Diffuse large B-cell lymphoma is the most common lymphoid malignancy in adults. It is a heterogeneous disease with variability in outcome. Genomic instability of a subset of proto-oncogenes, including c-MYC, BCL6, RhoH, PAX5, and PIM1, can contribute to initial tumor development and has been correlated with poor prognosis and aggressive tumor growth. Lymphomas in which these proto-oncogenes are unstable derive from germinal center B cells that express activation-induced deaminase (AID), the B-cell–specific factor that deaminates DNA to initiate immunoglobulin gene diversification. Proto-oncogene instability is evident as both aberrant hypermutation and translocation, paralleling programmed instability which diversifies the immunoglobulin loci. We have asked if genomic sequence correlates with instability in AID-positive B-cell lymphomas. We show that instability does not correlate with enrichment of the WRC sequence motif that is the consensus for deamination by AID. Instability does correlate with G-richness, evident as multiple runs of the base guanine on the nontemplate DNA strand. Extending previous analysis of c-MYC, we show experimentally that transcription of BCL6 and RhoH induces formation of structures, G-loops, which contain single-stranded regions targeted by AID. We further show that G-richness does not characterize translocation breakpoints in AID-negative B- and T-cell malignancies. These results identify G-richness as one feature of genomic structure that can contribute to genomic instability in AID-positive B-cell malignancies. [Cancer Res 2007;67(6):2586–94]
The ability of AID to initiate genomic instability has stimulated considerable interest in understanding how this factor is targeted to specific genes. Transcription of the target gene is a prerequisite for deamination by AID, reflecting preferential deamination of single-stranded rather than double-stranded DNA substrates (30). Nonetheless, transcription is not sufficient for deamination, and many genes that are transcribed in activated B cells are not targets for AID. Strikingly, in DLBCL, aberrant hypermutation is restricted to a subset of proto-oncogenes, including BCL6, c-MYC, PAX5, PIM1, and RhoH1; systematic analysis revealed no evidence of aberrant hypermutation in about a dozen other representative genes expressed at comparable levels in germinal center B cells, including a-MYC, CD10/Calla, NBS1, and L-Plastin (14). This suggests that identification of the features that distinguish unstable proto-oncogenes from other transcribed genes could provide insights into mechanisms that target AID to specific genes in activated B cells.

Two features of genomic sequence and structure could in principle contribute to AID attack on a transcribed gene: (a) abundance of the WRC sequence motif that is the preferential target for AID and (b) G-richness. Transcription of G-rich regions, like the S regions, results in formation of unusual DNA structures. These contain a stable RNA/DNA hybrid on the template strand and single-stranded regions interspersed with G4 DNA on the G-rich strand (31). G4 DNA is a four-stranded DNA structure in which interactions between strands are stabilized by G-quartets, planar arrays of four guanines (32, 33). The structure in transcribed S regions can readily be observed by electron microscopy as characteristic G-loops, which are several hundred base pairs in length (31). Systematic analysis has shown that G-loops readily form upon in vitro or intracellular transcription and form within a variety of G-rich repeats, including immunoglobulin S regions, the mammalian telomeric repeat TTAGGG, multimerized synthetic G-rich sequences, and a G-rich region of the human c-MYC proto-oncogene (31, 34). G-loop formation occurs only if the nontemplate strand is G-rich. G-rich regions are the sites of immunoglobulin class switch recombination, and this is the physiologic orientation for transcription of the S regions essential to induce switch recombination. G4 DNA within G-loops is recognized by factors that promote genomic stability, including MutSα (35) and the RecQ family helicase BLM (36). AID, which promotes genomic instability, binds to ssDNA and has been directly imaged to be bound within single-stranded regions of G-loops formed at c-MYC or the S regions (34).

We have now surveyed proto-oncogenes shown to be unstable in B-cell lymphomas to establish whether they are distinguished by abundance of WRC motifs or G-rich sequence composition. Using software that we developed to characterize both these features of genomic sequence, we show that proto-oncogenes that are targets of aberrant hypermutation are not characterized by an unusually high density of WRC motifs but are G-rich. Using electron microscopic imaging, we verify experimentally that transcription-induced G-loops form in two G-rich and unstable proto-oncogenes, BCL6 and RhoH, but not in a control gene, a-MYB, which is not a target of translocation or aberrant hypermutation. By further genomic analysis, we show that genes that are targets of aberrant hypermutation in normal B cells, CD95/Fas, B29, and MBI, are similarly G-rich but do not contain an unusual density of WRC motifs. Conversely, we show that G-richness is not characteristic of regions near 105 independent breakpoints within 49 breakpoint clusters in 15 different genes that undergo translocations in AID-negative B- and T-cell malignancies, including AF6, AF9, AML1, CBFB, E2A, ETO, ML1, MYH11, NUP98, PBX1, PML, RAP1GDS1, RARA, TOPI, and TEL1. These results establish that genes targeted for instability in B-cell malignancies and normal B cells contain G-rich regions, which can form G-loops upon transcription. Thus, G-rich sequence composition is one feature of genomic structure that can contribute to genomic instability in B cells and B-cell malignancies.

Materials and Methods

Quantification of WRC motifs using WRC-Finder. To determine the frequency of the AID deamination consensus sequence motif WRC, we designed the computer program WRC-Finder (Supplementary Fig. S1A). This program reads a user-defined number of characters from a FASTA formatted sequence file and records the occurrence of the sequence WRC (W = A or T; R = A or G). The window to be analyzed is shifted along the sequence in single-nucleotide increments and the process repeated until the end of the file is reached. Output files from WRC-Finder were imported into Excel (Microsoft Corp., Redmond, WA) and plotted as bar graphs displaying the number of WRC motifs per 300-bp window. The average number of WRC motifs per window analyzed for each gene was determined by summing the output in Excel and dividing by the number of windows. The average for a group of genes was determined as the sum of the averages of the group, divided by the number of genes in that group.

Quantification of density of G-runs using G-Finder. To assess the density of G-runs and potential of a gene for G-loop formation, we designed the computer program G-Finder (Supplementary Fig. S1B). This program reads 100 characters from a FASTA formatted sequence file and records the number of G-runs in this sequence window. A G-run is defined as a stretch of three or more consecutive guanines. Each run containing three or more consecutive guanines is counted once. The window to be analyzed is shifted along the sequence in single-nucleotide increments until the end of the sequence file is reached. Output files from G-Finder were imported into Excel and plotted as bar graphs displaying the number of G-runs per 100-bp window. The average number of G-runs per window analyzed per gene was determined by summing the output in Excel and dividing by the number of windows. The average for a group of genes was determined as the sum of the averages of the group, divided by the number of genes.

Genomic sequences. Human genomic sequences analyzed were immunoglobulin S regions, nucleotides (nt) 1 to 3,000 (GenBank accession no. X54713); c-MYC, nt 87,842 to 90,842 (GenBank accession no. AC1038153); BCL6, nt 37,338 to 40,338 (GenBank accession no. A0720222.19); RhoH1, nt 68,817 to 71,817 (GenBank accession no. AC095057.3); PIM1, nt 26,687 to 29,687 (European Molecular Biology Laboratory (EMBL) accession no. AL553579.17); PAX5, nt 122,041 to 125,041 (EMBL accession no. AL161781.12); a-MYB, nt 23,786 to 26,786 (GenBank accession no. A083928.11); CD10/Calla, nt 67,225 to 64,225 (GenBank accession no. AC117348.5); L-Plastin, nt 99,980 to 72,980 (EMBL accession no. AL137141); NBS1, nt 143,751 to 146,751 (GenBank accession no. AF069291); CD95/Fas, nt 144,196 to 147,696 (EMBL accession no. AL157394.15); B29, nt 42,083 to 39,084 (GenBank accession no. AC127029.12); MBI, nt 38,328 to 41,328 (GenBank accession no. AC010616.3); TBP, nt 76,089 to 79,090 (EMBL accession no. AL031259.1); and ribosomal protein S14, nt 9,559 to 12,558 (GenBank accession no. AC011388.7).

Statistical analysis. A Mann-Whitney test (unpaired two-tailed t test assuming the sample distribution is not Gaussian) was done on average G-runs per 100 bp values per gene, comparing unstable and stable proto-oncogenes in AID-positive lymphomas, and unstable proto-oncogenes in AID-negative lymphomas and germinal center B cells to unstable proto-oncogenes in AID-negative B-cell and T-cell malignancies.

Plasmids. pbCL6 contains a 2.6-kb fragment spanning exon 1 and intron 1 of the human BCL6 gene (nt 37,288–39,865; GenBank accession no. A0720222.19). PCR amplified from human genomic DNA (Promega, Instability of G-Rich Proto-Oncogenes
Figure 1. Density of WRC motifs does not correlate with proto-oncogene instability. A, two immunoglobulin variable regions, VH1-2 and VH2-5, were scored for the presence of WRC motifs on both the nontemplate (above) and template (below) strands. The presence of WRC motifs was scored in a 300-bp moving window and graphed to show occurrence of WRC motifs per 300 bp (Y axis) relative to position in nt (X axis). B, c-MYC, BCL6, RhoH, PIM1, and PAX5 were scored for WRC repeats. The region of each gene analyzed spanned 3.0 kb and was bounded upstream at the 5’ boundary of exon 1 (or exon 1b in the case of PAX5). Horizontal bars between charts for nontemplate and template strands indicate the region targeted by aberrant hypermutation (14). C, genes that are not targeted for hypermutation in DLBCL (14) were scored for WRC repeats.
Instability of G-Rich Proto-Oncogenes

Madison, WI) using synthetic oligonucleotide primers 5′-GGAGCAGGCCACTCCATGT and 5′-CTCTCTCTGCCCCACGTCTTTT. pRhoH contains a 4.3-kb region spanning exon 1 to intron 1 of the human RhoH gene (nt 68,546–72,845; GenBank accession no. AC095057.3); PCR amplified from human genomic DNA using primers 5′-TGGTAAATTTACCTCCAGGG and 5′-CAGTCTGACTTTACGATTACG. pa-MYB contains a 1.2-kb region spanning exon 1 to the 5′ region of exon 2 of human a-MYB (nt 23,786–24,997; GenBank accession no. AC083928.11); PCR amplified from human genomic DNA using primers 5′-AGTGAGATGAGATGAGCC and 5′-GTTAGGCTCAGATGTAAGCT. PCR products were cloned into the PCRII vector for Topo cloning according to the manufacturer’s protocol (Invitrogen, Carlsbad, CA).

Transcription and electron microscopy. Transcription was carried out for 15 min at 37°C in reactions containing 60 µg/mL supercoiled plasmid DNA, 1 mmol/L each nucleotide triphosphate, 40 mmol/L KCl, and 50 units/mL of either T7 RNA polymerase (for pa-MYB reactions) or SP6 RNA polymerase (for pBCL6 and pRhoH/TTF reactions). T7 and SP6 RNA polymerases (NEB) were added in manufacturer’s buffer. Free RNA was digested by incubation with 20 µg/mL RNase A for 15 min at 37°C. DNAs were linearized at unique restriction sites by digestion with restriction enzymes (NEB) in manufacturer’s buffer: pBCL6 with HindIII; pRhoH/TTF with EcoRV; and pa-MYB with BglII. Samples were spread for transmission electron microscopy as previously described (31, 34) and imaged using a JEOL 1010 transmission electron microscope at 60 kV. Images were captured with a Gatan ultracam camera (Gatan, Pleasanton, CA) and acquired with Gatan Digital Micrograph software. Size and location of loops relative to the unique restriction site for each plasmid were measured using ImageJ (NIH).

Results

Density of WRC motifs does not correlate with proto-oncogene instability. The WRC motif is the minimal consensus motif for deamination by AID (24). We used the program WRC-Finder to determine the density of WRC motifs on both the template and non-template strands of five proto-oncogenes shown to be targets for aberrant hypermutation, c-MYC, BCL6, RhoH, PIM1, and PAX5 (Fig. 1B), and, for comparison, four genes shown not to be mutated in DLBCL (14), a-MYB, CD10/Calla, L-Plastin, and NBS1 (Fig. 1C). WRC motifs were scored within a 3.0-kb region bounded by the 5′ border of exon 1, with the exception of PAX5, in which exon 1B defined the border, corresponding to the regions that are targeted for aberrant hypermutation and translocation in B-cell lymphomas. Calculation of the average density of WRC motifs per window analyzed showed that this density did not differ between unstable proto-oncogenes and controls (average WRC density, 16 ± 0.8 and 16 ± 1.0, respectively; Table 1) or between these genes and the immunoglobulin V regions, which are natural targets for hypermutation (average WRC density, 15 ± 2.0; Table 1). In only one case, PAX5, did the peaks of WRC motif density track with the zone of hypermutation (Fig. 1B). Thus, with the possible exception of PAX5, a bias in the number and distribution of WRC motifs does not seem to account, in general, for targeting of aberrant hypermutation to specific proto-oncogenes.

G-Finder identifies G-rich regions within genomic DNA. We developed a program, G-Finder, to quantitate G-richness and the potential for G-loop formation. G-Finder scores the number of G-runs (three or more consecutive G’s) in a 100-bp window that moves in 1-bp increments. For example, the output from G-Finder illustrates the well-documented G-richness of the human immunoglobulin S region, known to be a target of translocations leading to B-cell lymphomas (Fig. 2A). In contrast, human VH1-2 and VH2-5 contain very few G-runs (Fig. 2B).

Table 1. Quantitation of average densities of WRC motifs and G-runs in genes that are targets for translocation and/or aberrant hypermutation, and controls

<table>
<thead>
<tr>
<th>Immunglobulin loci</th>
<th>WRC motifs</th>
<th>G-runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunoglobulin V regions (human VH1-2 and VH2-5)</td>
<td>15 (± 2.0)</td>
<td>1.7 (± 0.1)</td>
</tr>
<tr>
<td>Human immunoglobulin Sµ region</td>
<td>—</td>
<td>6.9 (± 2.3)</td>
</tr>
<tr>
<td>DLBCL (AID positive)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstable (BCL6, c-MYC, PAX5, PIM-1, RhoH)</td>
<td>16 (± 0.8)</td>
<td>2.1 (± 0.6)</td>
</tr>
<tr>
<td>Controls (a-MYB, CD10/Calla, L-Plastin, NBS1)</td>
<td>16 (± 1.0)</td>
<td>0.9 (± 0.2)</td>
</tr>
<tr>
<td>Normal germinal center B cells (AID positive)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstable (B99, MB1/CD79a, CD95/Fas)</td>
<td>15 (± 1.2)</td>
<td>2.5 (± 1.0)</td>
</tr>
<tr>
<td>Controls (SH4, TBP)</td>
<td>17 (± 2.0)</td>
<td>1.4 (± 0.9)</td>
</tr>
<tr>
<td>Leukemias (AID negative)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF6, AF9, AML1, CBFB, E2A, ETO, MLL, MYH11, NUP98, PBX1, PML, RAP1GDS1, RARA, TOPI, TEL1</td>
<td></td>
<td>1.0 (± 0.4)</td>
</tr>
</tbody>
</table>

NOTE: Quantitation was carried out as described in Materials and Methods. BCL6, which is a target of instability in both DLBCL and normal germinal center B cells, was included in the former group in this analysis. The G-run average for genes from AID-negative cells is the average of all analyzed translocation breakpoint regions.
Proto-oncogenes that are unstable in DLBCL contain multiple G-runs as identified by G-Finder. The program G-Finder scored G-run content of genomic sequences. X axis, relative position (nt); Y axis, number of G-runs per 100 bp. In each region, G-richness of the non-template strand was scored. A, human IgH S\textsubscript{μ} switch region. B, human V regions, V\textsubscript{H}1-2 and V\textsubscript{H}1-22-5. C, proto-oncogenes c-MYC, BCL6, RhoH, PIM1, and PAX5, which are targets of instability in DLBCL. The region scored was bounded upstream by the 5' end of exon 1 (or exon 1b for PAX5). Horizontal bar below indicates the region of each gene targeted by aberrant hypermutation (14). D, a-MYB, CD10/Calla, L-Plastin, and NBS1, which are not targeted for hypermutation in DLBCL. The region scored was the same as in (C).
RhoH, PIM1, and PAX5. These genes are all transcribed in DLBCL (37) and thus satisfy one criterion for G-loop formation (31). We also examined as controls the a-MYB proto-oncogene, CD10/Calla, L-Plastin, and NBS1, which are transcribed but not targets of aberrant hypermutation in DLBCL (14, 37). G-runs were conspicuously more abundant in the unstable proto-oncogenes (Fig. 2C) than in the controls (Fig. 2D). The average density of G-runs was $2.1 \pm 0.6$ in proto-oncogenes, which are targets of instability, and $0.9 \pm 0.2$ in the control genes (Table 1). This 2.3-fold difference is highly significant ($P = 0.0159$). There were also considerable local differences in G-richness between genes in each class. c-MYC and BCL6 were notable in containing three or more regions in which there were $\geq 8$ G-runs per 100 bp (evident as peak heights $\geq 8$; Fig. 2C). RhoH was the least G-rich of the unstable proto-oncogenes surveyed but it did contain several extensive regions with $\geq 4$ G-runs per 100 bp. In comparison, three genes not targeted for aberrant hypermutation, a-MYB, CD10, and L-Plastin, contained at most three extended regions of G-runs, and the maximum density of G-runs in those peaks was only 4 runs per 100 bp (Fig. 2D). NBS1 did contain a G-rich region but this was located very close to the promoter and may correspond to a CpG island.

G-loops form in G-rich proto-oncogenes. Both c-MYC and the immunoglobulin S regions form G-loops upon transcription, and these loops are bound by AID (34). We asked if G-richness identified by G-Finder (Fig. 2) correlated with G-loop formation by analyzing structures formed on transcription of three proto-oncogenes: BCL6, one of the most G-rich of the unstable proto-oncogenes; RhoH, the least G-rich of this subset; and a-MYB, which is neither G-rich nor unstable. Regions of these genes were cloned into plasmid vectors just downstream of a promoter for in vitro transcription to create the corresponding plasmids pBCL6, pRhoH, and pa-MYB. Plasmid DNA templates were transcribed in vitro; free RNA was digested with RNase A; and DNA was then linearized at a unique restriction site and imaged by transmission electron microscopy. G-loops were evident in 10% to 20% of the BCL6 and RhoH DNA templates (in 100 and 330 molecules, respectively), as illustrated by representative images (Fig. 3A and B). No loops were evident within transcribed a-MYB templates (0 loops in >100 molecules examined), as shown by a

![Figure 3](image-url)

**Figure 3.** G-loops form in transcribed G-rich proto-oncogenes. Representative electron microscopic images of plasmid DNAs following in vitro transcription. Arrows, G-loops. White box, exon 1. Bar, 500 nm. A, transcribed pBCL6 template. B, transcribed pRhoH template. C, transcribed pa-MYB template. D, mapping of loops formed on transcribed pBCL6 and pRhoH/TTF templates (see Materials and Methods). Arrows above denote mapped loops on individual molecules. White box, exon 1. Bar below denotes the region targeted for aberrant hypermutation (9, 14); vertical lines, translocation sites in BCL6 (5).
representative image (Fig. 3C). The sizes of G-loops ranged from 110 to 1,280 bp in BCL6 and from 120 to 770 bp in RhoH. Smaller loops may have been present but would not have been detected by electron microscopy, as the minimum loop size visible is ~100 bp.

The distribution of G-loops in BCL6 and RhoH is similar to the distribution within c-MYC that we previously reported (34). Analysis with G-Finder showed that, like BCL6 and RhoH, c-MYC is G-rich (Fig. 2C). Thus, there is a good correspondence between G-richness, as established by G-Finder, and formation of G-loops on in vitro transcription.

G-loops map to regions associated with genomic instability. The positions of the G-loops in transcribed plasmids pBCL6 and pRhoH were mapped with respect to the restriction cleavage sites in the plasmid templates (Fig. 3D). G-loops mapped to regions of BCL6 and RhoH targeted for aberrant hypermutation and translocation (5, 8, 9, 14, 38). Transcription-induced G-loops in c-MYC similarly map to the zone associated with instability (34).

G-richness characterizes genes that mutate in normal germlinal center B cells. The results above show that G-richness correlates with genomic instability in DLBCL. Aberrant hypermutation can alter the sequences of non-immunoglobulin genes not only in tumors but also in normal germlinal center B cells, where one target is BCL6, which is G-rich (Fig. 1). Mutations have also been documented in CD95/Fas, B29, and MB1 (39, 40), and somatic mutation of CD95/Fas and B29 also occurs in malignancies including multiple myeloma, Hodgkin’s lymphoma, non-Hodgkin’s lymphoma, and chronic lymphocytic leukemia (39, 41–43). To establish how density of WRC motifs and G-richness might contribute to aberrant hypermutation of CD95/Fas and B29, and MB1, we analyzed these genes with WRC-Finder and G-Finder (Supplementary Fig. S2A). In comparison, we examined two genes shown not to be mutated in normal germlinal center B cells, TBP and S14 (44), which encode the TATA binding protein and ribosomal protein S14, respectively (Supplementary Fig. S2B). In each case, analysis focused on a 3-kb region bounded by the 5’ border of exon 1, which includes the hypermutating zone. The average density of WRC motifs was 15 ± 1.2 in genes targeted for aberrant hypermutation and 17 ± 2.0 in the control genes (Table 1). In contrast, the average density of G-runs was 2.5 ± 1.0 in the former group and 1.4 ± 0.9 in the controls (Table 1), which corresponds to a 1.6-fold difference in G-richness. Thus, G-richness correlated with instability in AID-positive B cells.

Local differences were noted within specific genes. In CD95/Fas, hypermutation is not uniformly distributed but concentrates within a short (400 bp) region near the 5’ end of the gene. CD95/Fas was not uniformly G-rich but contained a very G-rich region at the 5’ end, which coincided with the region prone to hypermutation. WRC motifs were uniformly distributed along the gene (Supplementary Fig. S2A), similar to the distribution of WRC motifs in V_{H}1-2 and V_{H}2-5 (Fig. 1A). Both B29 and MB1 were G-rich and both contained regions in which WRC motifs were quite abundant. Neither TBP nor S14 contained an unusual density of G-runs, although S14 did contain a 5’ G-rich region; this may correspond to a CpG island, as in NBS1 (Fig. 2D).

G-richness does not correlate with translocation breakpoints in AID-negative leukemias. Translocation of specific genes characterizes not only AID-positive B-cell lymphomas but also a variety of hematopoietic malignancies. To establish whether G-richness correlates with translocation in AID-negative tumors, we analyzed recurrent translocation breakpoints in 15 genes from 7 different AID-negative B- and T-cell malignancies, including acute myeloid leukemia, childhood and adult acute lymphoblastic leukemia, therapy-related myelodysplastic syndrome, T-cell acute lymphocytic leukemia, childhood acute lymphoblastic leukemia, and acute promyelocytic leukemia. AID is not expressed in the originating cell types, nor has AID expression been documented in the resulting malignancies. The genes analyzed were MLL, AF9, AF6, AML1, CBFB, E2A, ETO, MYH11, NUP98, PBX1, PML, RAP1GDS1, RARA, TEL1, and TOP1. Recurrent translocations have been mapped to clustered breakpoints within transcribed regions in all these genes. (See Supplementary Fig. S3 for references and breakpoint accession numbers.) Sequences within 1.5 kb upstream and downstream of breakpoints were analyzed by G-Finder, plotting G-runs per 100 bp separately for each breakpoint region analyzed. If more than one breakpoint fell within a 1.5-kb region, they were analyzed together as a cluster. A total of 105 breakpoints, within 49 separate clusters, were analyzed in these 15 genes (Supplementary Fig. S3). This analysis showed that the majority of breakpoints fell within regions in which G-runs peaked at or below 4 per 100 bp.

To compare G-richness of regions analyzed, we tabulated densities of G-runs as a per gene average (Supplementary Table S1). We graphed the average G-run density per unstable gene region analyzed for two groups of genes: those that are unstable in AID-positive DLBCL or in germlinal center B cells and those that are unstable in AID-negative B- and T-cell malignancies (Fig. 4). The average density of G-runs near breakpoints in AID-negative malignancies was 1.00 ± 0.4 (Table 1). This is 2.1-fold lower than the average for genes that are targets of instability in AID-positive DLBCL alone (P = 0.0009) and 2.26-fold lower than the average for genes that are targets of instability in all AID-positive B cells examined, including DLBCL and normal germlinal center B cells (P < 0.0001; Fig. 4). Thus, G-richness correlates with instability in AID-positive but not AID-negative B and T cells.

![Figure 4. G-richness correlates with instability in AID-positive but not AID-negative B and T cells.](image-url)

**Average: 2.26**

**AID-positive**

**AID-negative**

**P-value:** <0.0001

Bars, average for each group, also shown below the figure.
Figure 5. Model for genomic instability at G-rich regions in AID-positive B cells. Transcription of a G-rich region induces formation of G-loops, which contain a cotranscriptional RNA/DNA hybrid on the template strand, and G4 DNA interspersed with single-stranded regions on the G-rich strand. AID binds to single-stranded regions of the non-template strand and initiates lesions by deamination. Lesions created by AID are processed to produce single-strand breaks, which are targets for aberrant hypermutation or translocation.

Discussion

We have shown that the subset of proto-oncogenes targeted for aberrant hypermutation in AID-positive B-cell tumors and normal B cells, BCL6, c-MYC, PAX5, Pim1, and RhoH, the tumor suppressor gene CD95/Fas, and B-cell receptor proteins B29 and MB1, all share a key feature of genomic sequence: extended regions containing runs of consecutive guanines on the non-template DNA strand. We have directly shown that transcription of BCL6 and RhoH results in formation of G-loops, DNA structures containing a stable RNA/DNA hybrid on the template strand and G4 DNA on the non-template strand (Fig. 3). Identical structures form in transcribed S regions and c-MYC (31, 34). The proto-oncogene a-MIB is not a target for aberrant hypermutation in DLBCL, and it is not G-rich (Fig 2D; Table 1) nor does it form G-loops upon transcription (Fig 3C).

Genes that were unstable in AID-positive B cells were not distinguished by an elevated density of WRC sequence motifs, the consensus target of AID (Table 1). In certain instances, locally high densities of the WRC motif may increase instability within limited regions of specific genes. This could occur within a region of intron 1 of c-MYC, which is G-rich and also contains a very high local density of WRC motifs (which corresponds to nt 366–1,989 in Figs. 1B and 2C), and which is targeted for translocation (45).

Quantitation of densities of G-runs at regions of instability identified a significantly higher average density in genes that were targets of instability in AID-positive than in AID-negative cells. As shown in Fig. 4, the average density of G-runs near regions of gene instability was 2.26-fold higher in AID-positive than in AID-negative B and T cells (P < 0.0001). These results argue that instability is AID dependent in germinal center B cells and tumors arising from this cell type.

We propose that G-rich sequence composition contributes to genomic instability of specific oncogenes in B-cell lymphomas, just as it contributes to recombination of the G-rich mammalian switch regions, by enhancing AID-initiated DNA deamination. As outlined in Fig. 5, transcription of a G-rich region would result in G-loop formation, increasing accessibility of single-stranded regions on the non-template DNA strand to the B-cell–specific DNA deaminase AID. G-loops are targets for AID (34), and AID preferentially deaminates single-stranded regions of DNA (30). AID is known to be critical to translocations leading to lymphomagenesis in murine models (14, 23, 46). A link between AID and aberrant hypermutation has recently been documented in human B cells infected with EBV, which results in AID overexpression and increased mutation of both BCL6 and P53 (47). The mechanism diagrammed in Fig 5 could promote aberrant mutation as well as translocation.

Our results suggest that most translocations contributing to leukemias in AID-negative cells result from a mechanism that does not depend on high G-run content. However, we do not exclude the possibility that, in specific instances, high G-content may contribute to gene instability in AID-negative cells. Oncogenes are, as a class, more G-rich than the genomic average (48); thus, they may be readily targeted by any mechanism that promoted instability at G-rich regions. For example, one translocation pair analyzed from AID-negative leukemias, RARA and PML, contained G-rich breakpoint regions that were significantly more G-rich than all other breakpoint regions analyzed (Supplementary Fig. S3vi; Supplementary Table S1). G4 DNA structures formed within G-rich regions are targets for repair factors including the RecQ family helicases BLM and WRN (36, 49, 50) and MutSα (35). Analysis of tumors deficient in these factors may reveal other examples of instability at G-rich regions.

Acknowledgments

Received 7/3/2006; revised 1/16/2007; accepted 1/18/2007.

Grant support: NIH grants R01 41712 and R01 GM65988 (N. Maizels).

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

We thank the members of the Maizels laboratory, especially Johanna Eddy, for valuable discussions.

References

8. Dallery E, Galet-Cazes-Zouitina S, Collyn-d’Hooghe M, et al. TTF, a gene encoding a novel small G protein, fuses to...
G-Rich Proto-Oncogenes Are Targeted for Genomic Instability in B-Cell Lymphomas

Michelle L. Duquette, Michael D. Huber and Nancy Maizels


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/67/6/2586

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2007/03/12/67.6.2586.DC1

Cited articles
This article cites 49 articles, 27 of which you can access for free at:
http://cancerres.aacrjournals.org/content/67/6/2586.full.html#ref-list-1

Citing articles
This article has been cited by 17 HighWire-hosted articles. Access the articles at:
/content/67/6/2586.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.