G-Rich Proto-Oncogenes Are Targeted for Genomic Instability in B-Cell Lymphomas

Michelle L. Duquette,[1] Michael D. Huber,[2] and Nancy Maizels[2]

Departments of [1]Immunology and [2]Biochemistry, University of Washington School of Medicine, Seattle, Washington

Abstract

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, PIM1*, and *PAX5*, 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]

Introduction

Specific proto-oncogenes are targets of genomic instability in tumors that derive from activated B cells. Instability can be evident as translocation or as accumulation of point mutations, referred to as aberrant hypermutation. Translocation has been most extensively documented at c-*MYC*. c-*MYC* recombines with the immunoglobulin heavy chain locus (IgH) switch (S) regions in sporadic Burkitt's lymphoma, resulting in the t(8;14) translocation that is the hallmark of this malignancy (1). c-*MYC*/IgH translocations are also found in multiple myeloma (2, 3). c-*MYC*/IgH translocations can contribute significantly to tumorigenesis by juxtaposing c-*MYC* to the active immunoglobulin enhancer and promoting deregulated expression of this critical gene. The *BCL6* (4, 5), *PAX5* (6), *PIM1* (7), and *RhoH* (also known as *TTF*; ref. 8) proto-oncogenes have also been found to translocate in B-cell lymphomas. Aberrant hypermutation has been documented at *BCL6*, c-*MYC, PAX5, PIM1*, and *RhoH* in B-cell malignancies and at *BCL6* in normal human germinal center B cells (9–18).

Proto-oncogene activation by aberrant hypermutation and translocation has been causally related to B-cell lymphoma development and also correlates with aggressive tumor growth and poor disease prognosis. Somatic hypermutation of 5′ noncoding regions in *BCL6* and c-*MYC* can result in deregulation of proto-oncogene transcription, which contributes to transformation (9, 15, 19). Aberrant somatic hypermutation of c-*MYC, PAX5, PIM1*, and *RhoH*, as well as *BCL6* translocation, are associated with progression from follicular lymphoma to the more aggressive diffuse large B-cell lymphoma (DLBCL), and *PAX5*/IgH translocations have been found in a subset of aggressive non-Hodgkin's lymphomas (5, 20). Thus, it is of considerable interest to understand the mechanisms that promote both aberrant hypermutation and translocation of proto-oncogenes.

Aberrant hypermutation and translocation events that alter proto-oncogene sequence and structure in activated B cells seem to depend on mechanisms that promote changes in genomic sequence and structure essential to the immune response (21–23). In antigen-activated B cells, somatic hypermutation produces single-base changes in the rearranged and expressed variable regions and, when coupled with selection, generates antibodies with increased affinity and specificity. Class switch recombination deletes a large region of chromosomal DNA, replacing one constant region with another and thereby optimizing antigen clearance. Antigen-activated B cells occupy special microenvironments of secondary lymphoid tissue, called germinal centers. Germinal center B cells express a mutagenic factor, activation-induced deaminase (AID), which deaminates C to U in DNA to initiate both class switch recombination and somatic hypermutation. The sequence motif WRC (W = A or T; R = G or A) is a hotspot for hypermutation of immunoglobulin variable regions and a preferential target for deamination *in vitro* (24). A similar pattern of mutation is produced upon aberrant somatic mutation of proto-oncogenes and immunoglobulin variable regions: mutations are preferentially targeted to the WRC motif and single-nucleotide substitutions predominate, accompanied by occasional deletions and insertions (14, 25). AID is expressed in normal germinal center B cells, in B-cell tumors including germinal center B-cell non-Hodgkin's lymphomas, and in subsets of nongerminal center B-cell non-Hodgkin's lymphomas (26). AID is required for c-*Myc* translocation in the mouse model for Burkitt's lymphoma and pristane-induced plasmacytoma (27); translocation, in turn, promotes tumorigenesis (28). Translocations between c-*Myc* and IgH are induced rapidly on induction of AID expression in primary murine B cells and depend on AID deaminase activity (29).

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Present address for M.L. Duquette: Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037.

Present address for M.D. Huber: Department of Cellular Biology, The Scripps Research Institute, La Jolla, CA 92037.

Requests for reprints: Nancy Maizels, Department of Immunology, University of Washington School of Medicine, 1959 Northeast Pacific Street, Box 357650, Seattle, WA 98195. Phone 206-221-6876; Fax: 206-221-6781; E-mail: maizels@u.washington.edu.

doi:10.1158/0008-5472.CAN-06-2419

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 *RhoH*; 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-*MYB, 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 structures formed 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 they 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 *MB1*, 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, MLL, MYH11, NUP98, PBX1, PML, RAP1GDS1, RARA, TOP1*, 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. S1*A*). 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.

Quantitation 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. S1*B*). 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 Sμ region, nucleotides (nt) 1 to 3,000 (GenBank accession no. X54713); c-*MYC*, nt 87,842 to 90,842 (GenBank accession no. AC103819.3); *BCL6*, nt 37,338 to 40,338 (GenBank accession no. AC072022.19); *RhoH*, 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. AL353579.17]; *PAX5*, nt 122,041 to 125,041 (EMBL accession no. AL161781.12); a-*MYB*, nt 23,786 to 26,786 (GenBank accession no. AC083928.11); *CD10/Calla*, nt 67,225 to 64,225 (GenBank accession no. AC117384.5); *L-Plastin*, nt 69,980 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); *MB1*, nt 38,328 to 41,328 (GenBank accession no. AC010616.5); *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-positive 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. AC072022.19), PCR amplified from human genomic DNA (Promega,

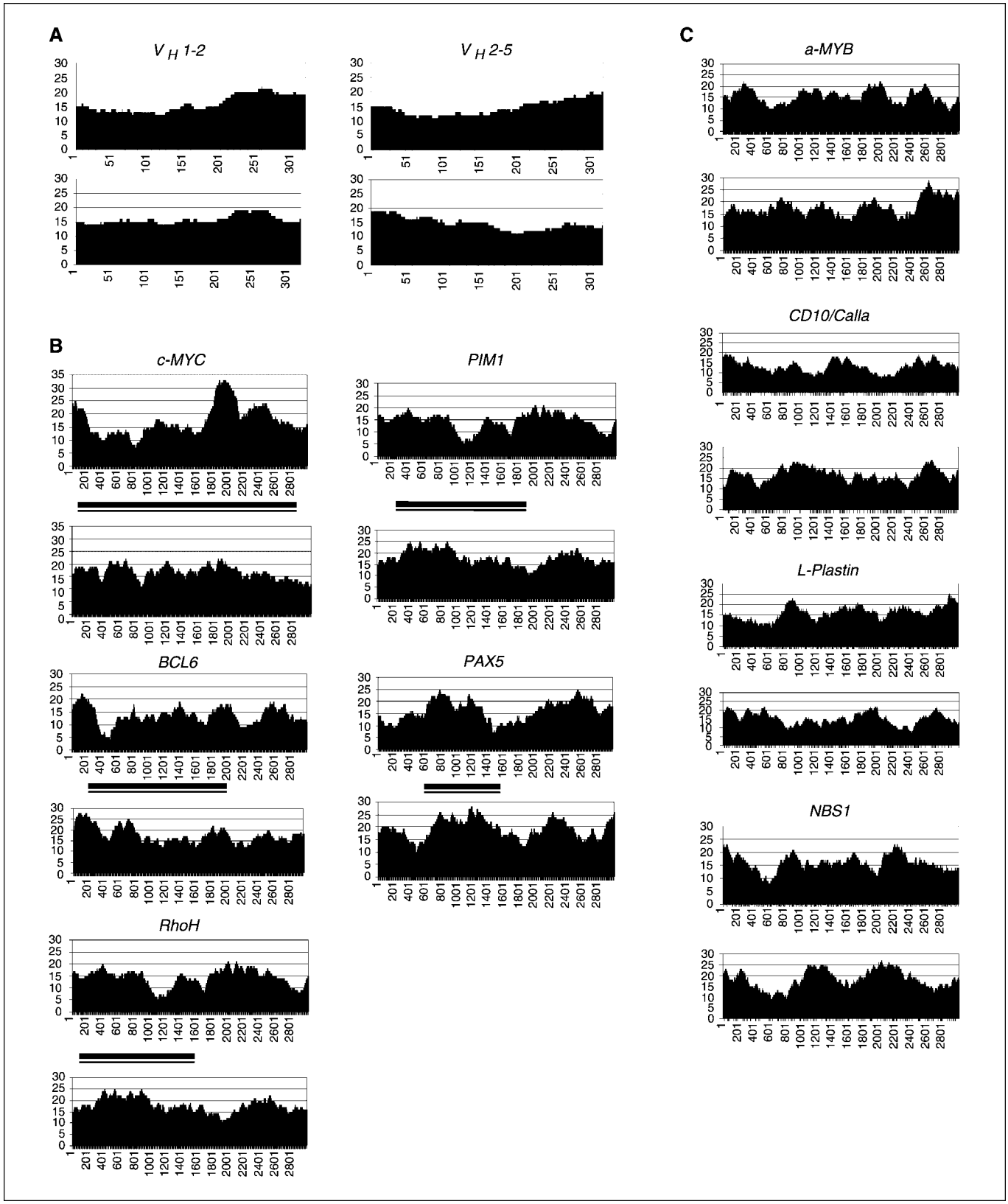

Figure 1. Density of WRC motifs does not correlate with proto-oncogene instability. *A,* two immunoglobulin variable regions, V_H1-2 and V_H2-5, were scored for the presence of WRC motifs on both the nontemplate (*above*) and template (*below*) strands. The presence of WRC motifs was scored in a 300-bp 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.

Madison, WI) using synthetic oligonucleotide primers 5′-GGAGCAGGCCATACCATCGT and 5′-CTCTCTCCTGCCCCACTTTT. 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′-TGGTAATTTTACTTCCATGAGG and 5′-CACTGTGACTTCAGTTTTACG. 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′-AGTGAGGATGAGGATGATGACC and 5′-GGTAAGCCTCAGATGATAAGCT. 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 *Hin*dIII; pRhoH/TTF with *Eco*RV; and pa-MYB with *Bgl*II. 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 ultrascan camera (Gatan, Pleasanton, CA) and acquired using 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 nontemplate strands of two human immunoglobulin V_H regions to serve as a reference, V_H1-2 (IGHV1-2, GenBank accession no. X07448) and V_H2-5 (IGHV2-5, GenBank accession no. X62111). The density of WRC motifs was relatively uniform on the two strands of each V_H region (Fig. 1*A*) and the average WRC density within these regions was 15 ± 2.0 (Table 1). WRC-Finder was also used to determine the density of WRC motifs on both the template and nontemplate strands of five proto-oncogenes shown to be targets of aberrant hypermutation, c-*MYC, BCL6, RhoH, PIM1*, and *PAX5* (Fig. 1*B*), and, for comparison, four genes shown not to be mutated in DLBCL (14), a-*MYB, CD10/Calla, L-Plastin*, and *NBS1* (Fig. 1*C*). 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. 1*B*). 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 Sμ region, known to be a target of translocations leading to B-cell lymphomas (Fig. 2*A*). In contrast, human V_H1-2 and V_H2-5 contain very few G-runs (Fig. 2*B*). Quantitation showed that the density of G-runs in the human Sμ region is much higher than in the V regions (6.9 compared with 1.7; Table 1).

G-richness correlates with genomic instability. We used G-Finder to examine proto-oncogenes shown to be targets for aberrant hypermutation in DLBCL (14), including c-*MYC, BCL6*,

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

	WRC motifs	G-runs
Immunoglobulin loci		
Immunoglobulin V regions (human V_H1-2 and V_H2-5)	15 (± 2.0)	1.7 (± 0.1)
Human immunoglobulin Sμ region	—	6.9 (± 2.3)
DLBCL (AID positive)		
Unstable (*BCL6*, c-*MYC, PAX5, PIM-1, RhoH*)	16 (± 0.8)	2.1 (± 0.6)
Controls (a-*MYB, CD10/CALLA, L-Plastin, NBS1*)	16 (± 1.0)	0.9 (± 0.2)
Normal germinal center B cells (AID positive)		
Unstable (*B29, MB1/CD79a, CD95/Fas*)	15 (± 1.2)	2.5 (± 1.0)
Controls (*S14, TBP*)	17 (± 2.0)	1.4 (± 0.9)
Leukemias (AID negative)	—	1.0 (± 0.4)
AF6, AF9, AML1, CBFB, E2A, ETO, MLL, MYH11, NUP98, PBX1, PML, RAP1GDS1, RARA, TOP1, TEL1		

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.

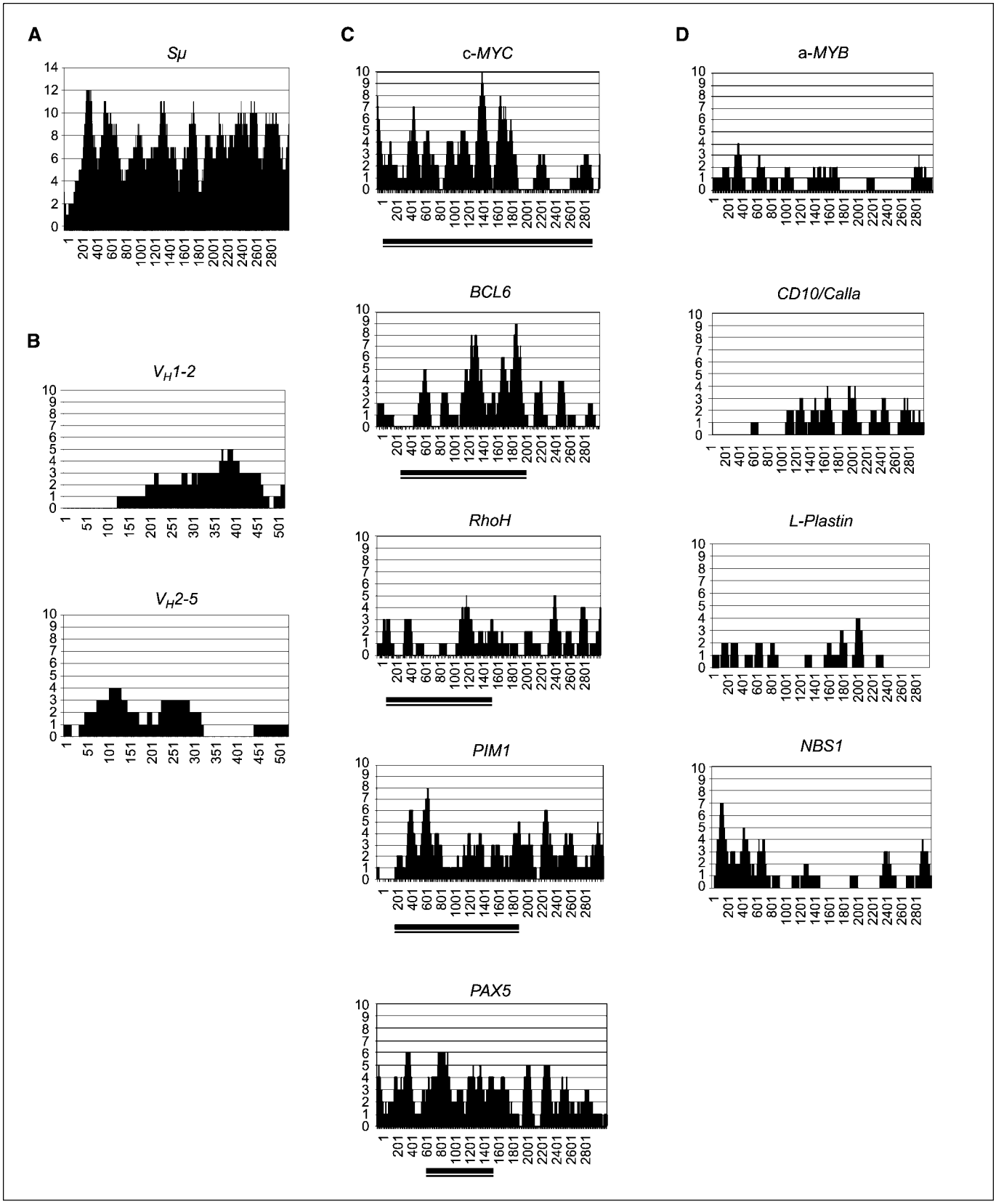

Figure 2. Proto-oncogenes that are unstable in DLBCL contain multiple G-runs as identified by G-Finder. The program G-Finder scored G-run content of genomic sequences. *X axis,* relative position (nt); *Y axis,* number of G-runs per 100 bp. In each region, G-richness of the nontemplate strand was scored. *A,* human IgH Sμ switch region. *B,* human V regions, V_H1-2 and V_H1-22-5. *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. 2*C*) than in the controls (Fig. 2*D*). 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. 2*C*). *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. 2*D*). *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. 3*A* and *B*). No loops were evident within transcribed a-*MYB* templates (0 loops in >100 molecules examined), as shown by a

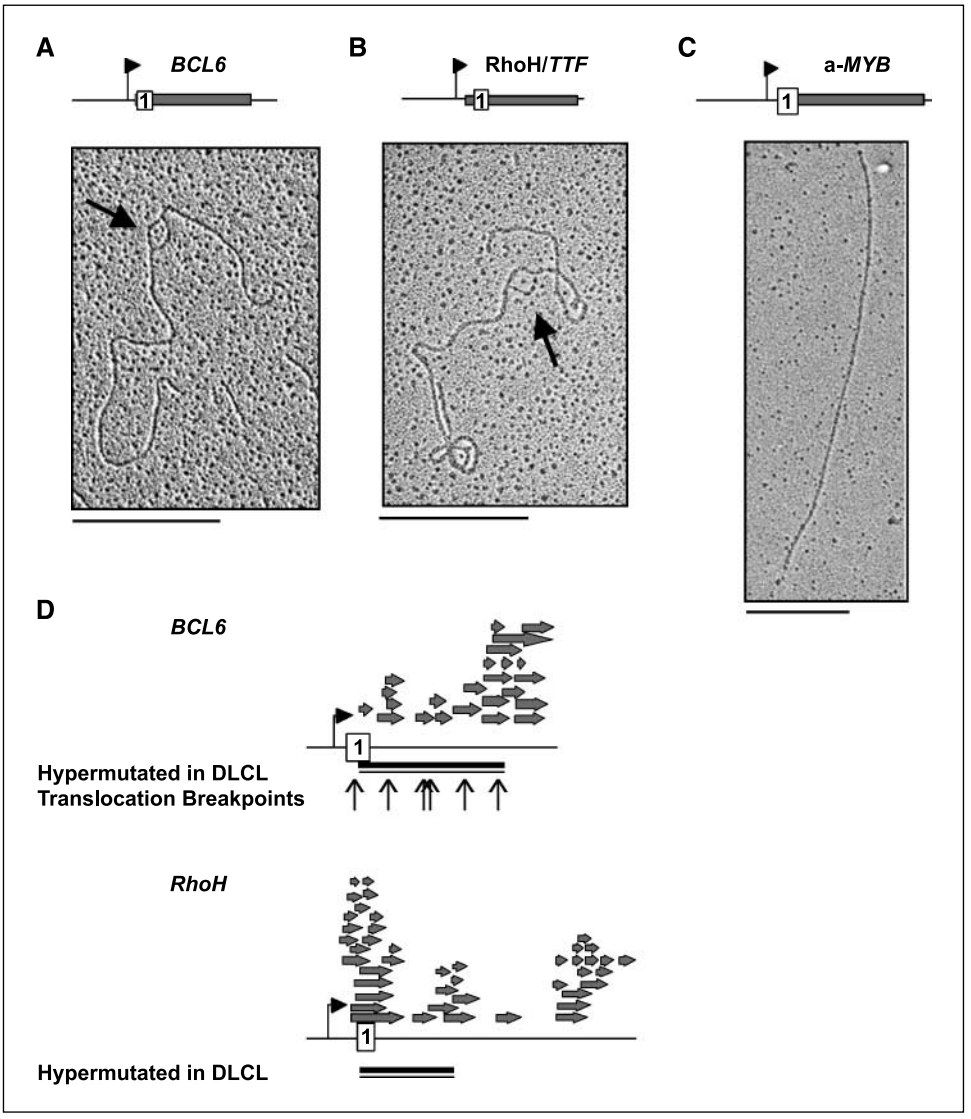

Figure 3. G-loops form in transcribed G-rich proto-oncogenes. Representative electron microscopic images of plasmid DNAs following *in vitro* transcription. *Arrows,* G-loops. *White box,* exon 1. *Bar,* 500 nm. *A,* transcribed pBCL6 template. *B,* transcribed pRhoH template. *C,* transcribed pa-MYB 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. 3*C*). 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. 2*C*). 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. 3*D*). 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 germinal 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 germinal 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, B29*, and *MB1*, we analyzed these genes with WRC-Finder and G-Finder (Supplementary Fig. S2*A*). In comparison, we examined two genes shown not to be mutated in normal germinal center B cells, *TBP* and *S14* (44), which encode the TATA binding protein and ribosomal protein *S14*, respectively (Supplementary Fig. S2*B*). 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. S2*A*), similar to the distribution of WRC motifs in V_H1-2 and V_H2-5 (Fig. 1*A*). 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. 2*D*).

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, PBX, 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 germinal 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 germinal 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.

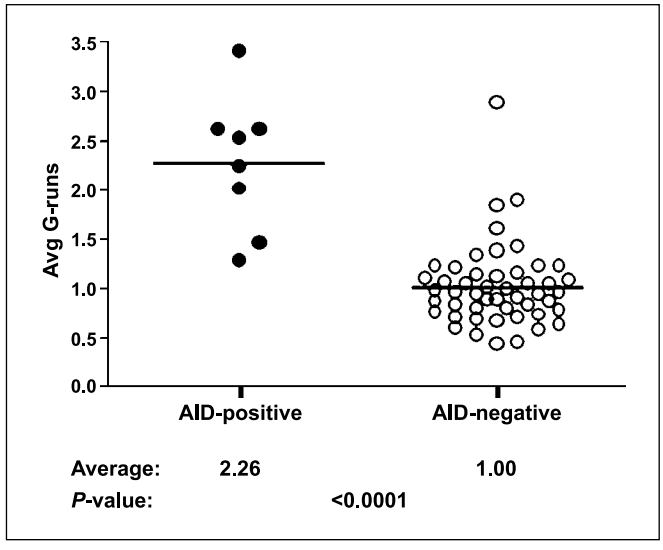

Figure 4. G-richness correlates with instability in AID-positive but not AID-negative B and T cells. *Left,* average G-richness per 100 bp of gene regions that are unstable in AID-positive cells, including *BCL6* (unstable in DLBCL and normal germinal center B cells); c-*MYC, PAX5, PIM1*, and *RhoH* (unstable in DLBCL); and *B29, CD79A*, and *CD95/FAS* (unstable in germinal center B cells). *Right, AF6, AF9, AML1, CBFB, E2A, ETO, MLL, MYH11, NUP98, PBX1, PML, RAP1GDS1, RARA, TOP1*, and *TEL1* breakpoint clusters (all of which are sites of translocation in AID negative leukemias). Average G-richness of all regions analyzed per gene was tabulated in Supplementary Table S1. *Bars,* average for each group, also shown below the figure.

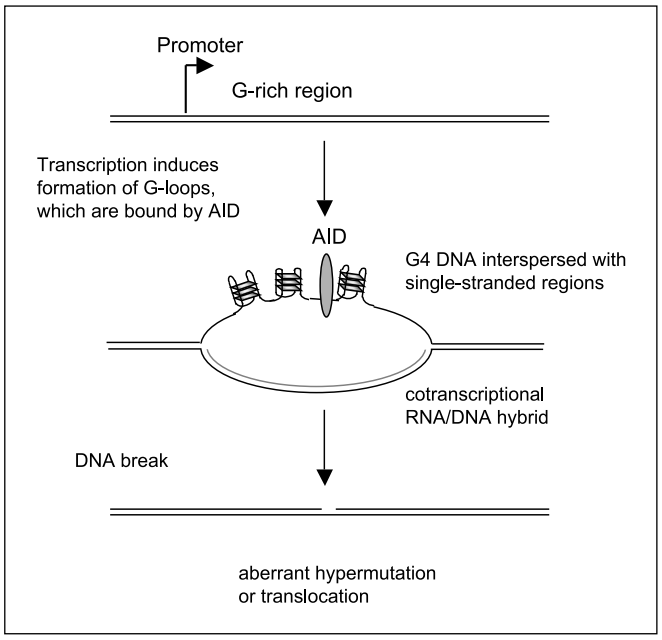

Figure 5. Model for genomic instability at G-rich regions in AID-positive B cells. Transcription of a G-rich region induces formation of G-loops, which contain a cotranscriptional RNA/DNA hybrid on the template strand, and G4 DNA interspersed with single-stranded regions on the G-rich strand. AID binds to single-stranded regions of the nontemplate 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, *BCL*6, 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 nontemplate 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 nontemplate strand (Fig. 3). Identical structures form in transcribed S regions and c-*MYC* (31, 34). The proto-oncogene a-*MYB* is not a target for aberrant hypermutation in DLBCL (14), and it is not G-rich (Fig. 2*D*; Table 1) nor does it form G-loops upon transcription (Fig. 3*C*).

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. 1*B* and 2*C*), 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 nontemplate 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. S3*vii*; 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

1. Kuppers R, Dalla-Favera R. Mechanisms of chromosomal translocations in B cell lymphomas. Oncogene 2001;20:5580–94.
2. Avet-Loiseau H, Gerson F, Magrangeas F, Minvielle S, Harousseau JL, Bataille R. Rearrangements of the c-myc oncogene are present in 15% of primary human multiple myeloma tumors. Blood 2001;98:3082–6.
3. Bergsagel PL, Kuehl WM. Chromosome translocations in multiple myeloma. Oncogene 2001;20:5611–22.
4. Butler M, Corbally N, Dervan PA, Carney DN. BCL-6 and other genomic alterations in non-Hodgkin's lymphoma (NHL). Br J Cancer 1997;75:1641–5.
5. Akasaka T, Lossos IS, Levy R. BCL6 gene translocation in follicular lymphoma: a harbinger of eventual transformation to diffuse aggressive lymphoma. Blood 2003;102:1443–8.
6. Poppe B, De Paepe P, Michaux L, et al. PAX5/IGH rearrangement is a recurrent finding in a subset of aggressive B-NHL with complex chromosomal rearrangements. Genes Chromosomes Cancer 2005;44:218–23.
7. Jacob AK, Sreekantaiah C, Baer MR, Sandberg AA. Translocation (1;6)(p12;p23) in ANLL. Cancer Genet Cytogenet 1990;45:67–71.
8. Dallery E, Galiegue-Zouitina S, Collyn-d'Hooghe M, et al. TTF, a gene encoding a novel small G protein, fuses to

the lymphoma-associated LAZ3 gene by t(3;4) chromosomal translocation. Oncogene 1995;10:2171–8.
9. Migliazza A, Martinotti S, Chen W, et al. Frequent somatic hypermutation of the 5′ noncoding region of the BCL6 gene in B-cell lymphoma. Proc Natl Acad Sci U S A 1995;92:12520–4.
10. Pasqualucci L, Migliazza A, Fracchiolla N, et al. BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. Proc Natl Acad Sci U S A 1998;95:11816–21.
11. Shen HM, Peters A, Baron B, Zhu X, Storb U. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science 1998;280:1750–2.
12. Peng HZ, Du MQ, Koulis A, et al. Nonimmunoglobulin gene hypermutation in germinal center B cells. Blood 1999;93:2167–72.
13. Capello D, Vitolo U, Pasqualucci L, et al. Distribution and pattern of BCL-6 mutations throughout the spectrum of B-cell neoplasia. Blood 2000;95:651–9.
14. Pasqualucci L, Neumeister P, Goossens T, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature 2001;412:341–6.
15. Pasqualucci L, Migliazza A, Basso K, Houldsworth J, Chaganti RS, Dalla-Favera R. Mutations of the BCL6 proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma. Blood 2003;101:2914–23.
16. Gaidano G, Pasqualucci L, Capello D, et al. Aberrant somatic hypermutation in multiple subtypes of AIDS-associated non-Hodgkin lymphoma. Blood 2003;102:1833–41.
17. Montesinos-Rongen M, Van Roost D, Schaller C, Wiestler OD, Deckert M. Primary diffuse large B-cell lymphomas of the central nervous system are targeted by aberrant somatic hypermutation. Blood 2004;103:1869–75.
18. Liso A, Capello D, Marafioti T, et al. Aberrant somatic hypermutation in tumor cells of nodular-lymphocyte-predominant and classic Hodgkin lymphoma. Blood 2006;108:1013–20.
19. Bhatia K, Spangler G, Hamdy N, et al. Mutations in the coding region of c-myc occur independently of mutations in the regulatory regions and are predominantly associated with myc/Ig translocation. Curr Top Microbiol Immunol 1995;194:389–98.
20. Rossi D, Berra E, Cerri M, et al. Aberrant somatic hypermutation in transformation of follicular lymphoma and chronic lymphocytic leukemia to diffuse large B-cell lymphoma. Haematologica 2006;91:1405–9.
21. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature 2003;422:726–30.
22. Maizels N. Immunoglobulin gene diversification. Annu Rev Genet 2005;39:23–46.
23. Ramiro AR, Nussenzweig MC, Nussenzweig A. Switching on chromosomal translocations. Cancer Res 2006;66:7837–9.
24. Pham P, Bransteitter R, Petruska J, Goodman MF. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature 2003;424:103–7.
25. Schwindt H, Akasaka T, Zuhlke-Jenisch R, et al. Chromosomal translocations fusing the BCL6 gene to different partner loci are recurrent in primary central nervous system lymphoma and may be associated with aberrant somatic hypermutation or defective class switch recombination. J Neuropathol Exp Neurol 2006;65:776–82.
26. Pasqualucci L, Guglielmino R, Houldsworth J, et al. Expression of the AID protein in normal and neoplastic B cells. Blood 2004;104:3318–25.
27. Ramiro AR, Jankovic M, Eisenreich T, et al. AID is required for c-myc/IgH chromosome translocations *in vivo*. Cell 2004;118:431–8.
28. Park SS, Shaffer AL, Kim JS, et al. Insertion of Myc into Igh accelerates peritoneal plasmacytomas in mice. Cancer Res 2005;65:7644–52.
29. Ramiro AR, Jankovic M, Callen E, et al. Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature 2006;440:105–9.
30. Pham P, Bransteitter R, Goodman MF. Reward versus risk: DNA cytidine deaminases triggering immunity and disease. Biochemistry 2005;44:2703–15.
31. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev 2004;18:1618–29.
32. Maizels N. Dynamic roles for G4 DNA in the biology of eukaryotic cells. Nat Struct Mol Biol 2006;13:1055–9.
33. Phan AT, Kuryavyi V, Patel DJ. DNA architecture: from G to Z. Curr Opin Struct Biol 2006;16:288–98.
34. Duquette ML, Pham P, Goodman MF, Maizels N. AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. Oncogene 2005;24:5791–8.
35. Larson ED, Duquette ML, Cummings WJ, Streiff RJ, Maizels N. MutSa binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions. Curr Biol 2005;15:470–4.
36. Huber MD, Duquette ML, Shiels JC, Maizels N. A conserved G4 DNA binding domain in RecQ family helicases. J Mol Biol 2006;358:1071–80.
37. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503–11.
38. Preudhomme C, Roumier C, Hildebrand MP, et al. Nonrandom 4p13 rearrangements of the RhoH/TTF gene, encoding a GTP-binding protein, in non-Hodgkin's lymphoma and multiple myeloma. Oncogene 2000;19:2023–32.
39. Muschen M, Re D, Jungnickel B, et al. Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. J Exp Med 2000;192:1833–40.
40. Gordon MS, Kanegai CM, Doerr JR, Wall R. Somatic hypermutation of the B cell receptor genes B29 (Igβ, CD79b) and mb1 (Igα, CD79a). Proc Natl Acad Sci U S A 2003;100:4126–31.
41. Landowski TH, Qu N, Buyuksal I, Painter JS, Dalton WS. Mutations in the Fas antigen in patients with multiple myeloma. Blood 1997;90:4266–70.
42. Payelle-Brogard B, Magnac C, Mauro FR, Mandelli F, Dighiero G. Analysis of the B-cell receptor B29 (CD79b) gene in familial chronic lymphocytic leukemia. Blood 1999;94:3516–22.
43. Gordon MS, Kato RM, Lansigan F, Thompson AA, Wall R, Rawlings DJ. Aberrant B cell receptor signaling from B29 (Igβ, CD79b) gene mutations of chronic lymphocytic leukemia B cells. Proc Natl Acad Sci U S A 2000;97:5504–9.
44. Shen HM, Michael N, Kim N, Storb U. The TATA binding protein, c-Myc and survivin genes are not somatically hypermutated, while Ig and BCL6 genes are hypermutated in human memory B cells. Int Immunol 2000;12:1085–93.
45. Muller JR, Janz S, Potter M. Differences between Burkitt's lymphomas and mouse plasmacytomas in the immunoglobulin heavy chain/c-myc recombinations that occur in their chromosomal translocations. Cancer Res 1995;55:5012–8.
46. Janz S. Myc translocations in B cell and plasma cell neoplasms. DNA Repair (Amst) 2006;5:1213–24.
47. Epeldegui M, Hung YP, McQuay A, Ambinder RF, Martinez-Maza O. Infection of human B cells with Epstein-Barr virus results in the expression of somatic hypermutation-inducing molecules and in the accrual of oncogene mutations. Mol Immunol 2007;44:934–42.
48. Eddy J, Maizels N. Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Res 2006;34:3887–96.
49. Sun H, Karow JK, Hickson ID, Maizels N. The Bloom's syndrome helicase unwinds G4 DNA. J Biol Chem 1998;273:27587–92.
50. Mohaghegh P, Karow JK, Brosh Jr RM, Jr., Bohr VA, Hickson ID. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. Nucleic Acids Res 2001;29:2843–9.