The Journal of Immunology

RESEARCH ARTICLE | FEBRUARY 15 2000

The Nucleotide-Replacement Spectrum Under Somatic Hypermutation Exhibits Microsequence Dependence That Is Strand-Symmetric and Distinct from That Under Germline Mutation[1] FREE

Lindsay G. Cowell; ... et. al

J Immunol (2000) 164 (4): 1971–1976.
https://doi.org/10.4049/jimmunol.164.4.1971

The Nucleotide-Replacement Spectrum Under Somatic Hypermutation Exhibits Microsequence Dependence That Is Strand-Symmetric and Distinct from That Under Germline Mutation[1]

Lindsay G. Cowell and Thomas B. Kepler[2]

Somatic mutation is a fundamental component of acquired immunity. Although its molecular basis remains undetermined, the sequence specificity with which mutations are introduced has provided clues to the mechanism. We have analyzed data representing over 1700 unselected mutations in V gene introns and nonproductively rearranged V genes to identify the sequence specificity of the mutation spectrum—the distribution of resultant nucleotides. In other words, we sought to determine what effects the neighboring bases have on what a given base mutates "to." We find that both neighboring bases have a significant effect on the mutation spectrum. Their influences are complicated, but much of the effect can be characterized as enhancing homogeneity of the mutated DNA sequence. In contrast to what has been reported for the sequence specificity of the "targeting" mechanism, that of the spectrum is notably symmetric under complementation, indicating little if any strand bias. We compared the spectrum to that found previously for germline mutations as revealed by analyzing pseudogene sequences. We find that the influences of nearest neighbors are quite different in the two datasets. Altogether, our findings suggest that the mechanism of somatic hypermutation is complex, involving two or more stages: introduction of mis-pairs and their subsequent resolution, each with distinct sequence specificity and strand bias. ***The Journal of Immunology,*** **2000, 164: 1971–1976.**

During affinity maturation, a subset of the Ag-responsive B cells experiences somatic hypermutation at the rearranged Ig locus (for a recent review, see Ref. 1). Somatic hypermutation is a fundamental component of the overall strategy of the immune system and is found in all organisms that possess rearranged Ag receptors, from cartilaginous fishes to humans. In spite of almost 30 years of effort, the molecular mechanisms involved remain shrouded in mystery. Clues to the identity of the "mutator" have thus been sought by indirect means. The mutational process exhibits a distinctive signature: particular microsequences are mutated at a higher frequency than others (2–4). In the absence of selection, the mutation rate at a given position depends on the chemical identity of the nucleotide at that position and on those of the nucleotides in its immediate neighborhood. Rogozin and Kolchanov (2) suggested two consensus motifs that promote hypermutation: RGYW and TAA. Subsequent research has focused primarily on the former motif, which is consistent with the commonly cited occurrence of hotspots at the serine codons AGY.

In addition, this sequence specificity has itself had a striking influence on the evolution of Ig V genes allowing an enhancement of their plasticity under affinity maturation. Codon bias differs between framework and complementarity-determining regions, with the result that the framework nucleotides are less mutable than those in the complementarity-determining regions (5–7). Direct counting of mutations accumulated in nonproductively rearranged Ig genes confirms that this difference hypothesized under relatively simple empirical models for mutability is indeed realized in a significant and observable way (8).

There is another aspect of the mutational mechanism that has the potential for providing a distinctive signature and thereby information about the underlying mechanisms of somatic hypermutation: the mutation spectrum. By this we mean the frequencies at which specific bases occur at a particular position, given that the original nucleotide at that position has mutated (again, under selection-free conditions). One aspect of this issue has been addressed already: the transition to transversion ratio has been determined to be approximately 2:1 (3, 9). We have analyzed a much larger dataset than has previously been considered and therefore can provide a detailed characterization of the mutation spectrum for all nearest-neighbor interactions. This dataset is comprised of four independent sets of nonfunctional Ig V sequences and Ig introns that have undergone somatic hypermutation free of selective pressure. Our analysis is based primarily on straightforward χ^2 tests of multidimensional contingency tables. We find that the spectrum depends not only on the identity of the mutating base, but also on the identity of the immediate 5′ and 3′ neighbors.

An effect of this sort has been documented for meiotic (germline) mutations by comparing a large dataset of human genes and their related pseudogenes (10). We find that the mutation spectrum and its context specificity for somatic hypermutation is very different from that observed in meiotic mutation. The context-dependent effects under somatic hypermutation can be very crudely summarized by the observation that the effect of flanking nucleotides is frequently to promote homogeneity of the local sequence. For example, a nucleotide within a homodimer is more likely to experience a transition than the same nucleotide in another context,

Biomathematics Program, Department of Statistics, North Carolina State University, Raleigh, NC 27695.

Received for publication August 13, 1999. Accepted for publication December 9, 1999.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by a Fulbright grant to L.G.C. and National Science Foundation Award MCB 9357637 to T.B.K.

[2] Address correspondence and reprint requests to Dr. Thomas B. Kepler, Biomathematics Program, Department of Statistics, Box 8203, North Carolina State University, Raleigh, NC 27695-8203. E-mail address: kepler@unity.ncsu.edu

0022-1767/00/$02.00

Table I. *Classification of nucleotides*

	σ	$\bar{\sigma}$	τ	$\bar{\tau}$
C	**C**	**G**	**T**	**A**
T	**T**	**A**	**C**	**G**
A	**A**	**T**	**G**	**C**
G	**G**	**C**	**A**	**T**

while a nucleotide neighbored by its complementary base is more likely to experience a transversion to its complement.

Somatic hypermutation shows a marked strand bias; adenines mutate more frequently than do thymines, for example. Recently, this simple observation has been complicated by the results of analyses finding correlation between the mutability of trinucleotide motifs (11) or quartets of the RGYW motif (12) and the mutability of the corresponding inverted complement motifs. This apparent symmetry is largely confined to symmetry between **G** and **C** nucleotides, so the tentative picture now drawn is that of a compound mechanism that mutates **A** and **T** in a strand-biased manner, but that mutates **G** and **C** without notable bias (13–15).

We find that the mutation spectrum exhibits very little strand bias at all. In particular, the symmetry between **A** and **T** is quite marked. We speculate that this difference between the targeting of somatic mutation and the resulting mutational spectrum is due to the action of multiple distinct mechanisms responsible for the biochemistry of somatic hypermutation, even beyond the multiple mechanisms hitherto postulated.

Materials and Methods

Datasets

The pooled set of somatically mutated sequences contains a total of 1721 mutations: 610 **A**, 452 **G**, 336 **C**, and 323 **T**. The sequences included in this analysis are as follows: murine *J-C* intron sequences containing 510 mutations (4), nonfunctionally rearranged human heavy chains containing 349 mutations (16), nonfunctionally rearranged human heavy and κ and λ light chains with 67, 319, and 84 mutations, respectively (8), and murine 3′ flanking region sequences (3′ VJ_κ1, 154 mutations (17, 18); 3′ J_H1, 162 mutations (19)) and *J-C* intron sequences (77 mutations (20, 21)). We performed comparisons with germline mutations using pseudogene data from Hess et al. (10).

Because the sequences come from a variety of sources and include both murine and human gene segments, we tested for whether the mutation spectrums from the different datasets have a similar distribution using the heterogeneity χ^2 test for pooling contingency tables (22) (see below for statistical methods). The tests found no differences at the 0.05 level for complete $4 \times 4 \times 4 \times 3$ tables for 5′ and 3′ adjacent nucleotides ($p = 0.167$ and $p = 0.158$, respectively). Thus, the pooling of data from all datasets is very unlikely to cause errors in the statistical inferences of interest to us here. This procedure does not provide an exhaustive comparison of the characteristics of each dataset, however, and should not be taken as positive evidence that they are identical in all respects. Further data might very well reveal differences between murine and human sequences or between exon and intron sequences, but the lack of differences under the heterogeneity χ^2 test does provide confidence that the effects discussed in this paper are not artifacts of the pooling process.

Statistical methods

To test independence of the mutation spectrum from the identity of the mutating nucleotide's 5′ or 3′ neighbor, we formed two $4 \times 3 \times 4$ contingency tables with 24 degrees of freedom each in which the rows categorize the identity of the 5′ or 3′ flanking base, respectively, the columns categorize the identity of the destination base (the base that a mutation is to) and the tiers categorize the mutating, or original, base. We used a χ^2 test of the null hypothesis of independence of neighboring base and destination base, conditional on mutating base (23). That is, for each mutating base *Y*, we tested whether the identity of the destination base is random with respect to the identity of the 5′ or 3′ neighbor base. When testing for conditional independence, the total χ^2 is the sum of the partial χ^2 values, which, in our case, represent the effect for each of the four mutating bases.

Table II. *Pooled somatic hypermutation data*[a]

	Mutating Base				
Effect of	**C**	**T**	**A**	**G**	Total
5′	12.16	11.45	**19.20**	**14.24**	**57.05**
	0.058	0.076	**0.004**	**0.027**	**0.0002**
3′	11.48	8.29	9.78	11.91	41.45
	0.075	0.218	0.134	0.064	0.015
Mutations	335	323	610	452	1721

[a] $\chi^2_{0.05,6} = 12.59$; $\chi^2_{0.05,24} = 36.42$. χ^2 and p values shown in bold indicate $p < 0.05$.

Under the null, each of these partial χ^2 values as well as the total is distributed like χ^2 with the appropriate number of degrees of freedom.

The available pseudogene data (10) take the form of substitution frequencies for the center base pair in each of the 32 base-pair triplets. The authors of that study "collapsed" the data by summing the frequencies for each triplet with that for its complement, thereby obviating the need to discriminate between the two DNA strands. For example, the mutation frequency of base **C** to base **T** in the context **ACG** was computed as

$$\frac{m(\mathbf{ACG} \rightarrow \mathbf{ATG}) + m(\mathbf{CGT} \rightarrow \mathbf{CAT})}{m(\mathbf{ACG}) + m(\mathbf{CGT})}$$

where $m(XYZ)$ is the total number of Y mutations when Y is flanked by X and Z, and $m(XYZ \rightarrow XY'Z)$ is the total number of mutations of Y to Y' when flanked by X and Z. To make comparison with this pseudogene data possible, the somatic hypermutation data was coerced into this format as well. For both the pseudogene data set and this collapsed somatic hypermutation data set, $4 \times 3 \times 2$ contingency tables were set up as described above, but in this case, there are just two mutating bases, **C** and **T**.

To test for differences between the two data sets, we constructed two 4-way tables ($2 \times 3 \times 2 \times 4$) in which the classifications are 1) data source (somatic hypermutation or pseudogene), 2) destination base, 3) mutating base, and 4) neighboring base (5′ and 3′ neighbors, respectively). We tested the null hypothesis of independence of data source and destination base conditional on both mutating base and neighboring base, that is, for each of the eight dimers XY (YZ), we tested whether the identity of the destination base is random with respect to the source of the data.

For graphical representation of the contingency tables, we computed the adjusted residuals[3] for each cell (23). These have the attractive property that they are distributed approximately as standard normal random variables under the null hypothesis. Thus, in studying the figures, values of z larger in absolute value than 1.96 are significant at the 0.05 level, those larger than 2.58 are significant at the 0.01 level, and so on.

We computed and tested Spearman's rank correlation statistic using SPLUS (MathSoft, Seattle, WA) to evaluate the degree to which the sequence specificity of the mutation spectrum is symmetric and to check for similarities between the mutation spectra of the somatically mutated Ig sequences and the germline mutated pseudogene sequences.

Parameterization of the mutation probabilities

For convenience, we have adopted the following notation. σ (for "self") will be used to designate the identity of the mutating base; $\bar{\sigma}$ (self-complement) will be used to designate the base's complement; τ (transition) will be used to designate the base's transition base; $\bar{\tau}$ (transition-complement) will be used to designate the complement of the base's transition base. For convenient reference, we have provided a translation guide in Table I.

The mutation spectrum for a given motif is characterized by the proportion of mutations of σ in the motif $X\sigma$ or σZ that are transitions

$$p_S = \frac{m(\sigma \rightarrow \tau)}{m(\sigma)},$$

[3] Adjusted residuals are defined as follows. For a contingency table with two factors, let the count in cell (i, j) be denoted n_{ij} and its expected value e_{ij}. Then the adjusted residuals z_{ij} are given by

$$z_{ij} = \frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}(1 - n_{i\cdot}/N)(1 - n_{\cdot j}/N)}}$$

where N is the total number of counts, $n_{i\cdot} = \Sigma_j n_{ij}$ and $n_{\cdot j} = \Sigma_i n_{ij}$. This rescaling of the residuals makes their marginal distributions under the null hypothesis approximately standard normal.

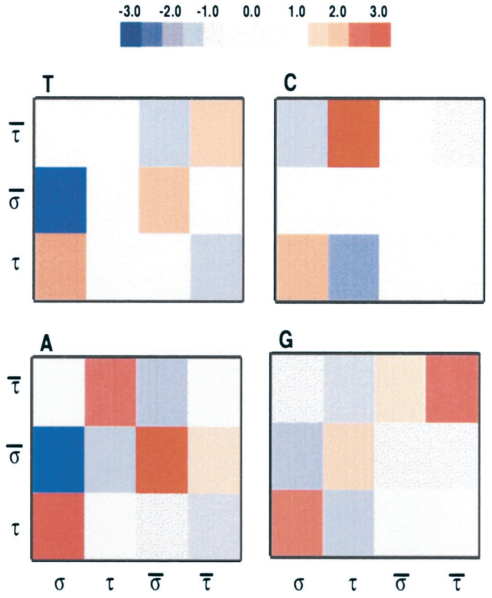

FIGURE 1. Adjusted residuals[3] for the test of conditional independence of mutation spectrum and 5′ neighbor. The residuals are encoded in colors given by the bar at the top of the figure. Under the null hypothesis (no effect of 5′ neighboring nucleotide), these residuals are distributed approximately as standard normals (*Z* statistics). Each of the panels represents a tier of the contingency table (see "Statistical methods") and is labeled in the *upper left corner* by the mutating base. Rows represent the destination base and columns represent the 5′ neighboring base; both are labeled according to the scheme described in Table I.

and the proportion of transversions of σ that are mutations to the complement of σ, $\bar{\sigma}$

$$p_C = \frac{m(\sigma \to \bar{\sigma})}{m(\sigma \to \bar{\sigma}) + m(\sigma \to \bar{\tau})}.$$

Results

Lack of independence between the destination base and the context of the mutating base

The mutation spectrum at a given base is clearly not independent of the microsequence containing that base. The χ^2 values for the test of the effect of both 5′ and 3′ nucleotides are significant at the traditional 0.05 level (Table II). The *p* value for the conditional independence of the 5′ base and the destination base is smaller than 0.001. In fact, the effect of the 5′ base seems to be larger than that of the 3′ base for all four mutating bases. For the effect of the 5′ nucleotides, all of the partial χ^2 values are much larger than expected under independence, as are the partials testing the effect of the 3′ nucleotide on **C** and **G**. The effect of the 3′ base on the mutation spectrum of **A** and **T** appears to be weak, consistent with the relative weakness of the 3′ nucleotide in general.

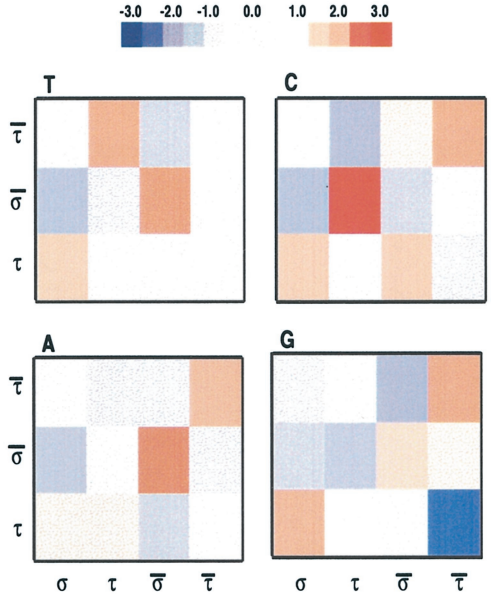

FIGURE 2. Adjusted residuals for the test of conditional independence of mutation spectrum and 3′ neighbor. Details as in Fig. 1.

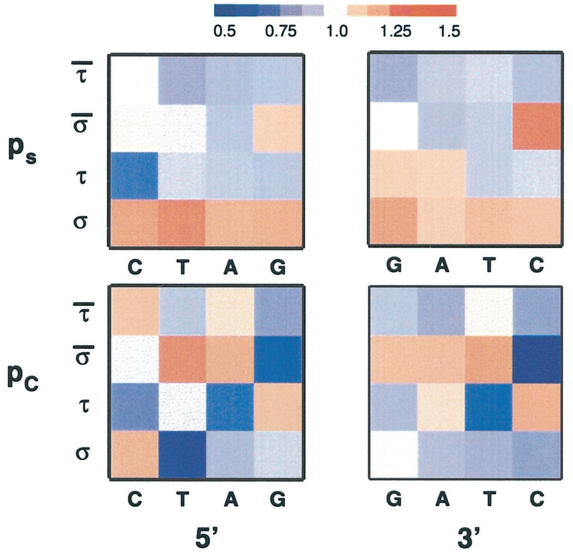

FIGURE 3. Scaled transition frequencies (p_S; *upper panels*) and transversion to complement (given transversion) frequencies (p_C; *lower panels*). Rows represent the neighbor and columns represent the mutating base. Note that the columns are reversed between the panels on the *left* (representing the effect of 5′ neighbors) and those on the *right* (representing the effect of 5′ neighbors); under complementation symmetry (no strand bias) the *left* and *right panels* will be identical but for sampling variability. Note that the color coding (legend above) is different from that in Figs. 1 and 2.

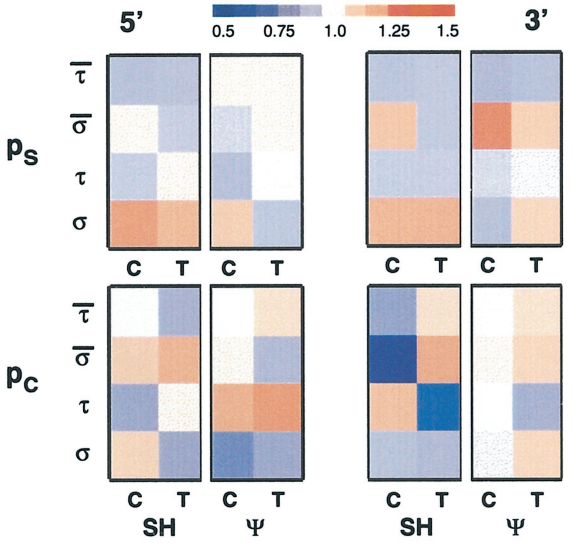

FIGURE 4. Comparison of somatic hypermutation data (SH; shown as the *left panel* of each paired set) and pseudogene mutation data (Ψ; shown as the *right panel* of each paired set). Other details of this figure are as in Fig. 3.

Adjusted residuals for each of the eight partial contingency tables are shown as contour plots in Figs. 1 and 2. Although each table shows several prominent deviations from the expected, the most consistent deviation, common to all the tables, is seen in the enhancement of the transition probability for homodimers (the lower left element of each plot). This effect appears to be stronger when the 3′ base in the dimer is mutating. The corresponding decrease in the transversion probability is concentrated into a reduction of the transversion to complement probability ($\sigma \rightarrow \bar{\sigma}$) while a reduction in the probability of $\sigma \rightarrow \bar{\tau}$ transversions is insignificant. This enhancement of the transition probability for homodimers is consistently observed when the four datasets are analyzed separately: 27 of the 32 relevant adjusted residuals are positive (9 of them significantly so), 2 of them are approximately 0, and 3 of them are negative (none of them significantly so; data not shown).

Effect on the transition frequency, p_S

Scaled transition frequencies, p_S, computed by dividing the transition frequency for the mutating base in a dimer by its background transition frequency, are plotted in the contour plots in Fig. 3, *upper panel*. The red row at the bottom of each figure reveals that both nucleotides in $\sigma\sigma$ dimers have an elevated transition frequency p_S. For example, the 3′ **C** in the homodinucleotide **CC** is more likely to mutate to **T** than would be expected based on the transition frequency of base **C** when considered out of context.

Effect on the proportion of transversions to $\bar{\sigma}$, p_C

The *lower panel* of Fig. 3 shows the scaled proportion of transversions that are to the complement of the mutating base, p_C, computed as described above for the scaled p_S. The lower left panel of this figure suggests that when **T** or **A** is preceded by its complement ($\bar{\sigma}\sigma$, for σ = **T, A**), the proportion of transversions to complement is enhanced. In fact, a preceding **T** may inhibit p_C for all bases except its own complement **A**, while it enhances p_C when preceding **A**. A similar pattern appears to hold for **A**. A 5′ **A** may enhance p_C for all bases except itself, while inhibiting p_C when preceding itself.

The row in the lower right panel of Fig. 3 corresponding to the mutating base being followed by its complement, $\sigma\bar{\sigma}$, suggests that a base having its complement as its 3′ neighbor base enhances the transversion to complement frequency. Bases **T**, **A**, and **G** mutate more often to their complement when they are followed by their complement. The exception is for the transversion **CG** → **GG**. Of the 336 mutations of **C**, only 8 occur for **C** in the dimer **CG**; only two of these are transversions, none of these is a mutation to **G**. The expected number of **CG** → **GG** transversions is 1.8.

T 3′ of **C**, **T**, or **G** inhibits the transversion to complement frequency, just as a 5′ **T**; a 3′ **A** appears not to have the same effect as a 5′ **A**.

Strand symmetry

One of the characteristic features of somatic hypermutation is its apparent strand asymmetry. For example, mutations are found at adenines much more frequently than at thymines (3, 9). This has been taken as evidence for strand bias of the mutator, that mutations are introduced preferentially into one strand. Recent analyses suggest a more complex picture than this; the sequence specificity of the mutator indicates some degree of symmetry, especially between **G** and **C** mutations (11, 12).

Our analyses indicate a high degree of strand symmetry in the effect that neighboring nucleotides have on mutation spectra. Inspection of Fig. 3 reveals a great deal of similarity between the effects of *X* on *Y* in the dinucleotide *XY* and those of the complements $\bar{X}$ on $\bar{Y}$ in the complementary dinucleotide $\bar{Y}\bar{X}$; e.g., the effect of **A** preceding **C** is quite similar to the effect of **T** following **G**. The figures have been constructed in such a way that corresponding plots will be identical if symmetry under complementation is exact.

Table III. *Pooled somatic hypermutation data, collapsed*[a]

Effect of	Mutating Base **C**	Mutating Base **T**	Total
5′	13.59	17.32	30.91
	0.035	0.008	0.002
3′	23.71	27.28	50.95
	10^{-4}	10^{-4}	10^{-6}
Mutations	788	933	1721

[a] $\chi^2_{0.05,6} = 12.59$; $\chi^2_{0.05,12} = 21.03$. χ^2 and p values shown in bold indicate $p < 0.05$.

Indeed, computing the Spearman's rank correlation for scaled p_S, in which the scaled p_S for *XY* (*Y* mutating) is paired with scaled p_S for $\bar{Y}\bar{X}$ ($\bar{Y}$ mutating), we find that there is a moderately high correlation, $r = 0.51$, and in spite of the small number of points, it is significantly different from zero ($p = 0.047$). This means that when the transition probability p_S for the dimer *XY* (*Y* mutating) is elevated (or inhibited), so is that for the complementary dimer $\bar{Y}\bar{X}$ ($\bar{Y}$ mutating) suggesting that *Y* in *XY* is replaced by *Y*′ on both strands of the DNA with similar probabilities. The complementary pairs for p_C, formed in just the same way are even more strongly correlated: $r = 0.77$, $p = 0.003$.

Collapsed somatic hypermutation data

To compare the mutation spectrum of the somatic hypermutation data with that of the pseudogene data, we first analyzed the somatic hypermutation data in the format of the pseudogene data as described above to ensure that the statistical patterns were not lost or changed by combining reverse complement motifs. The 4 × 3 contingency tables were analyzed as described in *Materials and Methods*, and, as can be seen from Table III, the mutation spectrum remains nonrandom with respect to the target context in the collapsed dataset. The total effect χ^2 values for 5′ and 3′ neighbors are significantly large, as are the partial χ^2 values.

The dominant patterns observed in the full somatic hypermutation dataset are evident in the collapsed dataset as well. Adjusted residuals for the four contingency tables reveal that in all four tables, the transition probability is enhanced for $\sigma\sigma$ homodimers, while the probability of $\sigma \rightarrow \bar{\sigma}$ transversions is reduced (data not shown). The enhancement of p_S for $\sigma\sigma$ homodimers is also evident in Fig. 4, and, as was the case with the full dataset, $\bar{\sigma}\sigma$ and **T**$\bar{\sigma}$ dimers have an enhanced $\sigma \rightarrow \bar{\sigma}$ transversion frequency (Fig. 4).

Table IV. *Pseudogene data*[a]

Effect of	Mutating Base **C**	Mutating Base **T**	Total
5′	165.96	96.00	261.96
	10^{-16}	10^{-16}	10^{-16}
3′	340.28	90.68	430.96
	10^{-16}	10^{-16}	10^{-16}
Mutations	10,517	8,327	18,844

[a] $\chi^2_{0.05,6} = 12.59$; $\chi^2_{0.05,12} = 21.03$. χ^2 and p values shown in bold indicate $p < 0.05$.

Table V. *A comparison of the collapsed somatic hypermutation data with the pseudogene data*[a]

Effect of	Mutating Base C	Mutating Base T	Total
5′	26.72	78.76	105.48
	10^{-3}	10^{-13}	10^{-13}
3′	31.13	64.84	95.97
	10^{-4}	10^{-10}	10^{-12}

[a] $\chi^2_{0.05,8} = 15.51$; $\chi^2_{0.05,16} = 26.30$. χ^2 and p values shown in bold indicate $p < 0.05$.

A comparison of the somatic hypermutation data with the pseudogene data

The mutation spectrum of the pseudogene data is dependent on the target context; the patterns of dependency, however, are not the same as those for the somatic hypermutation data. To test for independence of the mutation spectrum and the target context in the pseudogene data, 4×3 contingency tables were analyzed as described in *Materials and Methods*. The χ^2 values are shown in Table IV; all of the χ^2 values are highly significant indicating that the mutation spectrum does depend on the target context. To test for differences between the mutation spectra of the two datasets, 2×3 contingency tables were set up for each of the 8 dimer motifs, and their partial χ^2 values used to compute the relevant total χ^2 values (described in *Materials and Methods*). Both of the total effect χ^2 values are significant (see Table V), indicating that the mutation spectra of the two datasets do differ.

Having established that the effects are not the same in the two datasets, we now want to ask whether they are correlated. Although they are clearly not the same, there may be significant similarities. We tested this with two Spearman's correlation tests, one on the scaled p_S values and one on the scaled p_C values. Neither test showed any significant correlation (p_S: $r = 0.21$, $p = 0.42$; p_C: $r = 0.06$, $p = 0.82$). Thus, it appears that the mutation spectra of germline mutations and of somatic hypermutation are unrelated.

Discussion

We have shown that the mutation spectrum of somatically mutated Ig genes is non-random with respect to the primary sequence context of the mutating base. The identity of the destination base depends not only on the identity of the mutating base, but also on the identity of both the 5′ and the 3′ neighbor bases. There are several patterns that emerge. Rather that undertake to describe every detail, we have limited ourselves to description of the most prominent and potentially interesting patterns.

The most consistent dependence is in the increased tendency of homodimers to mutate via transitions and the attendant decrease of homodimer mutations to the complementary base. This is true regardless of the identity of the mutating base and whether it appears in the 5′ or 3′ position within the dimer. For example, both **A**s in the homodimer **AA** have an enhanced probability of mutating to **G** and a reduced probability of mutating to **T** when compared to **A**s in any other context.

Another feature is the tendency of **A/T** mixed dinucleotides to homogenize. That is, when **A** flanks **T** (or **T** flanks **A**), the mutating base tends to become that neighbor; e.g., **AT** → **AA** is enhanced. This effect is not seen for **G/C** dinucleotides.

There is a striking symmetry under complementation, especially for **A** or **T** mutating. This is in notable contrast to what has been suggested for the targeting of mutation, in which **A** and **T** seem to be more asymmetric than **G** and **C**.

We compared the effects of neighboring bases under somatic hypermutation to that observed in pseudogenes and found not only that the patterns differ, but that there is, in fact, no correlation between them. Thus, there is no evidence here to support the hypothesis that the mechanism of hypermutation is essentially related to normal DNA repair pathways. Our own analysis, however, of the targeting of somatic mutation shows a strong correlation between the microsequence specificity of the mutation targeting under somatic mutation and that under meiotic mutation.[4]

It is our hope that the patterns we have begun to elucidate will help identify the elusive mechanism(s) of somatic hypermutation. While we are not prepared to propose specific hypotheses in this regard, we would like to offer one general observation. The behavior of the mutation spectrum under complementation symmetry is rather different from the behavior of the targeting of mutations. Whereas the spectrum is strongly symmetric, especially between **A** and **T** nucleotides, there is a strong disparity between the targeting of mutation at **A** and **T** nucleotides. We suggest that this fact makes quite plausible the notion that the mechanism is complex, involving at least two stages, the introduction of mutations followed by their resolution. For example, the first stage might involve the insertion of mis-paired bases, in a way that depends on the local microsequence. A second stage might consist of the recognition and resolution of the noncanonical base pairs, again in a local microsequence-dependent manner, but one that is wholly different from that of the first stage. This scheme is consistent with our analysis[4] of hypermutation targeting, which further suggests that the first stage is closely related to the "targeting" of mutation under meiotic processes. Within the first stage, there may be two distinct mechanisms as suggested by others: one stage with strong strand bias, the other acting symmetrically (11); these two mechanisms may effect **A/T** and **G/C** nucleotides differently (13–15). We present evidence for an additional stage during which the distribution of resultant nucleotides is determined in a sequence-specific and strand-independent manner.

Acknowledgments

We thank Claudia Berek and Latham Claflin for sharing data prior to publication and for critical reading of the manuscript. We also thank Garnett Kelsoe for stimulating discussions.

References

1. Berek, C. 1999. Affinity maturation. In *Fundamental Immunology*, 4th ed. W. Paul, ed. Lippincott-Raven Publishers, Philadelphia, p. 863.
2. Rogozin, I. B., and N. A. Kolchanov. 1992. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighboring base sequences on mutagenesis. *Biochim. Biophys. Acta 1171:11*.
3. Betz, A. G., C. Rada, R. Pannell, C. Milstein, and M. S. Neuberger. 1993. Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. *Proc. Natl. Acad. Sci. USA 90:2385*.
4. Smith, A. S., G. Creadon, P. K. Jena, J. P. Portanova, B. L. Kotzin, and L. J. Wysocki. 1996. Di- and trinucleotide target preferences of somatic mutagenesis in normal and autoreactive B cells. *J. Immunol. 156:2642*.
5. Wagner, S. J., C. Milstein, and M. S. Neuberger. 1995. Codon bias targets mutation. *Nature 376:732*.
6. Kepler, T. B. 1997. Codon bias and plasticity in immunoglobulins. *Mol. Biol. Evol. 14:637*.
7. 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*.
8. Cowell, L. G., H. J. Kim, T. Humaljoki, C. Berek, and T. B. Kepler. 1999. Enhanced evolvability in immunoglobulin V genes under somatic hypermutation. *J. Mol. Evol. 49:23*.
9. Lebecque, S. G., and P. J. Gearhart. 1990. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5′ boundary is near the promoter, and 3′ boundary is about 1 kb from V(D)J gene. *J. Exp. Med. 172:1717*.

[4] T. B. Kepler, M. Oprea, and L. G. Cowell. The targetting of somatic hypermutation closely resembles that of meiotic mutation. *Submitted for publication.*

10. Hess, S. T., J. D. Blake, and R. D. Blake. 1994. Wide variations in neighbor-dependent substitution rates. *J. Mol. Biol. 236:1022*.
11. Milstein, C., M. S. Neuberger, and R. Staden. 1998. Both DNA strands of antibody genes are hypermutation targets. *Proc. Natl. Acad. Sci. USA 95:8791*.
12. Dörner, T., S. J. Foster, N. L. Farner, and P. E. Lipsky. 1998. Somatic hypermutation of human immunoglobulin heavy chain genes: targeting of RGYW motifs on both DNA strands. *Eur. J. Immunol. 28:3384*.
13. 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*.
14. Sale, J. E., and M. S. Neuberger. 1998. TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. *Immunity 9:859*.
15. Spencer, J., M. Dunn, and D. K. Dunn-Walters. 1999. Characteristics of sequences around individual nucleotide substitutions in IgV_H genes suggest different GC and AT mutators. *J. Immunol. 162:6596*.
16. 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*.
17. Rickert, R., and S. Clarke. 1993. Low frequencies of somatic mutation in two expressed V_κ genes: unequal distribution of mutation in 5′ and 3′ flanking regions. *Int. Immunol. 5:255*.
18. Weber, J. S., J. Berry, T. Manser, and J. L. Claflin. 1991. Position of the rearranged V_κ and its 5′ flanking sequences determines the location of somatic mutations in the J_κ locus. *J. Immunol. 146:3652*.
19. Weber, J. S., J. Berry, T. Manser, and J. L. Claflin. 1994. Mutations in Ig V(D)J genes are distributed asymmetrically and independently of the position of V(D)J. *J. Immunol. 153:3594*.
20. Weber, J. S., J. Berry, S. Litwin, and J. L. Claflin. 1991. Somatic hypermutation of the JC intron is markedly reduced in unrearranged kappa and H alleles and is unevenly distributed in rearranged alleles. *J. Immunol. 146:3218*.
21. Wu, P., and L. Claflin. 1999. Promoter-associated displacement of hypermutations. *Int. Immunol. 10:1131*.
22. Zar, J. H. 1996. *Biostatistical Analysis*, 3rd ed. Prentice-Hall, Upper Saddle River, NJ.
23. Everitt, B. S. 1992. *The Analysis of Contingency Tables: Monographs on Statistics and Applied Probability, 45*. Chapman & Hall, London.