Strand bias in Ig somatic hypermutation is determined by signal sequence within the variable region

Arthur Kar-Keung Ching, Pik-Shan Li, Wood-Yee Chan1, Chun-Hung Ma, Susanna Sau-Tuen Lee2, Pak-Leong Lim and Yiu-Loon Chui

Clinical Immunology Unit and Sir Y. K. Pao Centre for Cancer, Prince of Wales Hospital, 1Department of Anatomy and 2Department of Biochemistry, Chinese University of Hong Kong, Shatin, NT

Keywords: germinal centre B cells, κ light chain, mutation hotspot, transgenic mice, RGYW, WRCY

Abstract

Ig genes undergo hypermutation with a nucleotide preference of A over T for mutation on the coding strand. As only with concomitant strand bias can such nucleotide bias be observed, Ig gene hypermutation is generally accepted as a strand-specific process, for which the mechanistic basis remains unknown. It has previously been shown that different non-Ig sequences replacing the LVJ region of an Ig transgene to various extents are targeted for hypermutation with similar mutation frequencies. However, the nucleotide bias characteristic of Ig hypermutation was not found in two of the three such sequences studied. To test whether it is the DNA sequences of the non-Ig substrates that determine the pattern of nucleotide bias in hypermutation or whether the LVJ sequence may contain element(s) that confer strand bias, we have added back all the replaced LVJ sequences to one of the transgenes, $L_κ–Vgpt^*$, that expresses no strand bias in hypermutation and studied the outcome. The results show that the gpt sequence in the presence of the complete LVJ sequence hypermutates differently from the same sequence in $L_κ–Vgpt^*$ where 84% of the LVJ was replaced. The main difference is the resumption of strand bias characteristic of Ig hypermutation. Thus, whether or not a substrate sequence manifests strand bias in hypermutation is not inherently determined by the substrate DNA sequence. This indicates the presence of special element(s) within the LVJ that confer strand bias.

Introduction

Rearranged Ig genes in B cells stimulated by T-dependent antigens undergo high rates of mutation. The mutation domain, which centers at the antigen-binding V(D)J, is ~1–2 kb in length (1–5). Some of the mutations will result in higher binding affinities to the antigen, and B cells carrying these antibodies will survive and secrete the antibodies following a selection process that normally takes place in germinal centers. The improvement of serum antibody in binding antigens over the course of an immune response is known as affinity maturation (6,7). Although the molecular mechanisms of hypermutation are not yet understood, some important cis elements that influence hypermutation are now known. The essential role of Ig enhancers in driving hypermutation has been best studied in a murine κ light chain gene where the absence of either the intronic or the 3' enhancer drastically reduces or abolishes hypermutation (8–10). In that and other Ig transgene systems, a functional promoter has also been proved to be important in determining the position of the mutation domain (11,12). It has been shown that the promoter need not be of Ig origin nor of the RNA polymerase II-dependent type (10,13). The use of chimeric Ig transgenes where different parts of V(D)J were replaced by non-Ig sequences has, in addition, demonstrated that LVJ itself also does not contain elements required for targeting hypermutation (14).

Mutations found in the V(D)J region are not random, based on the analysis of silent mutations (15), non-productively rearranged alleles (16,17), flanking sequences in $J_κ$ clusters and 3’ sequences of rearranged V(D)J (18–20), as well as passenger Ig transgenes (14,21). These mutations were not subject to selection, and thus reflected only the interaction between the Ig sequence and the hypermutation machinery. Two general characteristics have been found. First, there is...
a higher frequency of mutation for the purine nucleotide, A, compared to the complementary pyrimidine, T, on the coding strand. For any preference in targeting one of the complementary nucleotides for mutation to be observable, a concomitant strand bias in the mutation has to take place. Thus, it has been generally accepted that hypermutation involves a strand-biased process (14,17,19–23). Second, there is a bias for mutation in some sequence motifs, of which the most notable is RGYW (R = A or G, Y = C or T, W = A or T) (14,16–21,24).

The use of non-Ig sequences as substrates for hypermutation allowed further probing into the intrinsic mutation preference of the hypermutation machinery. Using a series of Ig light chain (Lκ) transgenes in which the LVJ was replaced to different extents by non-Ig sequences such as bacterial gpt, nedd4 and human B- globin sequences, the non-random nature of mutations and the RGYW motif as a mutation hotspot have been confirmed (14,25). However, the strand bias shown in the hypermutation of Ig sequences is not found in the hybrid vigour F1CC mice were generated by crossing CBA/N male with C57BL/6 female mice. Fertilized eggs were obtained from mating of superovulated F1CC females with F1CC males. Vector-free Lκ-Vgpt*-3 was obtained by EcoRI digestion and electrophoresis into low melting point agarose gel (Gibco/BRL, Gaithersburg, MD), followed by Gelase (Epicerin, Madison, WI) digestion and centrifugation using a chromatrap 1000 column (Clontech, Palo Alto, CA). The DNA was then microinjected into the pronuclei of the fertilized eggs using a standard microinjection procedure. Transgenic mice were identified by PCR of the tail DNA using primers OX1BU and OX1BL, which flank the modified VκOX1/Lκ5 and confirmed by Southern blotting. Founders were then mated with BALB/c mice to obtain offspring for experiments.

Cell separation and sorting

For magnetic sorting of GL-7+ cells from Peyer’s patches, single-cell suspensions were first prepared from these lymphoid organs in sorting buffer (PBS containing 0.5% FCS, 2 mM EDTA and 0.01% NaN3, pH 7.2). The cell suspensions were then incubated with purified anti-mouse CD32/CD16 (Fcthe hypermutation of Ig sequences is not found in the

**Methods**

**Construction of transgene, Lκ-Vgpt*-3**

The transgene is a modified version of Lκ. Lκ is an EcoRI genomic fragment of the rearranged VκOX1/Jκ5 Ig gene containing the Cκ and both the intronic and 3’ enhancers (Fig. 1). The 459 bp gpt coding sequence containing an inactivation mutation (Val 86 GTT \rightarrow Asp GAT) near the middle was amplified from Lκ-Vgpt* (14) using primers NcoI (5’-CAG CCA CCA TGG GCG AAA AAT ACA TCG TCA CCT GGG ACG T) and OXP3 (5’-CTG GCA GTA ATA AGT GGC AGC ATC TTC). Lκ-Vgpt* is a transgene constructed by substituting part of the LVJ region of Lκ with the inactivated gpt* sequence (Fig. 1). After digestion with NcoI and EcoRV, the PCR product was used to exchange with the wild-type gpt (amplified from pSV2gpt) which had been inserted earlier into the FR1 of VκOX1/Jκ5 by sticky feet mutagenesis in M13mp18. The NcoI and EcoRV sites are respectively located at the 5’ end and within the 3’ region of gpt. The resulting fusion sequence of LVJ and the inactivated gpt was released from the vector by EcoRI and XbaI digestion and ligated to the XbaI–BamHI J-C intron/Cκ fragment and the ~8 kb SacI–EcoRI kE3’ sequence to form a single EcoRI fragment, Lκ-Vgpt*-3, in pUC18 (Fig. 1). The EcoRI and XbaI sites are respectively located at the 5’ end of the Ig fragment and within the J-C intron.

**Generation of transgenic mice**

Hybrid vigor F1CC mice were generated by crossing CBA/N male with C57BL/6 female mice. Fertilized eggs were obtained from mating of superovulated F1CC females with F1CC males. Vector-free Lκ-Vgpt*-3 was obtained by EcoRI digestion and electrophoresis into low melting point agarose gel (Gibco/BRL, Gaithersburg, MD), followed by Gelase (Epicerin, Madison, WI) digestion and centrifugation using a chromatrap 1000 column (Clontech, Palo Alto, CA). The DNA was then microinjected into the pronuclei of the fertilized eggs using a standard microinjection procedure. Transgenic mice were identified by PCR of the tail DNA using primers OX1BU and OX1BL, which flank the modified VκOX1/Lκ5 and confirmed by Southern blotting. Founders were then mated with BALB/c mice to obtain offspring for experiments.
transgenic Peyer’s patches was 8.4% (7.5–9.9%, n = 3). The mean percentage for the GL-7+ CD45R+ cells was 13.9% (6.2–21.1%, n = 8). All the selected cells were then digested using proteinase K in 500 µl of digestion buffer, and the DNA was precipitated and resuspended in 50 µl TE.

PCR and DNA sequencing

Amplification of the transgene was performed using the primers OX1BU and OX1BL which flank the modified VκOX1/Jκ5 region. OX1BU (5'-CGG AAT TCT TCT CTC AGG TAA TAA ATC G) primes at 276 bp 5’ to the transgene Vκ translation start site and OXBL (5'CCC CTC CAA ATC TCC CAC TT) is located within the J–C intron. 5 µl of genomic DNA was used for each 50 µl PCR reaction consisting of 35 cycles of 94°C for 60 s, 50.8°C for each 50 µL for 60 s, 72°C for 60 s and 72°C for 60 s with a last incubation step of 72°C for 10 min. The DNA polymerase used was either Taq polymerase (Gibco/BRL) or the proofreading enzymes, Pwo (Roche, Indianapolis, IN) and Pfu Turbo (Stratagene, La Jolla, CA). The PCR products were purified, digested by EcoRI and XbaI, and subcloned into M13mp18 for DNA autorosence (Applied Biosystems, Perkin-Elmer, Foster City, CA).

Major hotspots

The three major or most prominent hotspots for each mutation analysis data set were defined by the three nucleotide positions having the highest variability among the clones of the data set. For line 7.5 of Lκ-Vgpt*3-3, the three prominent hotspots Glu70(III) and Leu71(I) and Val73(III) were respectively found in 26, 40 and 32% of the 50 analyzed clones of the line. In mouse #103 of line 8.14 of Lκ-Vgpt*3-3, the three most prominent hotspots are the same Glu70(III) and Leu71(I) and Val73(III). These hotspots were respectively found in 17%, 9% and 15% of the 59 clones analyzed. The same three major hotspots were also found in one of the two independent data sets of the previously published Lκ-Vgpt*3-3. In mouse #103 of line 8.14 of Lκ-Vgpt*3-3, the two most prominent hotspots found among 39 clones are also identical to two of the three hotspots in Lκ-Vgpt*3-3. Glu70(III) and Leu71(I) were each present in 37% of the 33 analyzed clones for data set #1. For the other data set, #2, of Lκ-Vgpt*3-3 (14), the two most prominent hotspots found among 59 clones are also identical to two of the three hotspots in Lκ-Vgpt*3-3. Glu70(III) and Leu71(I) were each present in 38% of the 33 clones. The third and the weakest one, Ala107(II), is unique to this data set and was found among 18% of the clones. Val73(III) lagged behind at 13%.

Statistical analysis

The mutation frequencies between A and T, and between G and C in a 258 bp gpt region from nucleotide 148 to 405 (numbered from the first base of gpt ATG initiation codon) were compared using $\chi^2$ with Yates correction. For the Lκ-Vgpt*3-3 transgenic line 7.5 and mouse #103 of line 8.14, 50 and 59 clones respectively were included for analysis. After excluding the three major hotspots from analysis, the nucleotide composition of the 258 bp region for each clone was 60 for A, 69 for T, 67 for G and 59 for C. For the two mutation data sets of Lκ-Vgpt* obtained from line LxNG (14,19), 33 clones were represented in data set #1 and 39 clones in data set #2. After excluding the same three major hotspots, the nucleotide composition of the same 258 bp region of data set #1 is identical to that of Lκ-Vgpt*3-3. For data set #2 containing a different third major hotspot, there is one more G but one less C in the nucleotide composition after excluding the three major hotspots. The same $\chi^2$ method with Yates correction was used to compare the distribution of nucleotide substitutions between line 7.5 (Lκ-Vgpt*3-3) and the combined data sets, #1 and #2, of Lκ-Vgpt*3. After excluding the three major hotspots, the nucleotide composition for line 7.5 (as well as data set #1) within and outside the motifs respectively is 11 and 49 for A, 17 and 52 for T, 13 and 54 for G, and 16 and 43 for C. For data set #2, due to the change of one hotspot from a G in WRCY motif to a C outside the motif, the nucleotide composition concerning A and T remains unchanged, while the numbers of G and C within and outside the motifs are 14 and 54, 16 and 42 respectively. Values of $P < 0.05$ were considered as statistically significant.

Results

Design of the transgene construct, Lκ-Vgpt*3-3

The new transgene, Lκ-Vgpt*3-3, described here is analogous to the previously published Lκ-Vgpt* (14), both of which contain a non-Ig sequence, gpt. This gpt is the 459 bp complete amino acid coding sequence for the bacterial xanthine guanine phosphoribosyl transferase. The sequence contains a single base substitution near the middle to render the protein functionally inactive. The only difference between the two transgenes is that in the new one all the LVJ sequence remains, while in Lκ-Vgpt* 84% of the sequence was substituted by the gpt. As shown in Fig. 1, the gpt was incorporated into the LVJ region of the VκOX1 genomic fragment (Lκ), that contains all the known elements essential for transcription and hypermutation, in the transgenes in two different ways. In the previous transgene, Lκ-Vgpt*, the gpt replaced the LVJ region from its first base of initiation codon to approximately the middle of FR3. In the present transgene, Lκ-Vgpt*-3, the same gpt was inserted into the FR1 without substitution of any LVJ sequence. Both the parent Lκ and Lκ-Vgpt* transgenes have been shown to be capable of hypermutation (9,14). However, in the published 258 bp sequence segment of gpt from nucleotide 148 to 405 (numbered from the first base of gpt ATG initiation codon) of Lκ-Vgpt*, no mutational bias for A over its complementary T on the coding strand is found. This is in contrast with the Ig V(D)J sequence targets where the mutation of A is often found more frequently than that of T (14,17,19–22,25). Presented below are the results of our present study on hypermutation of the gpt sequence in the new transgene, Lκ-Vgpt*3-3.

Hypermutation of Lκ-Vgpt*3-3

Six mice from two independently derived Lκ-Vgpt*3-3 transgenic lines, 7.5 and 8.14, were used in the study. Line 7.5 was generated with co-injection of Lκ-Vgpt*3-3 and the parent Lκ constructs. It contains one to three copies of the transgenes. Line 8.14 was generated with microinjection of a single Lκ-Vgpt*3-3 construct. It contains three copies of the transgene. Peyer’s patch cells positively sorted for the GL-7+ phenotype by magnetic beads (MACS) were monitored by flow analysis for the percentage of the GL-7+ CD45R+ population (germinal centre B cells) (26–28). The variation in purity
of the GL-7⁺ CD45R⁺ population shown in Table 1 was due
to various degrees of contamination by the GL-7⁻ cells after
the magnetic sorting. For those GL-7⁺ cells, 92–96% of them
also expressed the CD45R⁻ phenotype. Transgenes in the
DNA of the sorted cells were amplified by PCR, subcloned
into M13 and sequenced for the same 258 bp gpt region as
the previously published Lκ–Vgpt* (14,19). Using the same
PCR conditions on transgenic mouse tail DNA, we have
determined the Taq polymerase error rate as 0.61/kb after 35
PCR cycles. To standardize the quality of the data used in
hypermutation analysis, we have uniformly applied the rule
that only those clones showing more than one mutation in the
258 bp gpt region were selected for hypermutation analysis.
This rule was applied to all the clones, regardless of whether
Taq or a proof-reading DNA polymerase (i.e. Pfu Turbo and
Pwo) was used, and also in selecting clones from the two
independent sets of previously published mutation data of
Lκ–Vgpt* (14,19). The databases, #1 and #2, were separately
constructed from two independent groups of mice of the
same transgenic line LκNG, bearing one copy of Lκ–Vgpt*.
The same number of PCR cycles as that for amplifying
the Lκ–Vgpt*-3 transgene (i.e. 35 cycles) was performed in
generating these databases. Taq DNA polymerase was used
in the PCR for database #1 and Pfu DNA polymerase for
database #2.

As shown in Table 1, there is no clear correlation between
the percentage of the clones carrying more than one mutation
(hypermutation-positive clones) and the purity of the GL-
7⁺ CD45R⁺ population among the five mice of the same transgenic line 7.5. The maximum percentage of the
hypermutation-positive clones was 36%. On the contrary,
the mutation frequencies shown by the mice of line 7.5 were
consistent among each other, averaging at 20.5/kb. This
figure is almost identical to the 20.3/kb shown by Lκ–Vgpt*
in the transgenic line LκNG carrying one copy of the transgene
(14). The mutation frequency of mouse #103 of line 8.14 is
~3-fold lower than that of line 7.5. This might reflect the
possibility that mouse #103 could be harboring more copies
of the Lκ–Vgpt*-3 transgene than the mice of line 7.5. As only
one or a small proportion of the multiple copies of the
transgene are preferentially targeted for hypermutation
(14,29), an increase in the number of the transgene copies,
as possibly in mouse #103 compared with the mice of line
7.5, should account for the reduction in average mutation
frequency of the transgene. Despite such a difference in
the average mutation frequency, the three most prominent
hotspots shown in the hypermutation of line 7.5 are identical
to those shown by mouse #103. The three hotspots are
Glu70(Ill) and Leu71(I) and Val73(Ill) (codon numbered from
the first ATG of the gpt sequence). Figure 2(A and B) shows
the nucleotide variability plots of the combined data of line
7.5 and mouse #103. The same three most prominent hotspots
were also found in one of the two sets of the mutation data

Table 1. Transgene mutations in PCR clones derived from the transgenic mice

<table>
<thead>
<tr>
<th>Line no.</th>
<th>Mouse no.</th>
<th>Purity of GL-7⁺ CD45R⁺</th>
<th>PCR enzyme used</th>
<th>No. (%) of clones with more than one mutationa</th>
<th>Total no. of mutations from clones carrying more than one mutation</th>
<th>Mutations per 10³ bp²</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5</td>
<td>62</td>
<td>92%</td>
<td>Pwo</td>
<td>13 (35)</td>
<td>82</td>
<td>24.4</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>70%</td>
<td>Pfu Turbo</td>
<td>20 (36)</td>
<td>90</td>
<td>17.4</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>77%</td>
<td>Pfu Turbo</td>
<td>5 (11)</td>
<td>24</td>
<td>18.6</td>
</tr>
<tr>
<td></td>
<td>162</td>
<td>81%</td>
<td>Pfu Turbo</td>
<td>7 (24)</td>
<td>38</td>
<td>21.0</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>36%</td>
<td>Pfu Turbo</td>
<td>5 (9)</td>
<td>31</td>
<td>24.0</td>
</tr>
<tr>
<td>8.14</td>
<td>103</td>
<td>NDb</td>
<td>Taq</td>
<td>59 (18)</td>
<td>114</td>
<td>7.5</td>
</tr>
</tbody>
</table>

aScored within the 258 bp sequence segment of the gpt region from nucleotide 148 to 405 (numbered from the first base of gpt ATG initiation codon).
bCell purity not checked.
cFor clones carrying more than one mutation.
Hypermute of L¢-Vgpt*-3 exhibits strand bias

When the numbers of A and T mutations relative to their respective unmutated nucleotides in the 258 bp gpt region of L¢-Vgpt*-3 in line 7.5 are compared using \( \chi^2 \) analysis, a statistically significant bias for the mutation of A is found (\( P = 0.0002 \)) (Table 2). No bias is found between the mutation of G and C (\( P = 0.3168 \)). This analysis is based on 220 mutations pooled from 50 clones (each harboring more than one mutation) of five mice of the same transgenic line 7.5. Mutations of the three most prominent hotspots, Glu70(III), Leu71(I) and Val73(III), are not included to avoid the bias caused by a high incidence of mutation in these positions. These three hotspots have no influence on the analysis of A and T mutations, as they are of G and C residues. However, mutations in the three positions alone account to 41 and 40% of all the C and G mutations respectively. In the same 258 bp gpt region, no mutational bias between A and T is found in the previous transgene L¢-Vgpt* nor between G and C (Table 2). This conclusion is based on the separate analysis of the two mutation data sets of L¢-Vgpt* (14,19), using the same criteria as that for analyzing L¢-Vgpt*-3 (i.e. including only those clones harboring more than one mutation and excluding the three most prominent hotspots). For one such data set, #1, 163 mutations pooled from 33 clones are used for the strand bias analysis. The same three most prominent hotspots, which are identical to those of the present transgene L¢-Vgpt*-3, are excluded. No mutational bias between A and T (\( P = 0.9571 \)) nor G and C (\( P = 0.5108 \)) is found. In another data set, #2, again no bias is found between A and T (\( P = 0.1285 \)) nor G and C (\( P = 0.4300 \)) using 167 mutations pooled from 39 clones. In this set of data, the third and the weakest of the three major hotspots is Ala107(II). The change in hotspot has no influence on the bias analysis between A and T, as Ala107(II) is a C. Taken together, these data show that the same gpt sequence in different transgenes can undergo two different modes of hypermutation, one with strand bias and another without.

To rule out the possibility that the strand bias in hypermutation of L¢-Vgpt*-3 is peculiar to the transgenic line 7.5, an independent transgenic mouse, #103 of line 8.14, was also studied. Mutation frequency of the transgene of this mouse is ~3 times lower than that of mice of line 7.5 (Table 1). Despite such a difference, a significant bias in mutation for A over T (\( P = 0.0102 \)) is found (Table 2). No bias is detected between G and C (\( P = 0.4300 \)). This analysis is based on 90 mutations collected from 59 clones, with exclusion of the three most prominent hotspots, Glu70(III), Leu71(I) and Val73(III).

### Table 2. Comparison of strand bias in hypermutation between L¢-Vgpt*-3 and L¢-Vgpt*

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Percentage of mutationsa (no.)b</th>
<th>Line 7.5</th>
<th>Line 18.4</th>
<th>Data set #1</th>
<th>Data set #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>L¢-Vgpt*-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>44% (92)</td>
<td>0.0002c</td>
<td>0.0102</td>
<td>28% (43)</td>
<td>0.9571</td>
</tr>
<tr>
<td>T</td>
<td>23% (56)</td>
<td>0.3168</td>
<td>0.4300</td>
<td>26% (50)</td>
<td>0.5108</td>
</tr>
<tr>
<td>G</td>
<td>18% (43)</td>
<td>0.0002c</td>
<td>0.0102</td>
<td>20% (34)</td>
<td>0.5108</td>
</tr>
<tr>
<td>C</td>
<td>14% (29)</td>
<td>0.3168</td>
<td>0.4300</td>
<td>24% (36)</td>
<td>0.5108</td>
</tr>
<tr>
<td>L¢-Vgpt*</td>
<td></td>
<td></td>
<td></td>
<td>32% (51)</td>
<td>0.1285</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( ^a \)The percentage of mutations for each base is corrected for the base composition of the region analyzed. The three most prominent hotspots are excluded from analysis [i.e. Glu70(III), Leu71(I) and Val73(III)] for L¢-Vgpt*-3 and for data set #1 of L¢-Vgpt* (19); Glu70(III), Leu71(I) and Ala107(II) for data set #2 L¢-Vgpt* (14).

\( ^b \)Number of mutations excluding the three most prominent hotspots. Only those clones harboring more than one mutation and excluding the three most prominent hotspots were analyzed in the 258 bp sequence segment of gpt from nucleotide 148 to 405.

\( ^c \)Probability value obtained by \( \chi^2 \) analysis to indicate whether there is a statistically significant difference in mutations between the complementary nucleotides on the coding strand. \( P \) values of statistical significance are in italics.

Mutations features associated with strand bias

As the mutation frequency of L¢-Vgpt*-3 shown by line 7.5 is very similar to that of L¢-Vgpt* in line LxNG (14) (20.5 versus 20.3/3kb), the distribution of nucleotide substitutions in the two transgenes that underlie their difference in strand bias manifestation can be analyzed quantitatively. Since RGYW and WRCY motifs stand out as the most preferred consensus sequences for mutation, substitutions within and outside the motifs were analysed as separate groups. As shown in Table 3, there are similarities and differences in the distribution of the nucleotide substitutions between the two transgenes. The similarity is that the substitutions for all the four nucleotides occurred preferentially to the RGYW and WRCY motifs. After correcting for the numbers of the nucleotides within and outside the motifs, we note that about two-thirds and three-quarters of all the mutations occurred to those motifs in L¢-Vgpt* and L¢-Vgpt*-3 respectively. The interesting difference is that the mutations are quite evenly distributed among the four nucleotides both within and outside the motifs for L¢-Vgpt*, but not for L¢-Vgpt*-3, which shows strand bias in hypermutation. For L¢-Vgpt*-3, there is a notable increase of A mutation in RGYW and WRCY motifs. Remarkable
Strand bias in hypermutation

Table 3. Comparison of the distribution of nucleotide substitutions in the same 258 bp gpt region between \( \lambda^c \)-V\( gpt^* \) and \( \lambda^c \)-V\( gpt^*\)-3\(^b\)

<table>
<thead>
<tr>
<th>Transgene</th>
<th>Corrected distribution of nucleotide substitutions(^b)</th>
<th>Neither RGYW nor WRCY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RGYW and WRCY</td>
<td>A</td>
</tr>
<tr>
<td>( \lambda^c )-V( gpt^* )</td>
<td></td>
<td>17%</td>
</tr>
<tr>
<td>( \lambda^c )-V( gpt^*)-3 (line 7.5)</td>
<td></td>
<td>27%</td>
</tr>
<tr>
<td>( P (\chi^2) )(^c)</td>
<td>0.0617(^<em>) 0.2836 0.975 0.3348 0.1274 0.0535(^</em>) 0.4539 0.0081</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Data for \( \lambda^c \)-V\( gpt\)-3 are drawn from line 7.5 which demonstrated an average mutation frequency that is almost identical to that of the \( \lambda^c \)-V\( gpt^* \) transgenic line \( \lambda^c \)-\( gpt \) \( (20.5/kb \) for the former compared with 20.3\( /kb \) for the latter). Data for \( \lambda^c \)-V\( gpt^* \) are pooled from the two data sets, \#1 and \#2. The three most prominent hotspots are excluded from analysis.

\(^b\)Corrected for the numbers of the nucleotides present within and outside of the RGYW and WRCY motifs (the numbers are shown in Methods).

\(^c\)\( P \) values for the difference in distribution of nucleotide substitutions between the transgenes. Value of statistical significance (<0.05) are in italics. Asterisked values are values near the threshold of statistical significance.

decreases, however, are noted outside the motifs. These occurred to the mutation of T and highly significantly so to the mutation of C. Although statistically significant bias between G and C mutations is not found in either \( \lambda^c \)-V\( gpt^*-3 \) or \( \lambda^c \)-V\( gpt^* \) (Table 2), this fall in the frequency of C mutations outside the motifs in \( \lambda^c \)-V\( gpt^*-3 \) suggests the presence of a strand-dependent component in C mutation in \( \lambda^c \)-V\( gpt^*-3 \). With the mutation database of \( \lambda^c \)-V\( gpt^*-3 \) increasing in size, we expect to see a bias between the G and C mutations reaching to a statistically significant level.

Discussion

By comparing in detail the hypermutation characteristics of a non-Ig sequence, \( gpt \), incorporated in two different ways in the LVJ region of a \( \lambda^c \)-OX1/J\( \beta \)-5 Ig transgene, one with retention of the whole set of LVJ sequence (\( \lambda^c \)-V\( gpt^*-3 \)) and another only 16% of it (\( \lambda^c \)-V\( gpt^* \)), we have obtained the following two major findings. First, the same non-Ig sequence substrate can undergo two different modes of hypermutation. The one retaining the full set of LVJ sequence hypermutates with manifestation of strand bias and the other one without. Second, manifestation of strand bias is accompanied by an increase of A mutation in the RGYW and WRCY motifs, and a decrease of T and C mutations outside the motifs. For the identical substrate manifesting no strand bias, the mutations are more or less evenly distributed among the four nucleotides both within and outside the motifs.

The first finding suggests that the LVJ region must contain an element(s) that confers strand bias to hypermutation. This element(s) is not present in \( \lambda^c \)-V\( gpt^* \) where only 16% of the LVJ region comprising the 3' half of FR3, CDR3 and J\( \beta \)-5 remains in the transgene. In one other previously published transgene, \( \lambda^c \)-V\( \beta \)G, where a human \( \beta \)-globin sequence has substituted almost the entire V sequence, leaving only a few nucleotides in the FR1, and the J\( \beta \)-5 region, hypermutation of the \( \beta \)-globin region again displayed no strand bias (14,19). On the other hand, hypermutation of ned (neomycin resistance gene), placed within the LVJ region with retention of the 5' region of the Ig sequence from the initiation codon to almost the middle of FR2 and J\( \beta \)-5, was found to display strand bias also manifested by the bias for A mutation over T (14,19). Although these previously published findings on gpt (\( \lambda^c \)-V\( gpt^* \)), \( \beta \)-globin (\( \lambda^c \)-V\( \beta \)G) and ned (\( \lambda^c \)-V\( neo^*\)\( \Delta[(XS)] \)) could already be used to imply that part of the LVJ sequence may have influence on whether the hypermutation proceeds with a strand-dependent component, it could not be certain whether it is the different DNA sequence of the substrates that was responsible for the difference shown in the strand bias of hypermutation (14). By using an identical gpt sequence and showing that reconstitution of the LVJ region resumes the strand bias in hypermutation, we can rule out the latter interpretation and establish the importance of the LVJ sequence in conferring strand bias to hypermutation. With this important point established, we can now proceed to identify which part of the LVJ is responsible for the strand bias. By comparing the LVJ sequence content of the two transgenes positive in hypermutational strand bias, \( \lambda^c \)-V\( gpt^*-3 \) and \( \lambda^c \)-V\( neo^*\)\( \Delta[(XS)] \), with the other two negative transgenes, \( \lambda^c \)-V\( gpt^* \) and \( \lambda^c \)-V\( \beta \)G (all are derived from the same \( \lambda^c \)-OX1/J\( \beta \)-5 parent transgene) (Fig. 3), it can be concluded that the stretch of LVJ sequence comprising the FR1, CDR1 and the 5' half of FR2 probably contains the signal element(s) for conferring strand bias. We are aware of a report concluding that the increase in A mutations compared with T is a result of antigen selection, rather than reflecting strand bias as an inherent feature of hypermutation (16,30). This conclusion was based on analyzing the proportion of mutations represented by each base in 37 non-productively rearranged human \( \beta \)-globin genes; no increase in mutation of A compared with that of T was found. However, the analysis has not taken into account the base compositions of the sequences under investigation as we have done here and in other reports supporting strand bias (14,19,20–22). It should also be noted that, like the non-productively rearranged Ig genes, both the \( \lambda^c \)-V\( gpt^*-3 \) and \( \lambda^c \)-V\( gpt^* \) are non-functional transgenes, the mutations of which are therefore not subject to selection.

The second finding that the strand bias is associated with a marked increase in A mutation in the RGYW and WRCY motifs, and decrease in T and C mutations outside the motifs suggests that a component of the hypermutation...
Transgenes positive in hypermutational strand bias

---

Transgenes negative in hypermutational strand bias

---

Fig. 3. Structural comparison between the \( L_\kappa \)-derived transgenes with and without manifestation of strand bias in hypermutation. Striped boxes represent the non-Ig substrate sequences; all other boxes correspond to the various coding sequences of \( L_\kappa \). Thick lines represent the leader intron. The \( L_\kappa \) sequence segment present in and unique to the transgenes showing hypermutational strand bias is the CDR1 and its flanking sequences: a large part of FR1 at the 5’ and almost half of FR2 at the 3’.

Transgenes positive in hypermutational strand bias

---

Transgenes negative in hypermutational strand bias

---

Acknowledgments

We are indebted to Drs César Milstein and Cristina Rada (MRC Laboratory of Molecular Biology in Cambridge, UK) for providing us the raw databases of the published hypermutation of \( L_\kappa\)-Vgpt*.

Abbreviations

LVJ leader variable joining (sequence region coding for the leader and rearranged variable region of Ig light chain)

PNA peanut agglutinin
References


37 Kim, N., Bozek, G., Lo, J. C. and Storb, U. 1999. Different mismatch repair deficiencies all have the same effects on somatic


