entire data set of 1,919 V genes analyzed included a total of 185,243 codons. Because each codon consists of three nucleotide positions, and each of these nucleotide positions can exchange to any one of the three other nucleotides, there is a total possibility of 1,667,187 potential exchanges: (185,243) × (3 positions) × (3 other bases) = 1,667,187). In addition, each C was scored for the frequency in context (within the V genes) that they were preceded by a W (A or T) and then an R (A or G) and thus were part of a hotspot for AID activity as previously described (23, 27). Each C was also scored for how frequently it occurred in a coldspot. Guanines were similarly scored for occurrence in reverse-complement hotspots and coldspots. For example, for FWR1 in Table S2 (available at http://www.jem.org/cgi/content/full/jem.20042483/DC1), the

A Predicted result of CDR C-to-T exchanges:

B Predicted result of FWR C-to-T exchanges:

C Predicted result of CDR G-to-A exchanges:

D Predicted result of FWR G-to-A exchanges:

E Predicted result of all other CDR exchanges:

F Predicted result of all other FWR exchanges:

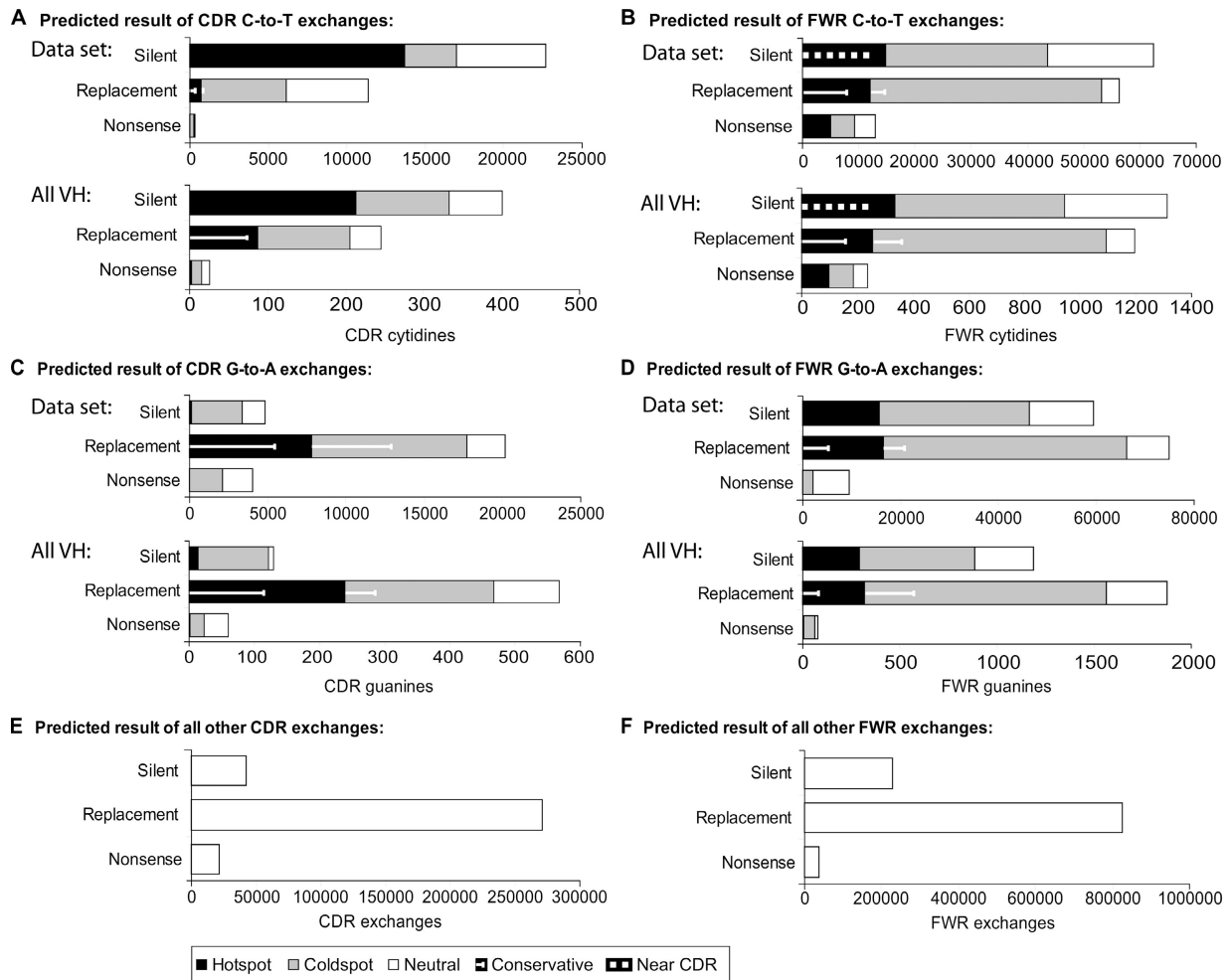■ Hotspot ▨ Coldspot □ Neutral ⊟ Conservative ⊞ Near CDR

Figure 3. Codon bias plus AID hotspot and coldspot targeting. Predicted effect of C-to-T (A and B), G-to-A (C and D), or other exchanges (E and F) in the CDRs and FWRs occurring in the germline (unmutated) counterparts of the 1,919 IgVH genes analyzed herein referred to as data set (compiled from the data in Tables S2 through S6). "All VH" genes refers to the 47 functional human germline VH genes to demonstrate that although the "Data set" contained certain VH genes more frequently than others, it is representative of the mutation pattern predicted for all VH genes. "Silent" bars indicate predicted frequency of silent mutations, "Replacement" indicates predicted frequency of amino acid replacement mutations, and "Nonsense" represents mutations where the exchange would result in stop codons. Black portions of bars represent proportion of Cs or Gs found within AID hotspot motifs, gray portions represent proportion of Cs or Gs in AID coldspots, and white represents Cs or Gs unaffected by either hotpots or coldspots (Neutral). White T bars represent the proportion of replacements that would result only in conservative amino acid replacements. The dashed line in B represents the proportion of Cs occurring in silent hotspots that are near the CDRs or in the hypervariable portion of FWR3 described in the text.

position two or "P2" of the TCC codon is listed as a coldspot 100% of the time. This means that every instance of TCC that occurred in the FWR1 of all 1,919 V_H genes analyzed was preceded by either a C or a G from position 3 of the previous codon such that the central C (T<u>C</u>C) was part of the AID coldspot motif: "ST<u>C</u>." Fig. 3 graphically summarizes the findings of Tables S2 through S6. To avoid confusion, hereafter "predicted mutations" will refer to the sum of all possible exchanges that could have occurred to the V genes analyzed, whereas "actual mutations" will refer to the 28,307 mutations that actually did occur. As not all VH genes are equally represented in the data set of 1,919 VH genes ana-

lyzed, there may be biases in the predicted frequency of mutations reported. To generalize the findings for all VH genes, one each of the 47 functional germline human VH genes were pooled and analyzed for comparison. Thus "data set" refers to the 1,667,187 exchanges that could have occurred to the germline counterparts (unmutated) of the 1,919 VH genes sequenced, and "all VH" refers to all possible exchanges that can occur in the 47 functional human VH genes.

Strikingly, within the CDRs, 67% (21,401/32,170) of the Cs in the data set are in positions where mutations to thymidine are silent, and 60% (12,929/21,401) of these same Cs are associated with AID hotspots (Fig. 3 A, Silent

bar). The frequency of C-to-T exchanges resulting in silent mutations is significantly greater than the expected frequency predicted by the genetic code or the frequency in non–Ig genes (Fig. 1 C), and more than any other exchange in the CDRs (Fig. 3, C and E; all comparisons by χ^2 with P < 0.0001). Of the remaining Cs in the CDRs of the data set that would produce replacement mutations when mutated to thymidine (10,477), half (5,101) are associated with AID coldspot motifs and thus are less likely to be mutated (Fig. 3 A, gray portion of the Replacement bar). It is also notable that although 700 of the 10,477 Cs that would cause replacements if mutated to thymidine are found in AID hotspot motifs (black portion of Replacement bar), most (539/700 or 77%) of these would result only in alanine to valine or hystidine to tyrosine exchanges and so are not as likely to disrupt the structure of the Ig molecule. The proportion of conservative replacement mutations are represented by white "Ts" within the black portion of the Replacement bar in Fig. 3. Thus, 99% (13,468/13,629) of all CDR-localized AID hotspot motifs are associated with Cs that if mutated to thymidine will result in silent or conservative amino acid exchanges. Targeting of C deamination to avoid amino acid replacements in the CDRs is evident primarily for the coding strand in that the predicted frequency of silent mutations due to the G-to-A transition (noncoding strand C-to-T) is similar to that of other exchanges (Fig. 3 C, Data set, and Fig. 3 E). However, it is notable that nearly half of the CDR replacements predicted to occur following a G-to-A transition (9,032/19,057), including 70% of those at AID hotspots (5,190/7,390), would result only in conservative amino acid replacements (V-to-I, S-to-N, and R-to-H; Fig. 3 C, Data set, white T bars). In addition, approximately half (9,354/19,057) of the Gs are in AID coldspots (Fig. 3, gray portion). Thus, G placement in the CDRs appears to avoid nonconservative amino acid replacements as well. In conclusion, the most striking result of this analysis is that the codons most commonly used to encode IgVH gene CDRs are highly biased such that most C nucleotides occur in positions on either DNA strand where if deaminated they either will not cause amino acid replacements or the replacements will be preferentially conservative in nature.

The region that most intricately supports AID induction of SHM without accumulation of amino acid replacements is CDR1 in which 99% (5,682/5,731) of Cs are in positions where mutation to thymidine will be silent or conservative, and of these 90% (5,116/5,682) are associated with AID hotspot motifs. Only 48 of 5,731, or <1% of all CDR1 Cs are in positions where AID deamination would cause a nonconservative replacement mutation, and of these, 39 of 48 (81%) are within AID coldspots and thus are not likely to be mutated. In conclusion, the Cs are intricately positioned in the CDR1 so there are ample sites for AID-induced C deamination to initiate the mutation process, but a minimum of AID-mediated replacement mutations.

Codon usage in the FWRs biases for cytidine placement in silent positions that are predominantly coldspots for AID

Like the CDRs, codon use in the FWRs is biased in that Cs are positioned so that transition to thymidine will result in a minimum (only 35% or 43,429/124,401) of nonconservative amino acid replacement (Fig. 3 B, Data set, Replacement, except portions overlaid with white T-shaped bars representing conservative replacements). Significantly, 84% (36,518/43,429) of these "C-to-T replacement" Cs are found in AID coldspot motifs (Fig. 3 B, Data set, Replacement, gray portion of bar) that we speculate further protect from the initiation of mutation in the FWRs. Thus Cs are placed such that only 35% of C-to-T mutations will cause amino acid replacements, which is significantly less frequently than for all other exchanges in the FWRs, and less than if all codons were used at a random frequency (56% for the genetic code; Fig. 1 C, all comparisons by χ^2 with P < 0.0001).

Also similar to the CDRs most (55% or 68,691/124,401) Cs in the FWRs would result in silent mutations or only conservative amino acid replacements upon mutation to thymidine (Fig. 3 B, Data set, Silent bar and white T regions of the replacement bar, χ^2, P < 0.05 compared with the predicted frequency of silent C-to-T mutations based on the genetic code). Note that unlike the CDRs, 43% of the "C-to-T silent" Cs are in AID coldspots (gray) and likely to avoid induction of SHM by AID activity on the FWRs. The 31% of the FWR Cs at C-to-T silent positions found in AID hotspots (black portion of Silent bar in Fig. 3 B, dashed bar) are found in positions adjacent to the CDRs and thus would prompt AID deamination of Cs targeting SHM to the CDRs, or are within the hypervariable portion of FWR3 (42). Interestingly, the remaining 10% (12,281/124,401) of Cs if mutated to thymidine would result in nonsense mutations (Fig. 3 B, Nonsense bar), which is significantly greater (χ^2, P < 0.0001) than the 1% (212/33,114) of CDR Cs leading to nonsense mutations (Fig. 3 A, Nonsense bar). It is notable that 6% (3 of 48) of Cs in the genetic code would result in nonsense mutations upon transition to thymidine, accounting for some of this bias. In fact, the most frequently occurring C in either FWR1 or FWR2 is associated with the glutamine codon CAG that if mutated to TAG (C-to-T exchange) forms a stop codon (Tables S2 and S6, available at http://www.jem.org/cgi/content/full/jem.20042483/DC1). Thus as discussed below, it appears that the FWRs may have evolved such that loss of function is preferable to excessive accumulation of replacements due to C-to-T exchanges following AID deamination.

The placement of G nucleotides in the FWRs is much more biased than in the CDRs such that G-to-A mutations were also predicted to be silent at a significantly greater frequency (P < 0.0001) than other exchanges (besides C-to-T), and thus replacement mutations due to noncoding strand C deaminations are avoided. In addition, as with C nucleotides, most of the Gs in the FWRs are found in reverse complement coldspot motifs (Fig. 4 D, gray portions of both
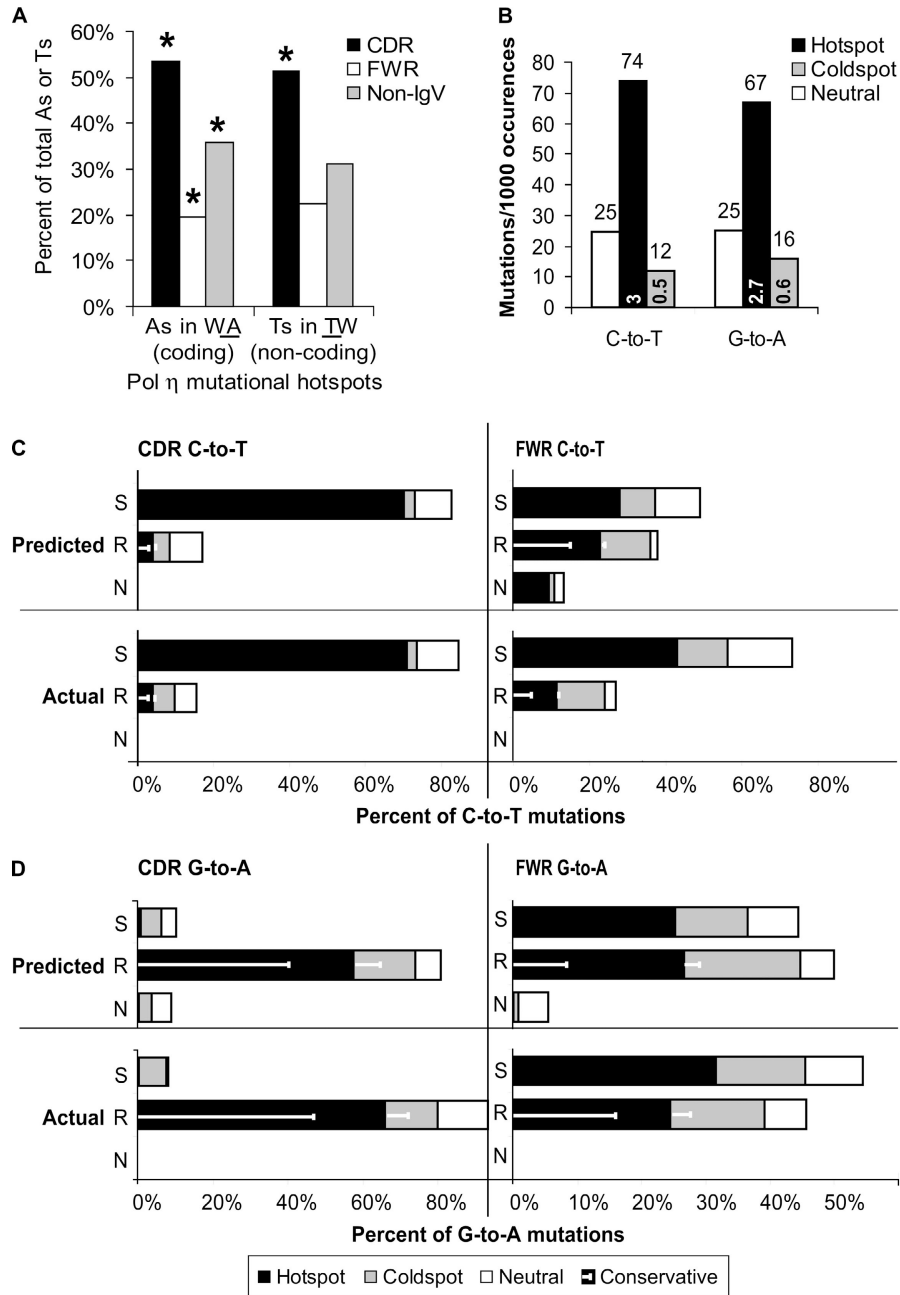
Figure 4.   Targeting of SHM. (A) Like C and G nucleotides, A and T nucleotides are also positioned to target SHM to the CDRs (black) and out of the FWRs (white). Polη tends to introduce point mutations at A in WA motifs during coding strand synthesis or at the reverse complement motifs (Ts in TW) during noncoding strand synthesis. Graph represents proportion of all As and Ts in WA or TW motifs from each region. The 17 non-Ig genes analyzed in Fig. 1 C were also analyzed for A and T positioning (gray). *, Significantly different by χ² test. (B) Analysis of silent mutation frequencies for Cs and Gs in AID hotpots, coldspots, and in neutral sequence contexts. The fold effect relative to neutral positions of hotspots and coldspots on mutation frequency is indicated at the base of the respective bars. (C and D) C-to-T and G-to-A mutation frequencies are remarkably predictable. There is no evidence for somatic selection of the CDR mutation frequencies and only moderate selection for increased silent and conservative mutations to the FWRs. The predicted frequency of AID-induced C-to-T (C) or G-to-A (D) in the CDRs or FWRs is based solely on codon composition of the unmutated germline genes compared with the "Actual" accumulation of C-to-T or G-to-A mutations in the 28,307 mutations sequenced. "Predicted" represents all possible mutations to the 1,919 VH genes analyzed (the data set values of Fig. 3, A and B) adjusted for the increased frequency of mutations at AID hotspots (3-fold for C-to-T and 2.7-fold for G-to-A) or decreased frequency at coldspots (0.5- or 0.6-fold for C-to-T and G-to-A).

Silent and Replacement bars), which would reduce the frequency of any mutations.

In conclusion, both C and G nucleotides appear to be positioned in the FWRs to avoid the initiation of SHM and to avoid replacements due to C deaminations on either DNA strand. Thus, the well-characterized reduced mutation frequency and preferential silent mutations in the FWRs are due not just to selection against loss-of-function mutations as previously believed, but also has evolved to utilize codons that target AID activity by precise placement of Cs and Gs in the V genes.

Prediction of mutations at cytidine and guanosine nucleotides for all VH genes is similar to the data set analyzed

As described above, the data set analyzed was not equally representative of all VH genes. The observation that most Cs and Gs are positioned to minimize the impact of C–to–T and G–to–A mutations on the Ig protein, whereas targeting the initiation of SHM to the CDRs and away from the FWRs is quite similar for all VH genes as for the data set (Fig. 3, compare Data set to All VH).

The AT phase of SHM is also targeted to the CDRs and out of the FWRs

As shown in Fig. 1 A and Fig. 2, there is a predominance of A–to–G and T–to–C exchanges over other AT mutations. As described above, polη is now known to be important for AT mutations and is suspected to polymerize excised patches of DNA following repair of AID-induced C deaminations (9–13, 18). The most common error of Polη is to mispair Gs rather than As with Ts, particularly if the preceding nucleotide is an A or T. This results in A–to–G mutations at W$\underline{A}$ hotspots (underlined A is mutated) if the coding strand is synthesized, and T–to–C mutations at $\underline{T}$W hotspots if the noncoding strand is synthesized. The intricate placement of C and G nucleotides would seem to preclude targeted placement of A and T nucleotides; however, this is not the case. Surprisingly, within the CDRs, 54% and 51% of As and Ts, respectively, are in W$\underline{A}$ or $\underline{T}$W Polη hotspots, whereas only 20% of As and 23% of Ts are found in polη hotspots of the FWRs (Fig. 4 A). For comparison, 36% of As are preceded by A or T (W$\underline{A}$) and 31% of Ts precede A or T ($\underline{T}$W) in the 17 non–IgV genes analyzed above. In conclusion, in addition to targeting AID-induced C and G mutations, the IgVH genes have also evolved to target AT mutations to the CDRs and to avoid AT mutations in the FWRs.

Hotspot/coldspot effects on mutation frequency

As described above, many C and G transitions likely arose due to the direct activity of AID (29). Analysis of only silent mutations provides relative mutation frequencies independent of selective processes. Thus analysis of silent C–to–T and G–to–A mutation frequencies should provide a direct assessment of AID targeting to hotspot and coldspot motifs. As

indicated in Fig. 4 B, these silent mutations included 2,098 C-to-T mutations in 28,430 hotspots (frequency = 74 mutations occurred for each 1,000 silent C-to-T hotspots), 383 mutations at 32,405 coldspot Cs (12 mutations occurred for each 1,000 silent C-to-T coldspots), and 596 mutations at 24,184 Cs that were neither hotspots nor coldspots ("neutral") giving a neutral frequency of 25 mutations per 1,000 C-to-T silent Cs. Thus, hotspot Cs were mutated to thymidine at threefold the frequency of neutral Cs, and coldspots were only mutated half as frequently. Similarly, silent G-to-A mutations occurred at 67 per 1,000 reverse complement hotspots (1,054/15,744), at 16 of 1,000 reverse complement coldspots (537/34,068), and at 25 of 1,000 neutral positions (303/12,022), or at a ratio of 2.7 to 1 mutations for hotspots versus neutral and 0.6 to 1 for coldspots. The frequency of hotspot and coldspot mutations corresponds well with previous in vivo and in vitro analyses (23, 32, 33).

The actual frequency and types of C–to–T and G–to–A mutations is highly predictable based solely on analysis of the VH gene sequences

The analyses above suggest that much of C and G transition mutations are actually targeted by the IgVH gene sequences themselves rather than selected for function during affinity maturation. To determine the relative impact of V gene sequence versus selection during an immune repsonse, the frequencies of C–to–T and G–to–A mutations predicted for the various regions (Fig. 3, Data set graphs) was adjusted to factor in the increased mutation frequency at hotspots (3-fold increase for Cs and 2.7-fold increase for Gs) and decreased for coldspots (0.5- and 0.6-fold decrease for Cs and Gs, respectively). For the CDRs, the adjusted prediction of C–to–T and G–to–A mutation spectrum was virtually identical to the actual frequencies observed (Fig. 4, C and D). The predicted mutations for the FWRs were also quite similar to the actual mutations, although some effect from selection was indicated by the increased accumulation of silent mutations beyond predicted (Fig. 4, C and D), and the preservation of critical cytseine residues in the FWRs (Table S1). In conclusion, as C and G transitions represent the initiating step of SHM, this analysis suggests that a good portion of mutational targeting to the CDRs and out of FWRs may actually have resulted due to evolutionary selection rather than somatic selection as commonly thought.

DISCUSSION

The initiating phase I of SHM is believed to result from the deamination of IgV gene Cs to deoxyuracil by AID that can lead to accumulation of C-to-T and G-to-A transitions if not repaired. Although targeting of SHM hotspots and coldspots has been the topic of previous analyses (33, 37, 38, 43), the selective pressures driving evolution of IgV gene sequences to best support the activity of a mutation process initiated by C deamination have not been previously considered. During phase II of SHM the resulting dU/dG mismatches can be re-

paired by error-prone DNA repair processes leading to accumulation of mutations of all nucleotides. Although these repair processes are error-prone, their purpose is repair, and so the fidelity of DNA synthesis cannot be too imprecise. Thus in order to introduce mutations other than those introduced from C deamination by AID (CG transitions), there must be a great excess of C deaminations. For example, short-patch excision repair in mammals involves excision and resynthesis of 25–30 nucleotides of DNA around a mismatch, and the polymerases involved such as Polη have an error rate of approximately one error per 500 nucleotides synthesized (18). Thus assuming no further modulation of the repair processes, for each deaminated C repaired, the chance that another error is introduced could be as low as only 5–6% (25/500 or 30/500). Thus as many as 20 C deaminations, most of which would presumably be repaired, would be required for each mutation introduced at A and T bases. With this in mind we hypothesize that without placement of C and G nucleotides to silent or conservative positions, there would be a bias in the spectrum of amino acid replacements to involve codons containing Cs and Gs. The result would be reduced diversity that would limit the generation of high affinity antibodies, causing a significantly increased length of time before a high affinity B cell clone could be generated to mediate secondary immune responses. Thus the observed targeting normalizes CG replacement mutations to correspond with the occurrence of AT replacements, significantly increasing the efficiency of affinity maturation.

Multiple factors appear to have simultaneously influenced the intricate evolution of IgVH genes to support the SHM process (summarized in Fig. S1, available at http://www.jem.org/cgi/content/full/jem.20042483/DC1). Although there was significant accumulation of C-to-T–induced silent mutations, due to precise placement of Cs, replacements due to C-to-T transitions were much less frequent than expected. Noncoding strand Cs were also clearly placed to avoid G-to-A transitions leading to replacements in the FWRs, and Gs are placed in the CDRs such that most G-to-A–induced replacements involve only conservative amino acid exchanges or are avoided completely by placement at coldspots. As the role of AID in SHM as an RNA or DNA editor is still controversial, it should be noted that other polymerases involved (ι and ζ) or other targeting mechanisms may preferentially introduce transitions at CG base pairs, accounting for the targeting observed. The complexity of these evolved biases in codon use are compounded by the precise concomitant hotspot/coldspot targeting of both AID activity and the errors typical of Polη to maximize the accumulation of mutations in the CDRs and minimize mutations to the FWRs. Thus, because of evolved biases in codon usage to target AID activity, the role of AID as an initiator of the SHM process is retained although its role as a direct mutator is reduced. Thus the V genes have evolved to normalize the replacement mutations induced directly by AID to occur at a similar frequency to replacements from the repair phases of the SHM mechanism.

As indicated in Fig. 4, the distribution and pattern of C and G transitions in the CDRs and FWRs is highly predictable based on analysis of the VH gene sequences alone. Thus, as C and G transitions are indicators of the AID induction of SHM, much of selection for functional mutagenesis to the CDRs and out of FWRs may have occurred during evolution rather than somatically. The IgVH gene data set includes the sum of 1,919 IgVH genes that were randomly cloned and sequenced from the functional B cell repertoire, and thus certain IgVH genes are represented more than others compared with the sum of 47 known functional VH genes (Fig. 4, All VH). As indicated in Fig. 3, the data set has both a modestly greater frequency of Cs placed in hotspots if silent and placed in coldspots if they would cause amino acid replacements, suggesting an even more dramatic avoidance of replacements due to C-to-T in the practical repertoire. This observation could be explained by somatic selection for B cells that use IgV genes that most efficiently support affinity maturation. However, it is well known that particular VH genes are preferentially recombined (44) and we and others have previously demonstrated that certain genes including the biased proportion of those analyzed herein are preferentially used in the B cell repertoire (39, 45). Thus the genes preferentially used for all B cells may be the best substrates of mutation, possibly representing an even greater complexity of evolutionary selection.

The most frequently used codon in the structurally important FWRs is CAG, which encodes glutamine. A C-to-T exchange of this codon will result in the TAG amber stop codon. Thus there may be an evolved bias to use codons that would cause stop mutations upon C-to-T exchange in the FWRs, suggesting a preference for loss-of-function rather than loading of AID and initiation of the mutator. This compares to the CDRs where initiation of SHM at Cs deaminated by AID without gross structural disruption is beneficial, and codons may be used to avoid C-to-T induced stops. There are only three codons in which C-to-T mutation will result in nonsense mutation including the other glutamine (Q) codon, CAA, and the arginine codon CGA. Structurally, the most similar and often interchangeable amino acid to Q is asparagine (N), however, C-to-T mutation of the N codons do not result in stop codons. Interestingly, by the standard CDR and FWR definitions, there are threefold more Qs than Ns used in the FWRs and inversely, there is 1Q-to-3Ns found in the CDRs of the 47 functional VH genes. If the hypervariable amino acids in FW3 (Fig. 2, positions 76–85) are counted as CDR4 as previously proposed (42), the observation is even more dramatic where there is a 7Q-to-1N ratio in the FWRs and inversely a 1Q-to-4N ratio in the CDRs. Similar analysis of the non-IgV genes described above demonstrates that in most genes Qs and Ns occur at equal frequencies. Thus nonsense mutations disrupting Ig translation and presumably death of the B cell clone may be preferred to induction of SHM in the FWRs. This adaptation may have arisen to avoid production of structurally aberrant antibodies

that could result in immunoglobulin deposition disease. Inversely, preferential use of N residues over Q residues in the CDRs may have evolved to allow efficient induction of the mutator without grossly disrupting Ig structure.

In conclusion, the IgV gene sequences appear to have evolved to direct the induction of SHM by AID into the CDR regions, but not the FWRs, while minimizing direct amino acid replacements from the resulting C deaminations in all regions. Juxtaposed on the targeting of AID is targeting of AT mutations introduced by Polη. The efficiency of affinity maturation appears to be highly reliant on targeting of SHM by the sequence of the V gene substrate. Thus, evolutionary pressures have selected IgV genes to both maximize and minimize SHM in critical regions of antibody molecules.

MATERIALS AND METHODS

Cloning and sequencing. Total RNA from various B cell populations sorted by flow cytometry were reverse transcribed and then subjected to PCR using VH gene leader region specific sense primers and isotype specific antisense primers as previously described (3, 39, 46). All cDNA clones were sequenced using automated DNA sequencers from Applied Biosystems (ABI-377 or ABI 3730 DNA sequencers). We have previously reported analyses of the frequency of nucleotide insertions and deletions (3), receptor revision (46), and V gene selection (39) concerning many of these sequences. All sequences were reanalyzed.

Analysis of IgVH genes for somatic mutations. All IgVH gene sequence analyses were performed using software developed for this purpose with the Microsoft Visual Basic and SQL query programming languages. VH gene sequences were compared with the V gene sequence set from the Immunogenetics database (41) using a stand-alone version of the NCBI BLAST search engine (47). VH genes were analyzed through the FWR3 excluding the CDR3 as this region includes the variable/diversity/junctional gene recombination junction, and thus it is not clear if mismatched nucleotides resulted from somatic mutation or if they arose during diversifying processes associated with VDJ recombination.

Statistical analyses. All statistics were done using JMP version 5.01 (SAS Institute Inc.) and Microsoft Excel.

Online supplemental materials. Supplemental materials include a Microsoft Excel file listing all of the 28,307 mutations and reference sequences (Table S1), the codons used by each variable gene region (Tables S2 through S6), and Fig. S1 and legend described in the discussion section. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20042483/DC1.

REFERENCES

1. Griffiths, G.M., C. Berek, M. Kaartinen, and C. Milstein. 1984. Somatic mutation and the maturation of immune response to 2-phenyl oxazolone. *Nature*. 312:271–275.
2. Rajewsky, K., I. Forster, and A. Cumano. 1987. Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science*. 238: 1088–1094.
3. Wilson, P.C., O. Bouteiller, Y.J. Liu, K. Potter, J. Bancherau, J.D. Capra, and V. Pascual. 1998. Somatic hypermutation introduces insertions and deletions into immunoglobulin V genes. *J. Exp. Med*. 187: 59–70.
4. Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo. 2000. Class switch recombination and hypermutation require activation–induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*. 102:553–563.
5. Revy, P., T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labelouse, A. Gennery, et al. 2000. Activation–induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell*. 102:565–575.
6. Honjo, T., M. Muramatsu, and S. Fagarasan. 2004. AID: how does it aid antibody diversity? *Immunity*. 20:659–668.
7. Petersen-Mahrt, S.K., R.S. Harris, and M.S. Neuberger. 2002. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature*. 418:99–103.
8. Rada, C., M.R. Ehrenstein, M.S. Neuberger, and C. Milstein. 1998. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity*. 9:135–141.
9. Pavlov, Y.I., I.B. Rogozin, A.P. Galkin, A.Y. Aksenova, F. Hanaoka, C. Rada, and T.A. Kunkel. 2002. Correlation of somatic hypermutation specificity and A-T base pair substitution errors by DNA polymerase eta during copying of a mouse immunoglobulin kappa light chain transgene. *Proc. Natl. Acad. Sci. USA*. 99:9954–9959.
10. Rogozin, I.B., Y.I. Pavlov, K. Bebenek, T. Matsuda, and T.A. Kunkel. 2001. Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. *Nat. Immunol*. 2:530–536.
11. Yavuz, S., A.S. Yavuz, K.H. Kraemer, and P.E. Lipsky. 2002. The role of polymerase eta in somatic hypermutation determined by analysis of mutations in a patient with xeroderma pigmentosum variant. *J. Immunol*. 169:3825–3830.
12. Zeng, X., D.B. Winter, C. Kasmer, K.H. Kraemer, A.R. Lehmann, and P.J. Gearhart. 2001. DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat. Immunol*. 2:537–541.
13. Wilson, T.M., A. Vaisman, S.A. Martomo, P. Sullivan, L. Lan, F. Hanaoka, A. Yasui, R. Woodgate, and P.J. Gearhart. 2005. MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. *J. Exp. Med*. 201:637–645.
14. Diaz, M., L.K. Verkoczy, M.F. Flajnik, and N.R. Klinman. 2001. Decreased frequency of somatic hypermutation and impaired affinity maturation but intact germinal center formation in mice expressing antisense RNA to DNA polymerase zeta. *J. Immunol*. 167:327–335.
15. Zan, H., A. Komori, Z. Li, A. Cerutti, A. Schaffer, M.F. Flajnik, M. Diaz, and P. Casali. 2001. The translesion DNA polymerase zeta plays a major role in Ig and bcl-6 somatic hypermutation. *Immunity*. 14:643–653.
16. Faili, A., S. Aoufouchi, E. Flatter, Q. Gueranger, C.A. Reynaud, and J.C. Weill. 2002. Induction of somatic hypermutation in immunoglobulin genes is dependent on DNA polymerase iota. *Nature*. 419:944–947.
17. McDonald, J.P., E.G. Frank, B.S. Plosky, I.B. Rogozin, C. Masutani, F. Hanaoka, R. Woodgate, and P.J. Gearhart. 2003. 129-derived strains of mice are deficient in DNA polymerase iota and have normal immunoglobulin hypermutation. *J. Exp. Med*. 198:635–643.
18. Kunkel, T.A., Y.I. Pavlov, and K. Bebenek. 2003. Functions of human DNA polymerases eta, kappa and iota suggested by their properties, including fidelity with undamaged DNA templates. *DNA Repair (Amst.)*. 2:135–149.
19. Ta, V.T., H. Nagaoka, N. Catalan, A. Durandy, A. Fischer, K. Imai, S. Nonoyama, J. Tashiro, M. Ikegawa, S. Ito, et al. 2003. AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat. Immunol*. 4:843–848.
20. Ramiro, A.R., P. Stavropoulos, M. Jankovic, and M.C. Nussenzweig. 2003. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat. Immunol*. 4:452–456.

21. 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.

22. Bransteitter, R., P. Pham, M.D. Scharff, and M.F. 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.

23. Pham, P., R. Bransteitter, J. Petruska, and M.F. Goodman. 2003. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*. 424:103–107.

24. Chaudhuri, J., M. Tian, C. Khuong, K. Chua, E. Pinaud, and F.W. Alt. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature*. 422:726–730.

25. 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.

26. Sohail, A., J. Klapacz, M. Samaranayake, A. Ullah, and A.S. Bhagwat. 2003. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res*. 31:2990–2994.

27. Yu, K., F.T. Huang, and M.R. Lieber. 2004. DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *J. Biol. Chem*. 279:6496–6500.

28. 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.

29. Rada, C., J.M. Di Noia, and M.S. Neuberger. 2004. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell*. 16:163–171.

30. Peters, A., and U. Storb. 1996. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity*. 4:57–65.

31. Rogozin, I.B., and N.A. Kolchanov. 1992. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim. Biophys. Acta*. 1171:11–18.

32. Neuberger, M.S., M.R. Ehrenstein, N. Klix, C.J. Jolly, J. Yelamos, C. Rada, and C. Milstein. 1998. Monitoring and interpreting the intrinsic features of somatic hypermutation. *Immunol Rev*. 162:107–16.

33. Shapiro, G.S., and L.J. Wysocki. 2002. DNA target motifs of somatic mutagenesis in antibody genes. *Crit. Rev. Immunol*. 22:183–200.

34. Spencer, J., M. Dunn, and D.K. Dunn-Walters. 1999. Characteristics of sequences around individual nucleotide substitutions in IgVH genes suggest different GC and AT mutators. *J. Immunol*. 162:6596–6601.

35. Rogozin, I.B., and M. Diaz. 2004. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J. Immunol*. 172:3382–3384.

36. Lam, K.P., R. Kuhn, and K. Rajewsky. 1997. In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death. *Cell*. 90:1073–1083.

37. Wagner, S.D., C. Milstein, and M.S. Neuberger. 1995. Codon bias targets mutation. *Nature*. 376:732.

38. Kepler, T.B. 1997. Codon bias and plasticity in immunoglobulins. *Mol. Biol. Evol*. 14:637–643.

39. Zheng, N.Y., K. Wilson, X. Wang, A. Boston, G. Kolar, S.M. Jackson, Y.J. Liu, V. Pascual, J.D. Capra, and P.C. Wilson. 2004. Human immunoglobulin selection associated with class-switch and possible tolerogenic origins for C(Delta) class switched B cells. *J. Clin. Invest*. 113:1188–1201.

40. Liu, Y.J., O. de Bouteiller, C. Arpin, F. Briere, L. Galibert, S. Ho, H. Martinez-Valdez, J. Banchereau, and S. Lebecque. 1996. Normal human IgD+IgM- germinal center B cells can express up to 80 mutations in the variable region of their IgD transcripts. *Immunity*. 4:603–613.

41. Lefranc, M.P. 2003. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res*. 31:307–310.

42. Capra, J.D., and J.M. Kehoe. 1974. Structure of antibodies with shared idiotypy: the complete sequence of the heavy chain variable regions of two immunoglobulin M anti-gamma globulins. *Proc. Natl. Acad. Sci. USA*. 71:4032–4036.

43. Oprea, M., and T.B. Kepler. 1999. Genetic plasticity of V genes under somatic hypermutation: statistical analyses using a new resampling-based methodology. *Genome Res*. 9:1294–1304.

44. Yu, K., A. Taghva, and M.R. Lieber. 2002. The cleavage efficiency of the human immunoglobulin heavy chain VH elements by the RAG complex: implications for the immune repertoire. *J. Biol. Chem*. 277: 5040–5046.

45. Brezinschek, H.P., R.I. Brezinschek, and P.E. Lipsky. 1995. Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction. *J. Immunol*. 155:190–202.

46. Wilson, P.C., K. Wilson, Y.J. Liu, J. Banchereau, V. Pascual, and J.D. Capra. 2000. Receptor revision of immunoglobulin heavy chain variable region genes in normal human B lymphocytes. *J. Exp. Med*. 191: 1881–1894.

47. Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389–3402.