View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Immunity, Vol. 16, 123–134, January, 2002, Copyright ©2002 by Cell Press

Effects of Sequence and Structure on the Hypermutability of Immunoglobulin Genes

Nancy Michael,[1] Terence E. Martin,[1] Dan Nicolae,[4] Nayun Kim,[2] Kris Padjen,[1] Ping Zhan,[4] Hanh Nguyen,[3] Carl Pinkert,[5] and Ursula Storb[1,3,6]

[1] Department of Molecular Genetics and Cell Biology
[2] Department of Biochemistry and Molecular Biology
[3] Committee on Immunology and
[4] Department of Statistics
University of Chicago
Chicago, Illinois 60637
[5] Department of Pathology and Laboratory Medicine
University of Rochester
Rochester, New York 14642

Summary

Somatic hypermutation (SHM) is investigated in related immunoglobulin transgenes that differ in a short artificial sequence designed to vary the content of hotspot motifs and the potential to form RNA or DNA secondary structures. Mutability depends on hotspots, not secondary structure. Hotspot motifs predict about 50% of the mutations; the rest are in neutral and coldspots. Clusters of mutations and the sequential addition of mutations found in cell pedigrees suggest epigenetic attributes of SHM. Sometime in SHM, an essential factor seems to become limiting. Particular error-prone DNA polymerases appear to create mutations in hotspots on the top and bottom DNA strands throughout the target and the SHM process. One transgene is superhypermutable in all regions, suggesting the presence of a *cis*-element that enhances SHM.

Introduction

Somatic hypermutation (SHM) of immunoglobulin (Ig) genes is a major mechanism for creating a vast repertoire of antibody specificities. The SHM process is turned on in B lymphocytes in developing germinal centers after interaction of the B cell with foreign antigen and T cells (Kelsoe, 1995). While Ig genes are the major targets for SHM, the protooncogene BCL6 is also mutable by this process (Pasqualucci et al., 1998; Peng et al., 1999; Shen et al., 1998). It appears that the specificity for targeting is determined by Ig enhancers and presumably their homolog in the BCL6 gene (reviewed in Storb, 1996). The Ig promoter can be replaced by other promoters without loss of mutability (Betz et al., 1994; Tumas-Brundage and Manser, 1997). The mutations start about 200 nucleotides (nt) from the transcription start site and extend for 1–2 kb. Thus, only the 5′ portion of the Ig genes, within and flanking the variable-joining (VJ) region, is mutated.

The primary sequence of the SHM target does not have to be an Ig sequence. Artificial substrates or bacterial sequences in the context of an Ig gene can be mutated as highly as the Ig VJ region (reviewed in Storb, 1996). However, the primary sequence of the target is important, as certain short sequences (di-, tri-, or tetranucleotides) were found to be either hotspots or coldspots for SHM (Bachl et al., 1997; Betz et al., 1993; Foster et al., 1999; Rogozin and Kolchanov, 1992; Shapiro et al., 1999; Smith et al., 1996). The 64 nt triplets have been ordered from the hottest to the coldest based on a survey of a large database of somatic mutations in Ig genes (Shapiro et al., 1999). The G of the hottest hotspot, AGC, has a mutability index (MI) of 3.36, meaning that it is 3.36 times more likely to be mutated than expected if targeting were insensitive to the sequence. The G in the context of the first nt of the coldspot, GAC, has a mutability index of 0.09, meaning that it is mutated about ten times less often than expected (Shapiro et al., 1999). The data were from untranslated regions; thus, they were unlikely to be selected. Instead, the hot/coldspots most likely reflect an intrinsic event of the mutation process.

Based on the finding that the duplication of the Ig promoter upstream of the Ig constant region leads to a wave of mutations over the constant region, which is normally not mutated, we have proposed a model in which SHM is linked to transcription (Peters and Storb, 1996; Storb et al., 1998a). In this model, through interaction with Ig enhancers, a mutator factor (MuF) becomes associated with the transcription complex near the promoter. As the Ig gene is transcribed, the MuF travels with the transcribing complex. We hypothesize that the MuF is deposited on the DNA within 1–2 kb 3′ of the promoter and cannot reassociate with an elongating transcription complex. The MuF is (or recruits) an endonuclease, which causes a single- or double-strand cut in the target DNA. Mutations are introduced into the DNA during repair of the lesion.

We had previously created a κ transgene, PEPS, with an insertion in the V region of 96 nt containing alternating EcoRV and PvuII restriction sites (EPS = Eco-Pvu sequence) (Klotz et al., 1998; Storb et al., 1998a). The artificial insert was found to be more highly mutable than the flanking VJ regions. The insert has a high proportion of SHM hotspots, but there are other similarly hot regions in the transgene. Also, as expected from the repeating EcoRV and PvuII restriction sites, the insert is predicted to form highly stable folded structures from single-stranded DNA or RNA transcripts, which, we postulated, might block transcription. We hypothesized that it was the combination of the hotspot concentration and the secondary structure that leads to the high mutability of the EPS insert (Storb et al., 1998a).

To test the relative roles of hot/coldspots and of potential secondary structure properties of the SHM target sequences, we created several novel Ig κ transgenes with insertions of short, defined sequences in the V region. The results clearly demonstrate the major importance of hotspots. The accumulation of a large dataset of mutations with one of the transgenes also provided unexpected new insights into the relationship of mutations in sequential cell generations, the organization of

[6] Correspondence: stor@midway.uchicago.edu

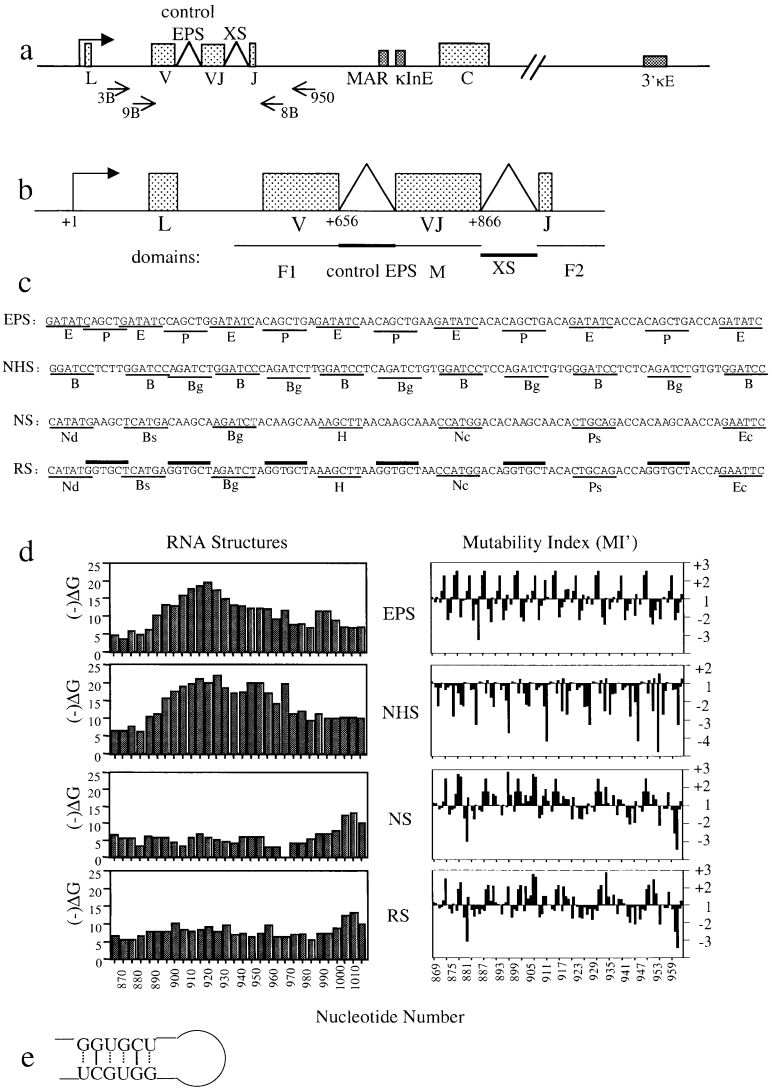

Figure 1. The Transgenes

(a) Diagram of the transgenes. Primers used in cloning and SSCP are indicated by straight arrows. The triangles represent the inserted control EPS and XS sequences. L, leader; V, variable; J, joining; C, constant gene segments; MAR, matrix attachment region; and κInE and 3′κE, κ gene enhancers.

(b) The mutable 5′ end of the transgenes. The numbers indicate the number of nucleotides from the start of transcription. Domains of transgenes analyzed for mutations are F1, control EPS, M, XS, and F2.

(c) Sequences of the XS inserts in the four different transgenes. E, EcoRV; P, PvuII; B, BamHI; Bg, BglII; Nd, NdeI; Bs, BspH1; H, HindIII; Nc, NcoI; Ps, PstI; Ec, EcoRI. In RS, the hexamer that allows RNA stem formation is overlined.

(d) Graphs of predicted RNA secondary structure and distribution of hot or cold nt in the XS mutation substrates. The predicted $-\Delta G$ values of RNA secondary structure and the mutability index (MI′) were determined as described in Experimental Procedures. The MIs are theoretical estimates of mutability based on previous studies (Shapiro et al., 1999).

(e) Base pairs of a stem formed between two of the RS-specific hexamers GGUGCU. Dotted lines, weak base pairing between G and U; solid lines, stronger base pairing between G and C.

the target genes into mutable domains, and the possible existence of a "mutation enhancer" element.

Results

Transgenic Mice

The four transgenes analyzed in this study are based on the PEPS transgene previously described (Klotz et al., 1998; Storb et al., 1998a). They contain the promoter, and the intron- and 3′ enhancers of the κ167 gene (Figure 1a). The transgenes are identical in all domains, except for the 96 test nt (XS) inserted 866–961 nt from the start of transcription (Figure 1b). As an internal control, the original EPS sequence is present at its original position 656–751 nt from the start of transcription. The XS sequences were designed to test the effect of sequence hotspots and predicted RNA secondary structure in targeting SHM. In the 2xEPS transgene the XS sequence is identical to EPS. As described before (Storb et al. 1998a), the EPS insert has a high level of predicted secondary structure and a relatively equal distribution of strong hotspots and coldspots (Figure 1d). Two other XS inserts were designed as relatively pure models of predicted secondary structure without hotspots (NHS = no hotspots) and of hotspots without secondary structure (NS = no stable secondary structure) (Figures 1c and 1d). Finally, the RS insert was designed to favor RNA, but not DNA, stem-loops. This discrimination is achieved by six repeats of GGTGCT so that the direct repeats of two such elements can form stem-loop structures, with a preponderance of GU base pairs (Figure 1e). The XS insert of the RS transgene has a moderate level of secondary structure stability and a similar distribution of hot and coldspots to the EPS insert (Figure 1d). The RS and NS transgenes lack the PvuII sites that account for 12 of the hot nt in the EPS transgene (6 × AGC, hot nt underlined). Instead, they have other hot nt that are part of known SHM hotspots, such as RGYW, WRCY, and WA (see below).

One transgenic line was analyzed in detail for the 2xEPS transgene (one copy) and the NS and NHS transgenes (two copies each). Two transgenic lines were ana-

Table 1. The Number of Expected and Observed Mutations in the Five Domains, by Length of Domain

Transgene	EPS		NHS		NS		RS	
Domain	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs
F1	25.24	19	73.02	73	38.16	21	231.67	183
Control EPS	12.82	22	37.09	52	19.38	29	117.67	158
M	15.22	16	43.27	63	23.01	20	137.28	141
XS	12.82	18	38.25	23	19.38	38	121.35	170
F2	17.89	9	51.38	32	27.05	19	163.03	119
Total	84		243		127		771	

Number of mutations in the five domains of the four transgenes. The expected number of mutations is calculated as (Ld/Lt)N, where Ld is the length of the domain, Lt is the length of the total sequence analyzed, and N is the total number of mutations observed.

lyzed for the RS transgene (G057 has two copies and G058 has ten copies). In this paper, only the data from the two-copy RS transgene are shown.

Mutability Depends on the Proportion of Hotspots, Not on Secondary Structure

The mutability of the transgenes was determined in PNA^{hi} B cells from Peyer's patches of 4- to 7-month old mice. DNA was amplified by PCR, cloned, screened for mutations by SSCP, and sequenced to identify mutations. In the analysis, we divided the sequenced 624 nt into five domains: F1, the region 5′ of the original EPS insert (187 nt); control EPS, the original EPS (96 nt); M, the region between control EPS and XS (114 nt); XS, the new insert (96 nt); and F2, the region 3′ of XS (131 nt) (Figure 1b) and counted the mutations by domain. The four transgenes are identical, except for the 96 nt of XS. As shown in Table 1, mutations do not occur at a constant frequency across these five domains, suggesting that the primary sequence or the DNA/RNA structure influences mutability.

In Figure 2a data are normalized to F1. The control transgene, 2xEPS, shows that the EPS insert in the XS position, like the control EPS insert, is still more mutable than its flanks. Thus, the strategy using transgenes with a second experimental insert (XS) provides a useful test of the relevance of hotspots and secondary structure.

In the NS transgene, the XS-NS insert, which has very little predicted secondary structure, is even more highly mutable than the control EPS insert, suggesting that secondary structure of the DNA or the RNA sequence is not required for SHM. That the NS sequence is more highly mutated than the EPS sequence is likely due to the scarcity of coldspots in NS (Figures 1d and 2a).

In the NHS transgene the XS-NHS insert, which has no hotspots but can form RNA or DNA stem-loop structures, is mutated less than the control EPS (Figures 1d and 2a; XS domain of the NHS transgene). Thus, the secondary structure of the DNA or RNA alone is not sufficient to cause a high level of SHM. Interestingly, the few mutations in the NHS insert are seen in neutral and coldspots at the frequencies expected from their MI (Figure 2c), suggesting that the intrinsic MIs are an indication of the primary sequence motifs that target SHM and that coldspots are indeed avoided. There is no evidence that the coldspots or the predicted secondary structure negatively affect SHM in other domains.

In the RS transgene, the RS insert which has a similar content of hotspots as its EPS insert is mutated to a similar degree, even though the secondary structure stability of the RS insert is considerably lower (Figures 1d and 2a).

Thus, the frequency of mutations of the XS domain compared with the control EPS domain is directly related to the relative frequencies of hotspots in the two domains: the XS/control EPS mutation ratios are NS $>$ RS $>$ EPS $>>$ NHS, and the hotspot ratios are in a similar relationship NS $>$ RS $=$ EPS $>>$ NHS. The order of the stability of secondary structures of the XS domain is rather different, NHS $>$ EPS $>>$ RS $>$ NS, suggesting that in the targeted sequence, the major determinant of mutability is the DNA sequence not its secondary structure or that of the encoded RNA.

The RS Transgene Is Particularly Hypermutable

As shown in the normalized data (Figure 2a), the mutability of the XS-RS domain compared with the control EPS domain in the RS transgene follows the general rule that the concentration of hotspots determines the mutation frequency. However, when the absolute frequencies of mutations of the transgenes are compared, the RS transgene has five to ten times more mutations than the other transgenes (Figure 2b). This was true for two independent RS transgenic lines, G057 (two copies) and G058 (ten copies), and for many individual mice. The mutation enhancement is therefore due to a *cis* element in the transgene and not to effects of the insertion site or a general hypermutable phenotype. All five domains of the RS transgene are more highly mutated. Since the only difference between the RS and the other transgenes is the 96 nt XS-RS insert, this insert must exert a generalized enhancement effect. This question is under further investigation.

Pedigrees of Related Mutations

Since the RS transgene was highly mutable, we collected a large number of mutated sequences. Many of the mutated RS transgene sequences could be grouped into seven pedigrees that appear to have arisen from a common ancestor (Figures 3a and 3b). In pedigree A, for example, the common ancestral sequence must have had one deletion of 9 nucleotides and 15 point mutations, since these changes are present in all other members of this pedigree. The ancestral sequence was not found for this pedigree, suggesting that the cell, or a daughter cell containing this sequence, may not have

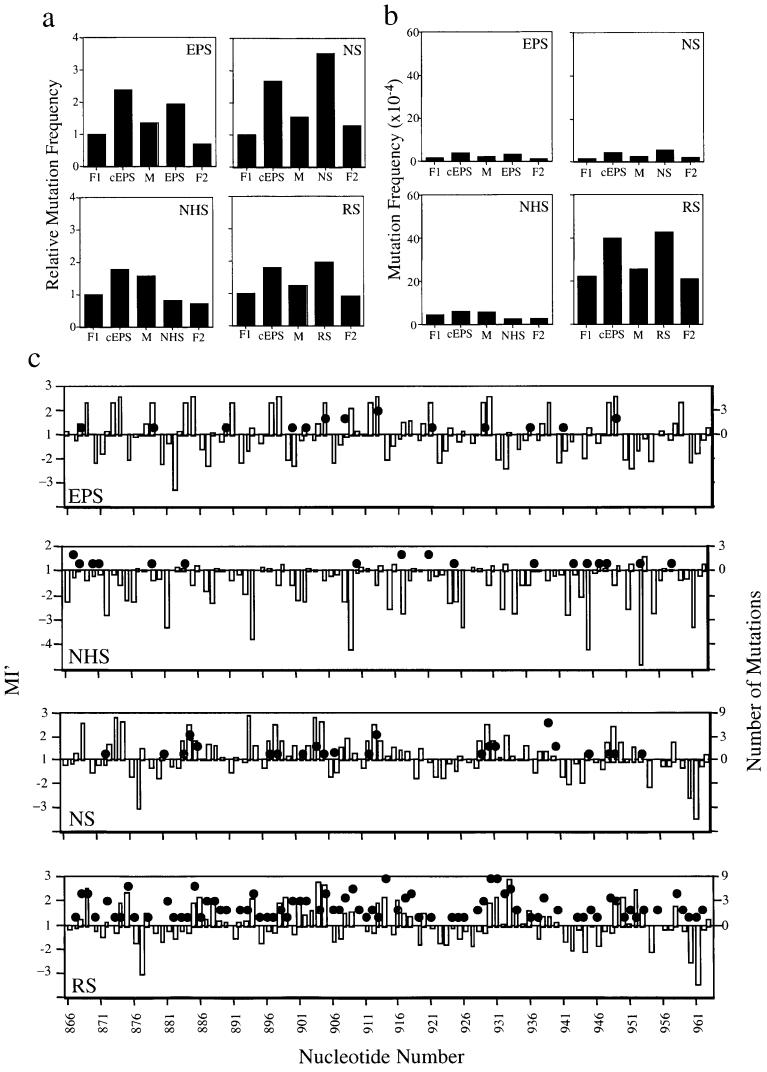

Figure 2. Mutability of the Four Transgenes

(a and b) For each transgene, mutations were identified by sequencing starting at nt 469 in the V region. The analyzed region was divided into five domains (see Figure 1b). (a) The mutation frequency was calculated as the number of mutations/(number of clones screened) × (length of domain) and normalized to F1 = 1. (b) Mutation frequencies, not normalized. (c) Graphs of mutability index (MI′, shown as bars) and number of mutations observed (shown as closed circles) for the XS domain of the four transgenes. Note that the right scale is contracted by a factor of 3 for the number of mutations between 3 and 9.

been preserved without additional mutations. The founder gave rise to at least seven different derivatives with different additional mutations. One of these derivatives (this sequence was not recovered) accumulated seven additional mutations and gave rise to two additional derivatives. These patterns indicate that the mutation process must have continued for at least seven cell generations (see Discussion). The pedigrees are unlikely due to chimeric molecules formed during PCR amplification because both the original and the new mutations are spread throughout the sequence (Figure 3b).

The Mutations Occur in Clusters

Considering individual sequences from the four transgenes, it appears that mutations occur in irregular clusters (example of RS clones from one mouse, Figure 4a). To determine the significance of this observation, we first checked the distribution of single mutations. In all four transgenes, single point mutations were found in hot-, neutral, and coldspots at the expected frequencies (Table 2), with a random pattern in the position of a single mutation along the transgenes (Figures 4a and 4b).

We then compared the spacing between mutations in sequences that have more than one mutation. The data were sorted by distances of 1–22 nt between two consecutive mutations and compared with 1000 simulations for each distance. For the RS, NHS, and 2xEPS transgenes, mutation doublets with the distance of 1 (we name them "1+1," meaning mutations at a given position "1" and its neighbor) and the distance of two nucleotides ("1+2," meaning mutations at a given position "1" and another 2 nt distant) were found more often than expected. For the NS transgene, the data were not statistically significant, although the trend was the same. Figure 4c shows the results for the RS transgene. We found 37 1+1 events and 43 1+2 events (distances 1 and 2, respectively, in Figure 4c). In 1000 simulations, 37 or more 1+1 events occurred only 14 times and 43 1+2 events did not occur a single time. Thus, the narrow clustering is significant ($p = .028$ and $p < .002$, respectively), suggesting that the mutations at two closely spaced positions are due to either simultaneous or repeated targeting at the same location. It should be noted that these spacings occurred anywhere throughout the sequenced region of the transgenes and were different for different DNA clones (e.g., Figures 3b and 4a).

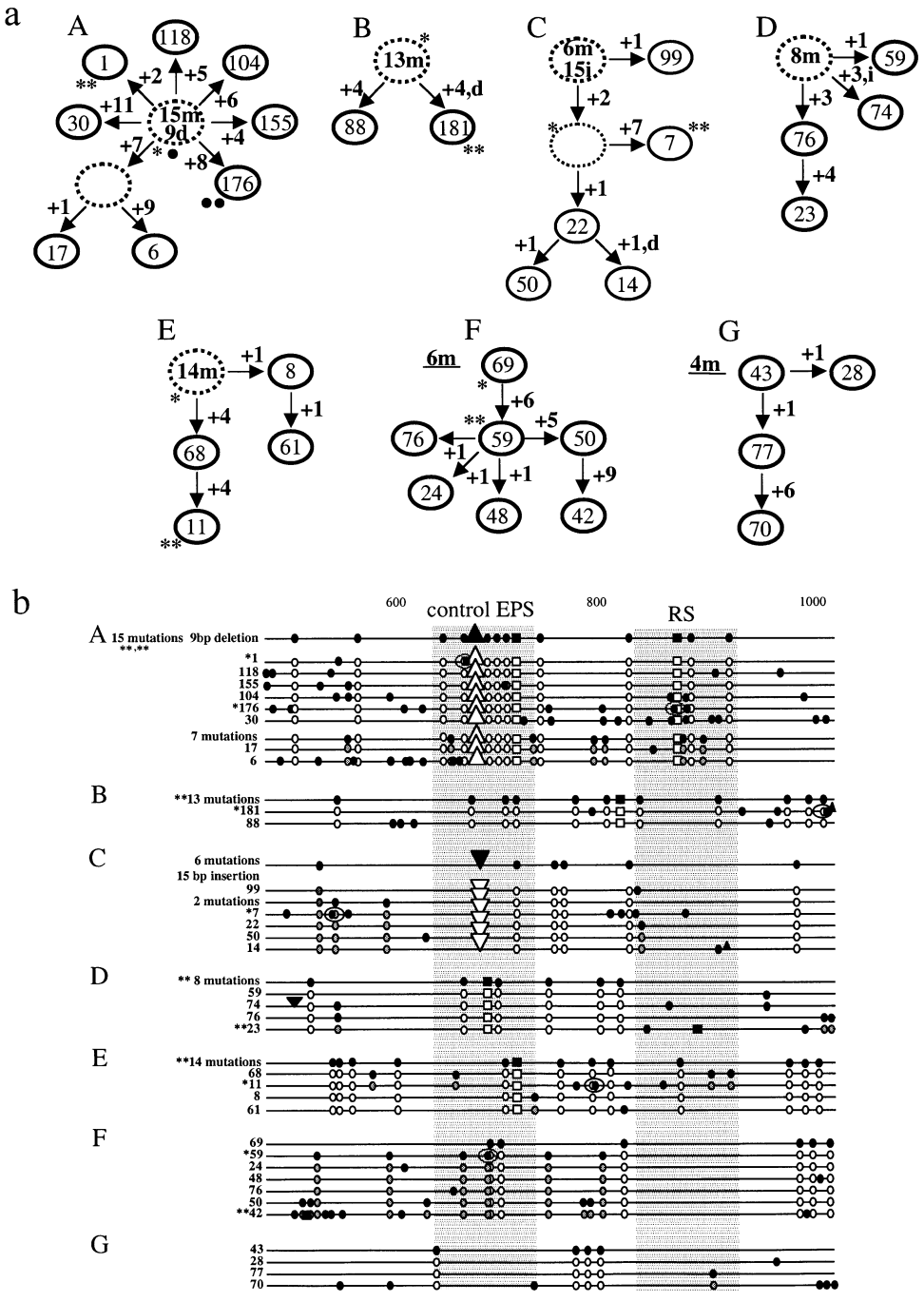

Figure 3. Pedigrees of Sequence Lineages from RS Mice

(a) Genealogy of sequences. The solid circles indicate observed mutated sequences. The stippled circles indicate precursor sequences (cells) that were not found but have mutations that are shared by all subsequent sequence (cell) generations. Numbers within the stippled circles indicate the numbers of point mutations (m) or deleted (d) or inserted (i) nucleotides. Sequences (cells) that were found are shown in solid circles with the clone number inside; for "parental" clones F and G, the number of mutations is noted outside and underlined. For daughter clones the + numbers indicate the number of mutations between recovered subsequent generations. The *(•) and ** (••) symbols indicate the precursor-progeny sequence relationships, respectively, in cases where a new mutation arose adjacent to an older one in a later generation. (b) Location of mutations in pedigree DNA clones of the RS transgene. Individual DNA clones are designated at left. In each pedigree, the top line represents a putative (A–E) or found (F and G) precursor sequence to all sequences in the given pedigree. Additional putative intermediate precursor clone sequences are indicated by the additional mutations, without a clone designation. * and elliptical circles, new mutation next to previous mutation; ** and squares, two adjacent new mutations. Triangles on base, deletions; triangles on tip, insertions. Closed circles and squares, new mutations. Open circles and squares, mutations present in the founder sequence. Shaded circles, mutations introduced in an intermediate generation.

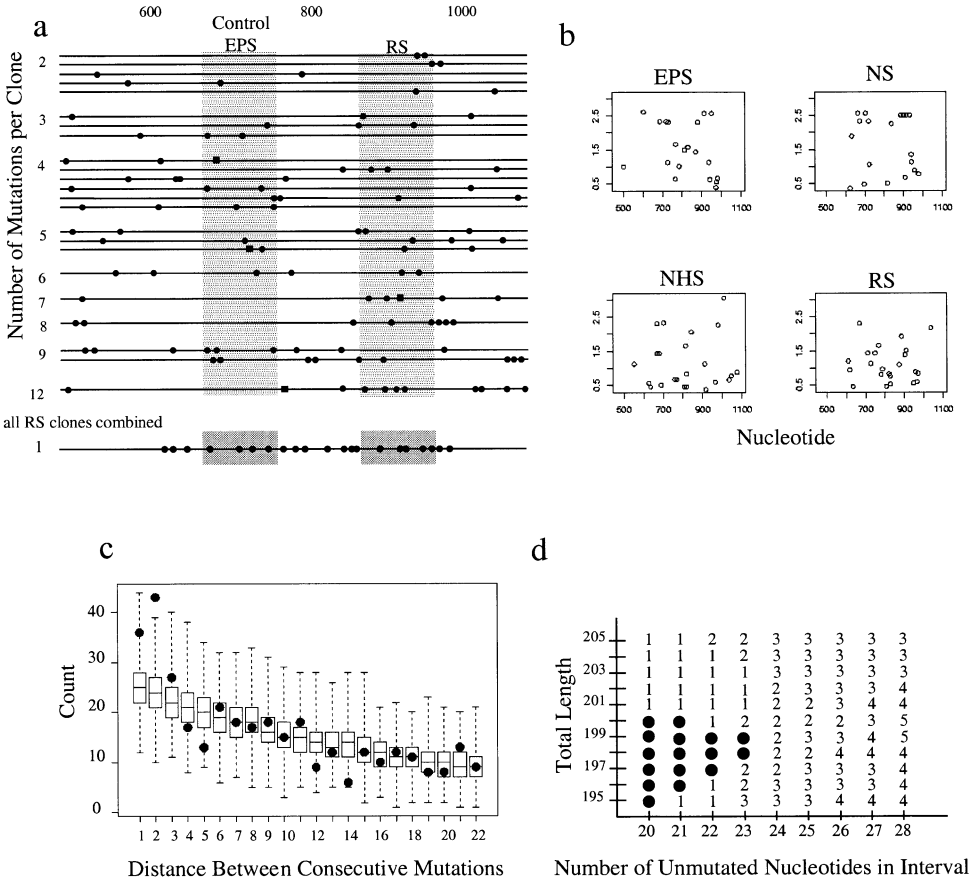

Figure 4. Clusters of Mutations

(a) RS clones from one mouse with two or more mutations. The data are grouped by numbers of mutations (2–12) in the 624 nt sequenced. Each line is one DNA clone, except the bottom line, which shows the location of each mutation in all clones with a single mutation. Each circle is one point mutation, and the squares are two adjacent mutations.

(b) Position of mutations and MI of the unmutated nt for sequences with only a single mutation of the four transgenes.

(c) Consecutive mutations in RS clones with at least two mutations (the individual clones had 2–50 mutations). The closed circles show the number of times two consecutive mutations with the indicated distances were observed. For each distance, 1000 simulations were performed, and the number of times two consecutive mutations with the given distance are expected was calculated. The boxplots describe the simulated data sets. The central line in each box is the median; the top and bottom lines are 75% and 25% of the simulations, respectively.

(d) Long-range cluster analysis for the RS gene. The numbers in the figure represent the numbers of clones that do not fit the corresponding pattern. The closed circles represent the patterns that fit all 96 clones analyzed. Each pattern is described by two numbers: the length of the interval of unmutated nt (*x* axis) and the total length of the interval, mutated and unmutated (*y* axis).

Given the high frequency of 1+1 and 1+2 spacings of mutations, we considered that the unexpected close spacing might indicate a very short patch of error-prone DNA repair. This assumes that when two mutations were near each other, they arose in the same mutation event. It was possible to investigate this in the RS pedigrees. Interestingly, in six cases (indicated in Figure 3a and summarized in Table 3, "1+1") a new mutation appeared adjacent to an older one in a later generation. Furthermore, in six cases, a new mutation occurred within two nucleotides of an original one in a later generation (Table 3, "1+2"). In seven additional sequences, two mutations were found next to each other in the same "generation" (Figure 3b), and one 1+2 appeared together in the same generation (data not shown). Since it is unlikely that cells from every generation were recovered, some or all of the 1+1 doublets and the 1+2 doublet appearing together in our sample may also have arisen in different cell generations (see Discussion). The later targeting at the very same location was not due to the creation of a hotspot by the first mutation (Table 3). In nine cases, the mutability index for the nt that was mutated later was essentially unchanged by the first mutation; in one case the MI went from neutral to moderately hot; in one case the MI dropped from hot to moderately hot; and in two other cases the MI dropped from moderately hot to neutral or cold. The 1+1 (Figure 3b) and 1+2 (data not shown) mutations occurred in all five domains and almost all were at different positions in different clones. Curiously, the mutation occurring first was generally a transition (82%) and the second a transversion mutation (75%) (data not shown).

We further determined in the RS transgenes whether 1+1 or 1+2 doublets of mutations were due to an abundance of hotspots at the mutated positions (Table 4). For the 1+1 doublets, the average MI was just barely

Table 2. Single Point Mutations in Hot-, Neutral, and Coldspots

EPS	Hot	Neutral	Cold	Total
Exp	8.4	6.2	4.4	
Obs	11	4	4	19
P		.721		
NHS				
Exp	5.9	6	4.1	
Obs	6	4	6	16
P		.733		
NS				
Exp	11.1	7.7	5.3	
Obs	17	3	4	24
P		.063		
RS				
Exp	10.9	7.6	5.5	
Obs	9	8	7	24
P		.894		

Distribution in hot-, neutral, and coldspots of single point mutations in the four transgenes.

Table 4. Mutation Index of Mutation Doublets and Adjacent Nucleotides

MI	5′Up Nt	5′	Middle Nt	3′	3′Down Nt
1+1 Doublets (32)					
Mean	1.09	1.34		1.39	1.16
SD	+/−0.59	+/−0.66		+/−0.68	+/−0.72
Median	0.85	1.21		1.15	0.95
Minimum	0.21	0.37		0.40	0.38
Maximum	2.33	2.57		2.70	2.89
1+2 Doublets (22)					
Mean	1.20	1.22	1.41	1.68	0.82
SD	+/−0.62	+/−0.48	+/−0.58	+/−0.74	+/−0.54
Median	0.93	1.01	1.43	1.95	0.62
Minimum	0.55	0.51	0.59	0.47	0.33
Maximum	3.06	2.15	2.57	2.89	2.27

Mutation Index (MI) of mutation doublets and the adjacent nt. Includes only the nonpedigree data of the RS transgene where the order in which the two mutations arose is not known. 5′, 3′, the mutated nt; 5′Up Nt, upstream nt; 3′Down Nt, downstream nt. In ( ), number of total doublets.

in the hotspot range, while the nt upstream and downstream of the doublet had a neutral MI. However, the values varied enormously (very high SD), the median MI was neutral, and the lowest and highest MIs were similar for all four positions. For the 1+2 pairs, the unmutated nt between the two mutated ones had a higher average MI than the 5′ mutation, and again, all the data points were extremely variable.

Combining all the findings with the 1+1 and 1+2 mutations, it is therefore likely that most paired mutations do not occur at the hottest hotspots, nor do they require that both nucleotides be hotspots, or that the first mutation creates a hotspot for the second, and, in cases where the genealogy is known, very often do not arise at the same time. They are thus most often not an indication of a single mutational event in which multiple errors are introduced in a short patch of DNA. Rather, most of the 1+1 and 1+2 doublets (at least in the pedigrees) seem to relate to potentially epigenetic conditions that favor clustering of mutations. If there is an effect by the first mutation on the targeting of the second, it may be that a factor involved in the first mutation remains bound at the site or that the first mutation causes a change in overall structure or context of the gene. Alternatively, there may be a preset structure context that favors the targeting of both the first and the second mutation.

Table 3. Change in MI for the Second Mutation by the First Mutation in 1+1 and 1+2 Doublets in RS Pedigrees

Family	Doublet	MI Change		
A	1+1	0.86 > 1.54	+	neutral > moderately hot
		1.38 > 1.51	=	moderately hot
A	1+2	1.92 > 1.66	−	hot > moderately hot
		0.80 > 1.28	=	neutral
		1.58 > 1.66	=	moderately hot
		2.57 > 2.45	=	very hot
		2.09 > 1.99	=	very hot
B	1+1	1.37 > 1.34	=	moderately hot
C	1+1	1.51 > 1.58	=	moderately hot
E	1+1	1.43 > 0.54	−	moderately hot > cold
F	1+1	1.43 > 0.96	−	moderately hot > neutral
F	1+2	0.4 > 0.45	=	cold

Changes in Mutation Index (MI) for the second mutation by the first mutation in 1+1 and 1+2 doublets in RS transgene pedigrees. MI change, the first number indicates the MI of the unmutated nt of the second mutation in the transgene; the second number indicates the MI of the unmutated nt of the second mutation taking into account the new nt created in its vicinity by the first mutation.

Besides the tight doublets, we also observed a larger pattern of clustering of mutations (Figure 4a). In the search for a possible determining structure, we analyzed the size and spacing of the larger clusters of mutations. With the assumption that nucleosome spacing may impose the pattern, we determined whether the data fit certain arrangements of mutated regions and unmutated gaps. A total of 154 different patterns were tested containing between 175 and 185 nt of mutated region and 20–33 nt gaps for a total periodicity of 195–218 nt. For the EPS, NHS, and NS transgenes that have moderate numbers of mutations per sequence, the data fit a large number of patterns. However, for the RS transgene, the data fit only the following patterns (Figure 4d): six patterns of a gap of 20 unmutated nt with 175–180 nt mutated, for a total unit length of 195–200 nt; five patterns of a gap of 21 unmutated nt with 175–179 nt mutated, for a total of 196–200 nt; three patterns of a gap of 22 nt with 175–177 nt mutated, for a total of 197–200 nt; and two patterns of 23 unmutated nt with 175–177 nt mutated, for a total of 198–199 nt. Thus, all the RS data fit 16 patterns totaling unit lengths of mutated region plus gaps of 195–200 nt. We then created 100 simulated data sets using as a base the same number of DNA clones, the same number of mutations, and the same distribution of the mutability indices as in the experimental data. In none of the simulations was such a large number of patterns found that fit all clones. Furthermore, the 16 compatible experimental patterns were grouped into a tight total length of 195–200 nt (Figure

4d). We believe that the relatively large number of patterns that fit the experimental data compared to the simulations is an indication of some higher order structure that organizes the mutation target (see Discussion).

Discussion

Hypermutability Depends Mainly on Predicted Hotspots

The analysis of the four transgenes with insertions of 96 nt containing various proportions of hotspots and of potential secondary structure of the DNA or RNA transcript clearly shows that the proportion of known SHM hotspots is the major determinant of the mutability of a region by the SHM process. There was no significant effect of potential secondary structure on the mutability of the respective insert or its flanking regions in the transgenes. Thus, if as postulated (Storb et al., 1998a), pausing by the RNA polymerase causes deposition of a mutator factor, the secondary structure predicted by RNA folding potential is not an indicator for pausing sites. It is still possible that secondary structures serve as signals for pausing or cleavage by a putative endonuclease, but some minimal stability may be enough for such signals and increasing stability may have little impact. However, the large-scale clustering of mutations may indicate RNA polymerase pausing near nucleosomes (see below).

Which DNA Strand Is Mutated, and Does the Mutation Mechanism Change in the Course of SHM?

Recently, several error-prone DNA polymerases have been implicated in SHM (Diaz et al., 2001; Frank et al., 2001; Poltoratsky et al., 2001; Rogozin et al., 2001; Storb, 2001; Zan et al., 2001; Zeng et al., 2001; Spencer et al., 1999). Different polymerases are most likely responsible for introducing errors in the SHM hotspots RGYW versus WA (Rogozin et al., 2001). By comparison with the known error patterns of these polymerases in vitro, it appears that mutations in the sequence WA or TW are introduced for A in the top or bottom strands, respectively (W is A or T, and the underlined A or T is the mutable nucleotide). Mutations in the sequence RGYW or WRCY are introduced for G on the top or bottom strands, respectively (R is purine, Y is pyrimidine). One of these hotspots is present in each of the hotspots defined by the Wysocki analysis. That analysis includes an ordering of all nt triplets, including those that do not conform to these four hotspot motifs. Interestingly, the XS-NHS insert in our NHS transgene, which contains none of these hotspot motifs is mainly mutated at the warmer nt (Shapiro et al., 1999); only 24% of the mutations are in really cold spots (Figure 2b). This suggests that there may be additional rules and perhaps different DNA polymerases that discriminate between the cold and warmspots.

It has been shown that double-strand breaks and perhaps single-strand breaks can be detected in the mutable V region, but not the C region of Ig genes undergoing SHM (Bross et al., 2000; Kong and Maizels, 2001; Papavasiliou and Schatz, 2000; Sale and Neuberger, 1998). It is not known if there is a unique single- or double-strand break required for SHM. It is also not clear which kind of break initially occurs. The repair may be error prone for some types of breaks but not others. To begin to address these questions, it is useful to gain independent evidence for strand targeting. Other studies have suggested strand-specific (reviewed in Storb et al. 1998b), as well as strand-independent targeting of SHM (Foster et al. 1999; Spencer et al. 1999). We have analyzed our transgene data, assuming that mutations found in the underlined nucleotides in the sequences RGYW, WRCY, WA, and TW indicate mutations in the top, bottom, top, and bottom strands, respectively. About half (47%–61%) of all mutations occurred in these hotspot motifs (Table 5). The percentage of hotspot mutations of the total mutations was very similar as determined by the RGWY, WRCY, WA, and TW analysis (51%, 47%, 61%, and 60% for EPS, NHS, NS, and RS, respectively) or by using the Wysocki compilations with a cut off of hotspots at $\geq$1.30 (54%, 41%, 61%, and 53%, respectively).

Considering the mutations in the five domains in the RS gene, 45%–71% of the hotspot mutations occurred on the top strand (Table 5). The ratios of top and bottom strand targeting vary somewhat in the different transgenes (Table 5), but overall, either strand appears to have a similar chance to be targeted. In all four transgenes, the F2 domain is mainly targeted on the bottom strand. This domain (962–1092 nt from the promoter) is located in a region where the overall mutation frequency is on the decline (Both et al., 1990; Gearhart and Levy, 1991; Lebecque and Gearhart, 1990; Rada et al., 1994). One possible explanation for the inverted T/A ratio would be that only a specially modified transcription complex can target this downstream region, resulting in a different strand distribution.

In most cases, there is a high A/T ratio (except, curiously, the F2 domain) but a variable G/C ratio of mutations in hotspot motifs (Table 5, underlined numbers). This suggests that DNA polymerases involved in WA hotspot mutations may introduce errors preferably in the top strand, while enzymes that introduce errors in RGYW may introduce errors variably in either strand.

There is no consistent change in targeting one strand or the other as SHM progresses through multiple generations (Table 6). However, given the patterns of differential top and bottom strand targeting, it is likely that in SHM conditions vary so that in slightly more than half the time, an error is introduced in the top strand and in the rest of the time, the bottom strand, presumably depending on the type of the break and the factors involved in its repair.

The Mutations Are Clustered

Clusters of mutations have been observed in previous studies (Diaz et al., 1998; Gearhart and Bogenhagen, 1983). The analysis of our transgenes suggests that there is an organization imposed on the DNA sequence that determines where the mutations fall. Clusters of mutations are at different positions in different DNA clones and therefore independent of the primary sequence of the transgene.

We consider it possible that the large-scale clustering of mutations may reflect a higher-order structure that is unrelated to the tight clustering in the 1+1 and 1+2 mutations. The large clusters in various regions and in

Table 5. Distribution of Mutations in Hotspot Motifs RGYW, WRCY, WA, and TW

	Transgene	EPS (51)[a]	NHS (47)	NS (61)	RS[b] (60)
Domain	NT[c]	Hot Mutations (%)[d]	Hot Mutations (%)	Hot Mutations (%)	Hot Mutations (%)
All	G	13	39	43	25
	C	50	31	27	33
	A	23	17	17	26
	T	14	13	12	16
	% Top[e]	36	56	60	51
F1	G	20	54	39	45
	C	63	25	31	18
	A	12	15	12	26
	T	5	6	17	11
	% Top	32	69	51	71
Control EPS	G	31	41	36	17
	C	31	27	26	31
	A	26	18	21	31
	T	13	14	18	22
	% Top	57	59	57	48
M	G	32	33	14	19
	C	0	<33	35	36
	A	68	30	45	26
	T	0	15	6	19
	% Top	100	>63	59	45
XS	G	10	0	61	22
	C	48	0	19	34
	A	25	0	12	27
	T	17	0	8	16
	% Top	35	NA	73	49
F2	G	0	10	34	27
	C	0	46	<39	33
	A	39	8	7	19
	T	61	37	20	20
	% Top	39	18	>41	46

Distribution of mutations in hotspot motifs RGWY, WRCY, WA, and TW in the five domains of the four transgenes. Underlined, more frequently mutated NT in G plus C or A plus T; e.g. EPS All, C > G and A > T.

[a] In ( ) % of total mutations that are in hotspots.

[b] Excludes RS pedigrees.

[c] G, C, A, and T mutated in RGWY, WRCY, WA, and TW, respectively.

[d] Mutated hotspots corrected for the content of the particular hotspot in the particular domain.

[e] % Top, % of total hotspot mutations that are in RGWY or WA and can thus be attributed to having arisen in the top strand (Top).

different positions in different sequences are likely due to an epigenetic structure. There are several ways in which this cell-specific structure may arise, for example, by individual patterns of DNA replication, or transcription, and/or chromatin organization.

Although elegant evidence has been presented which suggests that SHM generally takes place outside of global replication (Bertocci et al., 1998), it is possible that sometimes an SHM complex encounters a replication complex. Different cell clones may have established different origins of replication in the vicinity of the rearranged mutable Ig sequence, and a given clone may retain the same origin over time. However, the relationship of multiple clusters of mutations in one sequence to a given origin of replication is difficult to envisage.

Transcription is required for SHM, but the mechanism

Table 6. Top- versus Bottom-Strand Mutations in Multiple Generations in RS Pedigrees A–G

Generation	Parental		≥2nd		≥3rd		≥4th	
Strand								
T	24	36%	23	36%	9	18%	6	35%
B	16	24%	9	14%	15	30%	3	18%
N	26	39%	32	50%	26	52%	8	47%

Summary of top- and bottom-strand mutations (as derived from targeted hotspot motifs) in multiple generations in the pedigrees of the RS transgene. "Parent," earliest sequence recovered; ≥2nd to ≥4th generations, sequences with sequentially added mutations whose precursors are found (pedigrees F and G) or assumed (pedigrees A–E) (see Figures 3a and 3b). T, top strand mutated (G or A in RGWY or WA); B, bottom strand mutated (C or T in WRCY or TW); N, mutations not in RGWY, WRCY, WA, or TW hotspots.

is unknown. If, as postulated, the pausing of RNA polymerase were a potential mechanism for releasing a mutator factor, the clustered mutations would require that in the same cell and its progeny, transcription complexes progress in a specific pattern determined perhaps by the dosage of certain factors loaded at the promoter or at the start of "productive" elongation (Marshall and Price, 1992). It has been shown in a totally artificial system that different molecules of identical RNA polymerases progress at different rates and pause at different sites (Davenport et al., 2000). It is possible that different cells produce different levels of certain transcription factors, both activating and inhibitory, and that these various balances influence the SHM process, for example, where the postulated mutator factor is deposited along the transcribed gene.

We found that the mutation pattern is compatible with a stretch of 175–180 mutated nt with unmutated short gaps of 20–23 nt (Figure 4d), which may suggest an effect of nucleosome spacing. It is possible that, as the mutation apparatus interacts with nucleosomes, the 20 or so nucleotides associated with the center of the nucleosome where the DNA is most tightly wrapped around the core histones are not accessible. In analogy, in an in vitro system of V(D)J recombination, certain recombination signal sequences, when located in the center of the nucleosome dyad, are not accessible to the V(D)J recombination machinery (Kwon et al., 1998). In this scenario, for SHM the nucleosome pattern would have to be established differently in different precursor cells but then remain essentially unchanged during consecutive cell generations. In general, in pools of cells, fixed nucleosome patterns are seen within promoter and enhancer regions, but random patterns are seen within the transcribed gene (McPherson et al., 1993). The latter can be attributed to the transient displacement of nucleosomes by the RNA polymerase transcribing the gene (Studitsky et al., 1994) and/or the sampling of a large number of cells, each of which may have different nucleosome spacings. It is plausible that the displaced nucleosomes are reorganized at basically the same positions after passage of the transcription complex in a given cell lineage. Alternatively, nucleosomes may shift to slightly different positions after each passage. In that case, the gap of an unmutated sequence would become shorter with the mutation process. This is compatible with the narrowing of the gaps observed in later generations of the pedigrees (Figure 3b). In any case, nucleosomes are likely to cause pol II pausing, especially if certain factors are limiting (LeRoy et al., 1998), and thus may aid the local release of a mutator factor.

The reason for the 1+1 and 1+2 pair targeting may be different than for the large scale clusters. In this context, it is especially interesting that mutations of 1+1 or 1+2 pairs that are known to occur in different cell generations can occur either on the top or bottom strand in either of the generations (data not shown). Furthermore, in doublets 1+1 or 1+2 with unknown order of appearance, very often one hotspot is mutated on the top strand and the other on the bottom strand (data not shown). These closely spaced mutations therefore could not have arisen in the same DNA repair track. Thus, while it is likely that some doublet mutations arise in the same repair track, as some of the DNA polymerases implicated in SHM have an extremely high error rate, the majority of 1+1 and, expecially, 1+2 mutations appear to occur in different mutation events and even different cell generations. This may suggest that factors involved in creating or repairing the original break may remain bound to the DNA at a site where they originally acted and facilitate another mutation event in another round of transcription or subsequent cell generation.

Clonal Relationships of Mutations

An interesting question raised by the SHM pedigrees is the implied relationship between a mutating precursor cell and its clonal derivatives. The number of generations one can deduce from the number of differently mutated offspring depends on what happens in the cell division of the founder or intermediary cell; does each daughter cell get a mutated chromosome, or does one get a mutated and the other an unmutated chromosome? The latter seems to occur at a reasonable frequency, since we do find actual precursor sequences (in pedigree A = clone 6; D = 76, C = 22, E = 8 and 68; F = 59 and 50; and G = 43 and 77; Figures 3a and 3b). Recovering a precursor sequence supports the notions that (a) when in SHM an error arises in one DNA strand, it does not get corrected on the other strand, or (b) SHM occurs after the S phase. On the other hand, when the ancestral sequence is not found, besides technical reasons, this may indicate that (a) the mutations arose early in the cell lineage and both daughters of a mutating cell continued to mutate or (b) the errors arose in G1 and a mismatch repair mechanism "corrected" the wild-type sequence to fix the mutation before the S phase. Future experiments will have to determine which of these possibilities is correct.

In five pedigrees, a precursor cell that shares mutations with all members of the pedigree was not found (Figure 3a). However, in pedigrees F and G, a precursor sequence was recovered. Furthermore, in six of the seven pedigrees, we recovered intermediate sequences that gave rise to sequences with additional mutations. In our sample of one mouse from which the pedigrees F and G were derived, we sequenced every mutant DNA clone that was obtained. In these pedigrees, no "ghost" intermediates had to be introduced. In fact, the "original clone," #43 in pedigree G, was obtained six times, and the intermediate clone, #59 in pedigree F, was obtained seven times (Figures 3a and 3b). These findings suggest that some daughters of B cells engaged in somatic mutation divide without further mutation, whereas their siblings continue to mutate. It is possible that an essential factor for SHM (perhaps the postulated mutator factor) becomes limiting and packaged in different quantities in different daughter cells.

Experimental Procedures

Mutation Substrates in Transgenic Mice

The Vk167/PEPS sequence (Klotz et al., 1998) was variously modified to make four new mutation substrates (2xEPS, NHS, NS, and RS) by the insertion of the 96 bp DNA fragment XS (where XS = EPS, NHS, NS, or RS) in the BspHI site, 106 nt 3′ of the original EPS insert (Figure 1a). The DNA fragments were made by annealing single-strand oligonucleotides, synthesized to contain the desired sequence (Figure 1b). Transgenic mice on a (B6/SJL)F2 background

were produced by standard techniques. Founder animals were identified by PCR of mouse tail DNA. Transgene copy number was estimated by Phosphorimager quantitation of Southern blots of mouse tail DNA digested with BamHI and probed with ^{32}P-labeled Cκ (Klotz et al., 1998). The founders and subsequent generations of transgenic animals were bred to C57BL/6 mice and maintained on food that was not irradiated. Animal procedures conformed to University of Chicago and University of Alabama approved protocols.

DNA Clones of Transgenic Sequences from B Cells

B cells were prepared from Peyer's patches removed from transgenic mice aged 4–7 months. Cells were stained with PE-anti mouse CD45R/B220 (Pharmingen, San Diego, CA), FITC-antiGL7 (Pharmingen), and FITC-PNA (Sigma-Aldrich, St Louis, MO) and sorted by FACStar^{PLUS} (Becton Dickinson) in the Immunology Applications Core Facility at the University of Chicago. The B220^{+}, PNA/GL7^{hi} and B220^{+}, and PNA/GL7^{lo} populations of cells were collected. DNA, prepared by lysing the cells in 0.25% Tween-20 and 50 μg/ml Proteinase K, was used as the template in PCR, with primers 3B/950 (Figure 1a) and Pfu polymerase for 25–27 cycles. The PCR product was gel purified and cloned into the vector provided in the Zero Blunt TOPO PCR Cloning Kit (Invitrogen). Individual clones were picked, streaked, and grown on a 6 × 6 grid plate.

Identification of Somatic Mutations

Mutations in a transgene were detected by SSCP (Orita et al., 1989). Plasmid DNA of transgene clones was amplified with primers 8B/9B (Figure 1a) in a PCR that included Taq DNA polymerase and α^{32}P dCTP. The PCR product was digested with HincII into fragments of 248 and 160 bp, diluted 30-fold in 50% formamide, denatured by heating at 98°C for 15 min, and resolved by electrophoresis on a 6% polyacrylamide-10% glycerol gel. Mutant clones were detected by the differential migration of the single-strand fragments and were sequenced by the University of Chicago Cancer Research Center DNA Sequencing Facility. Individual mutations were identified with Sequencher. The data were tabulated, stored, and manipulated in Excel.

Measures of Mutability and RNA Secondary Structure

The MI indicates the relative likelihood that each nucleotide will or will not be mutated by somatic hypermutation and can be used to predict the mutation frequency for each nucleotide in its sequence context. Shapiro et al. (1999) determined the values from an analysis of unselected mutations by calculating the ratio of observed/expected mutations in the context of nucleotide triplets. If mutations occurred at random, the ratio would be 1. A MI $>$ 1 describes a preferred site of mutation. Likewise, a MI $<$ 1 describes a site that seems to be avoided by the mutation process. In the current study, the triplet values were adapted to predict the MI of the central nucleotide in a pentanucleotide context (G. Shapiro, personal communication) (Shen et al., 2000). When the MIs are displayed graphically, they are recalculated (MI′): if MI $\geq$ 1 MI′ = MI, and if MI $<$ 1 MI′ = −(1/MI) so that hotspots are graphed above and coldspots below the *x* axis at value 1.

RNA secondary structure is predicted by mfold (Zucker, 1989), modified by Martin and Wang to reiteratively scan a sequence (Storb et al., 1998a). This program displays the predicted structures formed by a designated length of sequence and its $-\Delta G$ (relative Gibb's free energy), advances a given number of nucleotides, and repeats the predictive process. We used a window of 50 nt repeated at 5 nucleotide intervals. The greatest $-\Delta G$ for each interval is assigned to the nucleotide at the 3′ end of that interval and used as a measure of the likelihood that the sequences in that interval can form a stable RNA structure.

Statistical Analysis

For Table 1, the expected number of mutations was calculated for each domain by assuming that each nt is equally likely to be mutated with the total number of mutations equal to the observed number of mutations. Thus, Expected = (Ld/Lt)N, where N is the total number of mutations observed, Ld is the length of the domain, and Lt is the total number of nt analyzed (624).

For Table 2, the data were analyzed based on the MI. Nt with an MI $\geq$ 1.3 were grouped as "hotspots," those with an MI between 0.78 and 1.29 as "neutral," and those with an MI $\leq$ 0.77 as "coldspots." The expected number of mutations in "hotspots," "neutral," and "coldspots" were calculated for each domain based on the MI of each nt and compared to the observed number of mutations. The significance of the difference between the expected and observed number of mutations was assessed using a likelihood ratio test on a multinomial distribution (Rice, 1995).

For Figure 4c, we investigated the spacing between mutations that occurred within 1–22 nt of each other. To evaluate the significance of our findings, we compared the observed counts to counts from simulated data sets. For each transgene, we simulated 1000 data sets that have the same number of clones, the same number of mutations within each clone, and the same MI as the observed transgene data. p values were calculated as twice the proportion of simulations in which we obtain a number of consecutive mutations with a given distance that is more extreme than in the observed data set (e.g., distance 1 had 38 observed cases, 14/1000 simulations were $>$38, thus p = 2 × 14/1000 = .028).

For Figure 4d, we investigated the significance of the number of patterns that fit the observed data. A pattern consists of two intervals of nt and can be described by two nonnegative integers, one that defines the length of the interval that is free of mutations and one that defines the total number of nt in the pattern. The first mutation in a clone is the starting point for the pattern configuration. As controls, we simulated 1000 data sets and compared the number of clones that fit certain patterns to those of the experimental data.

Acknowledgments

We are grateful to L. Wysocki and G. Shapiro for unpublished mutability indices and A. Longacre for stimulating discussions concerning breaks and DNA polymerases and for critical reading of the paper. This study was supported by NIH grants AI47380 and GM38649 (to U.S.). N.K. and K.P. were supported by NIH Training Grants GM07183 and Medical Scientist NRS Award 5 T32 GM07281, respectively. Transgenic mice were generated by the NICHD Transgenic Mouse Development Facility operated by the University of Alabama at Birmingham (contract NO1-HD-5-3229).

Received September 24, 2001; revised November 8, 2001.

References

Bachl, J., Steinberg, C., and Wabl, M. (1997). Critical test of hot spot motifs for immunoglobulin hypermutation. Eur. J. Immunol. *27*, 3398–3403.

Bertocci, B., Quint, L., Delbos, F., Garcia, C., Reynaud, C.-A., and Weill, J.-C. (1998). Probing Ig gene hypermutation with microsatellites suggests a nonreplicative short patch DNA synthesis process. Immunity *9*, 257–265.

Betz, A.G., Neuberger, M.S., and Milstein, C. (1993). Discriminating intrinsic and antigen-selected mutational hotspots in immunoglobulin V genes. Immunol. Today *14*, 405–411.

Betz, A., Milstein, C., Gonzalez-Fernandes, R., Pannell, R., Larson, T., and Neuberger, M. (1994). Elements regulating somatic hypermutation of an immunoglobulin K gene: critical role for the intron enhancer/matrix attachment region. Cell *77*, 239–248.

Both, G.W., Taylor, L., Pollard, J.W., Steele, E.J. (1990). Distribution of mutations around rearranged heavy-chain antibody variable-region genes. Mol. Cell. Biol. *10*, 5187–5196.

Bross, L., Fukita, Y., McBlane, F., Demolliere, C., Rajewsky, K., and Jacobs, H. (2000). DNA double-strand breaks in immunoglobulin genes undergoing somatic hypermutation. Immunity *13*, 589–597.

Davenport, R., Wuite, G., Landick, R., and Bustamante, C. (2000). Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase. Science *287*, 2497–2500.

Diaz, M., Greenberg, A.S., and Flajnik, M.F. (1998). Somatic hypermutation of the new antigen receptor gene (NAR) in the nurse shark does not generate the repertoire: possible role in antigen-driven reactions in the absence of germinal centers. Proc. Natl. Acad. Sci. USA *95*, 14343–14348.

Diaz, M., Verkoczy, L.K., Flajnik, M.F., and Klinman, N. (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.

Foster, S., Doerner, T., and Lipsky, P. (1999). Somatic hypermutation of VkJk rearrangements: targeting of RGYW motifs on both DNA strands and preferential selection of mutated codons within RGYW motifs. Eur. J. Immunol. *29*, 4011–4021.

Frank, E.G., Tissier, A., McDonald, J.P., Rapic-Otrin, V., Zeng, X., Gearhart, P.J., and Woodgate, R. (2001). Altered nucleotide misinsertion fidelity associated with poliota-dependent replication at the end of a DNA template. EMBO J. *20*, 2914–2922.

Gearhart, P.J., and Bogenhagen, D.F. (1983). Clusters of point mutations are found exclusively around rearranged antibody variable genes. Proc. Natl. Acad. Sci. USA *80*, 3439–3443.

Gearhart, P.J., and Levy, N.S. (1991). Kinetics and molecular model for somatic mutation in immunoglobulin variable genes. In Somatic Hypermutation in V-regions, E.J. Steele, ed. (Boca Raton, Fl: CRC Press), pp. 29–39.

Kelsoe, G. (1995). The germinal center reaction. Immunol. Today *16*, 324–326.

Klotz, E., Hackett, J.J., and Storb, U. (1998). Somatic hypermutation of an artificial test substrate within an Ig kappa transgene. J. Immunol. *161*, 782–790.

Kong, Q., and Maizels, N. (2001). DNA breaks in hypermutating immunoglobulin genes: evidence for a break-and-repair pathway of somatic hypermutation. Genetics *158*, 369–378.

Kwon, J., Imbalzano, A., Matthews, A., and Oettinger, M. (1998). Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1. Mol. Cell *2*, 829–839.

Lebecque, S., and Gearhart, P. (1990). Boundaries of somatic mutation in rearranged immunoglobulin genes: 5′ boundary is near the promoter, and 3′ boundary is ~1 kb from V(D)J gene. J. Exp. Med. *172*, 1717–1727.

LeRoy, G., Orphanides, G., Lane, W.S., and Reinberg, D. (1998). Requirement of RSF and FACT for transcription of chromatin templates in vitro. Science *282*, 1900–1904.

Marshall, N., and Price, D. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. Mol. Cell. Biol. *12*, 2078–2090.

McPherson, C.E., Shim, E.Y., Friedman, D.S., and Zaret, K.S. (1993). An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. Cell *75*, 387–398.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989). Detection of polymorhisms of human DNA by gel electrophoresis as single-strand conformation polymorphism. Proc. Natl. Acad. Sci. USA *86*, 2766–2770.

Papavasiliou, F.N., and Schatz, D.G. (2000). Cell-cycle-regulated DNA double-stranded breaks in somatic hypermutation of immunoglobulin genes. Nature *408*, 216–221.

Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., Chaganti, R., Klein, U., Kueppers, R., Rajewsky, K., et al. (1998). BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutationa cting outside Ig loci. Proc. Natl. Acad. Sci. USA *95*, 11816–11821.

Peng, H.-Z., Du, M., Koulis, A., Aiello, A., Dogan, A., Pan, L., and Isaacson, P. (1999). Nonimmunoglobulin gene hypermutation in germinal center B cells. Blood *93*, 2167–2172.

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

Poltoratsky, V., Woo, C., Tippin, B., Martin, A., Goodman, M., and Scharff, M.D. (2001). Expression of error-prone polymerases in BL2 cells activated for Ig somatic hypermutation. Proc. Natl. Acad. Sci. USA *98*, 7976–7981.

Rada, C., Gonzalez-Fernandez, A., Jarvis, J.M., and Milstein, C. (1994). The 5′ boundary of somatic hypermutation in a Vk gene is in the leader intron. Eur. J. Immunol. *24*, 1453–1457.

Rice, J. (1995). Mathematical Statistics and Data Analysis, 2nd Edition (Belmont, CA: Duxbury Press).

Rogozin, I., and Kolchanov, N. (1992). Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighboring base sequences on mutagenesis. Biochim. Biophys. Acta *1171*, 11–18.

Rogozin, I., Pavlov, Y., Bebenek, K., Matsuda, T., and Kunkel, T. (2001). Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. Nat. Immunol. *2*, 530–536.

Sale, J., and Neuberger, M. (1998). TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. Immunity *9*, 859–869.

Shapiro, G., Aviszus, K., Ikle, D., and Wysocki, L. (1999). Predicting regional mutability in antibody V genes based solely on di- and trinucleotide sequence composition. J. Immunol. *163*, 259–268.

Shen, H., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998). The BCL6 gene in normal memory B cells is mutated by the process of somatic hypermutation of immunoglobulin genes. Science *280*, 1750–1752.

Shen, H., Michael, N., Kim, N., and Storb, U. (2000). The TATA binding protein, c-myc, and survivin genes are not somatically hypermutated, while Ig and BCL-6 genes are hypermutated in human memory B cells. Int. Immunol. *12*, 1085–1093.

Smith, D., Creadon, G., Jena, P., Portanova, J., Kotzin, B., and Wysocki, L. (1996). Di- and trinucleotide target preferences of somatic mutagenesis in normal and autoreactive B cells. J. Immunol. *156*, 2642–2652.

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

Storb, U. (1996). The molecular basis of somatic hypermutation of immunoglobulin genes. Curr. Opin. Immunol. *8*, 206–214.

Storb, U. (2001). DNA polymerases in immunity: profiting from errors. Nat. Immunol. *2*, 484–485.

Storb, U., Klotz, E., Hackett, J., Kage, K., Bozek, G., and Martin, T.E. (1998a). A hypermutable insert in an immunoglobulin transgene contains hotspots of somatic mutation and sequences predicting highly stable structures in the RNA transcript. J. Exp. Med. *188*, 689–698.

Storb, U., Peters, A., Klotz, E., Kim, N., Shen, H.M., Kage, K., and Rogerson, B. (1998b). Somatic hypermutation of immunoglobulin genes is linked to transcription. Curr. Top. Microbiol. Immunol. *229*, 11–19.

Studitsky, V., Clark, D., and Felsenfeld, G. (1994). A histone octamer can step around a transcribing polymerase without leaving the template. Cell *76*, 371–382.

Tumas-Brundage, K., and Manser, T. (1997). The transcriptional promoter regulates hypermutation of the antibody heavy chain locus. J. Exp. Med. *185*, 239–250.

Zan, H., Komori, A., Li, Z., Cerutti, A., Schaffer, A., Flajnik, M.F., Diaz, M., and Casali, P. (2001). The translesion DNA polymerase zeta plays a major role in Ig and bcl-6 somatic hypermutation. Immunity *14*, 643–653.

Zeng, X., Winter, D., Kasmer, C., Kraemer, K., Lehmann, A., and Gearhart, P. (2001). DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. Nat. Immunol. *2*, 537–541.

Zucker, M. (1989). On finding all suboptimal folding of an RNA molecule. Science *244*, 48–52.