

The Strength of an Ig Switch Region Is Determined by Its Ability to Drive R Loop Formation and Its Number of WGCW Sites

Zheng Z. Zhang,^{1,2,3} Nicholas R. Pannunzio,^{1,2,3} Li Han,⁴ Chih-Lin Hsieh,^{1,2,3} Kefei Yu,^{4,*} and Michael R. Lieber^{1,2,3,*}

¹USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90033, USA

²Molecular and Computational Biology Section, Department of Biological Sciences, USC Dornsife, University of Southern California, Los Angeles, CA 90033, USA

³Departments of Pathology, Biochemistry & Molecular Biology, Molecular Microbiology & Immunology, and Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

⁴Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA

*Correspondence: yuke@msu.edu (K.Y.), lieber@usc.edu (M.R.L.)

<http://dx.doi.org/10.1016/j.celrep.2014.06.021>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

SUMMARY

R loops exist at the murine IgH switch regions and possibly other locations, but their functional importance is unclear. In biochemical systems, R loop initiation requires DNA sequence regions containing clusters of G nucleotides, but cellular studies have not been done. Here, we vary the G-clustering, total switch region length, and the number of target sites (WGCW sites for the activation-induced deaminase) at synthetic switch regions in a murine B cell line to determine the effect on class switch recombination (CSR). G-clusters increase CSR regardless of their immediate proximity to the WGCW sites. This increase is accompanied by an increase in R loop formation. CSR efficiency correlates better with the absolute number of WGCW sites in the switch region rather than the total switch region length or density of WGCW sites. Thus, the overall strength of the switch region depends on G-clusters, which initiate R loop formation, and on the number of WGCW sites.

INTRODUCTION

Immunoglobulin heavy chains (IgH) have a variable exon assembled by V(D)J recombination and constant domain exons (Chaudhuri et al., 2007; Hackney et al., 2009; Figure S1A). The IgH constant domains can be changed at the DNA level by Ig class switch recombination (CSR), and this allows the same antigen binding pocket to be directed to different locations in the organism and to elicit different immune responses. The initial immunoglobulin, IgM, is encoded by the C μ constant domain exons that are immediately downstream of the V(D)J exon. However, other IgH isotypes, including IgG, IgA, and IgE, are located over a region of several hundred kilobases downstream. Each set of IgH isotype exons for the constant region is preceded by a sterile transcript promoter and a switch region that is repetitive and rich in WGCW sites.

CSR is particularly efficient and well developed in mammals, where a majority of the antibody in the blood is generated from an IgH locus that has undergone CSR (Murphy, 2011). Efficient CSR may be essential for organisms with well-developed lungs to transport IgA into the airways to minimize pulmonary infections from viruses, bacteria, and fungi. The rapid IgH switch from IgM to IgA requires DNA recombination at the IgH locus. Rapid IgH isotype switch is also important in the gastrointestinal and urogenital tracts in mammals (Murphy, 2011).

AID is a cytidine deaminase that initiates both somatic hypermutation (SHM) and CSR (Muramatsu et al., 1999, 2007). AID requires single-stranded DNA (ssDNA) to deaminate C to U (Bransteitter et al., 2003; Pham et al., 2011; Yu et al., 2004). How ssDNA arises for SHM is not clear but it may be a simple consequence of transcription. Such a mechanism of generating ssDNA would be inefficient, and this may explain why SHM occurs relatively slowly, over a period of weeks in a typical immune response (Murphy, 2011). In contrast, CSR occurs much faster over a period of days, typically 48 to 72 hr, using the same AID enzyme. A preferred site of action by AID is the WGCW motif, shown both biochemically (Bransteitter et al., 2003; Yu et al., 2004) and within the switch regions in the genome (Han et al., 2011).

The basis for the difference in efficiency between mammalian CSR and SHM has not been clear. One difference may be that R loops form at mammalian IgH switch region DNA when B cells are activated for switch region transcription via cytokine stimulation (Yu et al., 2003; Yu and Lieber, 2003). The R loops form when transcription extends through the repetitive switch regions, and the nascent RNA threads back (anneals) onto the template DNA strand before the nontemplate DNA and template DNA strands can reanneal (Roy and Lieber, 2009; Roy et al., 2008, 2010). Thus, an R loop is a three-stranded nucleic acid structure consisting of the two DNA strands of the normal duplex, but with the newly transcribed RNA base paired as a duplex with the template DNA strand. Therefore, this leaves the nontemplate DNA strand entirely single-stranded (Roy and Lieber, 2009; Roy et al., 2008, 2010; Yu et al., 2003; Yu and Lieber, 2003). As mentioned, a promoter is located upstream of each Ig class switch region, and yet no protein is produced from this

transcript; hence, this is called a sterile transcript (Chaudhuri et al., 2007; Stavnezer et al., 2008; Wabl and Steinberg, 1996). The function of sterile transcripts has been unclear. We have proposed that one major function is to generate R loops because these would provide a stable site of ssDNA for AID action in a very efficient manner, especially for organisms that require rapid IgH switching (Yu et al., 2003; Yu and Lieber, 2003). Independent genome-wide methods to document ssDNA confirm that Ig switch regions are among the most single-stranded regions in the genomes of mammalian B cells (our analysis of Kouzine et al., 2013, K.L. and M.R.L., unpublished data). It is readily apparent that an R loop makes the nontemplate DNA strand available in ssDNA form for AID (Yu et al., 2005). Regarding the template DNA strand, we have described elsewhere how RNase H action can create zones of ssDNA on this remaining template strand (Roy and Lieber, 2009; Roy et al., 2008, 2010; Yu et al., 2003; Yu and Lieber, 2003).

In biochemical studies with purified DNA and prokaryotic RNA polymerases, we showed that G-clusters (consecutive Gs) are important for the newly transcribed RNA to thread back onto the template DNA strand to form an R loop (Roy and Lieber, 2009; Roy et al., 2008, 2010). This is almost indistinguishable from the formation of R loops at DNA replication origins in prokaryotes, where a G-cluster of a newly transcribed RNA must anneal to the template DNA strand to initiate DNA replication (Masukata and Tomizawa, 1990). Factors that favor annealing of the RNA to the DNA template, such as negative superhelicity, nontemplate DNA strand nicks, and proximity of the G-cluster to the 5' end of the RNA all favor more efficient R loop formation (Baaklini et al., 2008; Broccoli et al., 2004; Drolet et al., 2003; Roy and Lieber, 2009; Roy et al., 2008, 2010). For switch sequences, we term the G-cluster-rich region responsible for initiating the R loops the "R loop initiation zone" or RIZ. Once initiated in the RIZ, we find that the R loop can extend downstream of the RIZ into a region with no G-clusters but in which the G-density is relatively high (35%–50% G). We call this region the "R loop elongation zone" or REZ.

Despite having shown that R loops form in the genomes of mammalian B cells at Ig switch regions, the relationship between G-clusters and R loop formation within cells has not been tested. This link is a critical one, especially because recent reports have proposed R loops in key biological processes far outside of CSR (Balk et al., 2013; El Hage et al., 2010; Mischo et al., 2011; Nakama et al., 2012; Pfeiffer et al., 2013; Skourti-Stathaki et al., 2011; Stirling et al., 2012; Sun et al., 2013; Wahba et al., 2011; Yang et al., 2014). Genome biologists have highlighted Ig class switch-like sequences at other locations in the genome for the possibility that they have a propensity for DNA breakage; the hypothesis is that such breaks would direct chromosomal changes during evolutionary speciation (Bailey and Eichler, 2006). Other studies claim that half the genes in the genome are predicted to form R loops (Wongsurawat et al., 2012). Some studies claim that R loops exist at most CpG island promoters (Ginno et al., 2012). R loops have been associated with genome instability (Li et al., 2005; Stirling et al., 2012; Wahba et al., 2011). Despite these additional roles, none of these studies has defined what are the essential DNA

sequence elements or features needed to cause R loop formation in the genome.

Here we examine the effects of the following on CSR: (1) G-clusters interspersed among WGCW sites; (2) G-clusters physically separated from the WGCW sites; (3) WGCW site number and density; and (4) total switch region length (determined by the number of switch repeats). We find that CSR depends on the presence of G-clusters, which are the key determinant of R loop initiation, regardless of whether the G-clusters are interspersed within the switch repeats (as they are naturally in mammalian switch regions) or restricted to the upstream boundary in an RIZ. Once an R loop is initiated, the R loop final length determines how much ssDNA is exposed for AID action. AID preferentially acts at WGCW sites in ssDNA. Within the total switch region lengths studied here, we find that the total number of WGCW sites is more important than their density or the DNA length over which a given number of WGCW sites is distributed. These findings are important because they define what DNA sequence features are needed for efficient Ig CSR.

RESULTS

Experimental Strategy and Rationale

To study the factors important for R loop formation and for Ig CSR, we used a system that allows us to exchange out the genomic switch region and replace it with any desired sequence (Han et al., 2011). The mouse B cell line, CH12F3.2a, is able to specifically and efficiently switch to IgA upon cytokine stimulation (hereafter called the cellular system). Using this cellular assay system, the endogenous $S\alpha$ locus in CH12F3.2a was replaced with a positive-negative selection cassette (Puro Δ TK, provides puromycin resistance and ganciclovir sensitivity) flanked by two different loxP sites. The resulting cell line is called 1F7 (Han et al., 2011; Figure S1B). The sequence of interest (Table S1) is cloned into an exchange vector with the same loxP sites as 1F7 cells. This exchange vector is cotransfected into 1F7 cells along with a Cre-expressing vector. Cre mediates recombination between the exchange vector and the 1F7 chromosome at the corresponding loxP sites. The successful replacement of the selection cassette by the sequence of interest gives rise to ganciclovir-resistant clones, which are screened with a puromycin sensitivity test and PCR analysis for all clones (Table S2). Selected clones are confirmed with Southern blot.

Here, we tested parameters that may be important for IgH CSR using the cellular assay system. CSR as a function of G-clusters number was a focus of primary interest. Given that efficient CSR requires WGCW sites (Han et al., 2011), we tested synthetic switch regions that had two fundamentally different configurations: (1) G-clusters interspersed with WGCW sites (designated the interspersed configuration) and (2) G-clusters in an upstream RIZ, followed by WGCW sites in a downstream REZ, which has no G clusters but has a 45% G-density (designated the RIZ/REZ configuration).

Naturally occurring mammalian switch regions consist of repetitive units of 25–80 base pairs (bp), and the natural mammalian $S\alpha$ switch region has an 80 bp repeat (Dunnick et al., 1993;

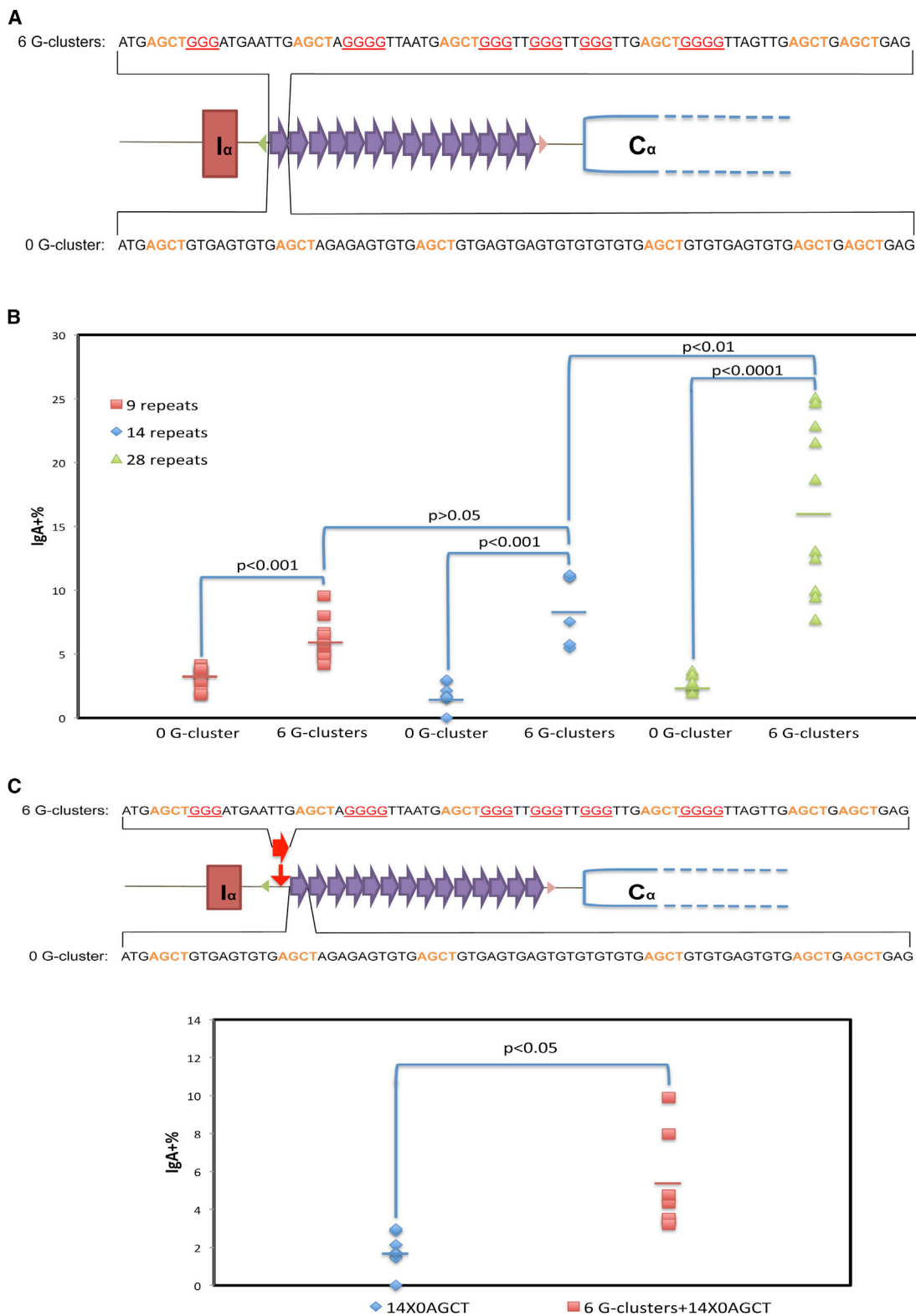


Figure 1. Effect of G-Clusters Interspersed within Switch Repeat Unit

(A) Construction of synthetic switch regions and the genomic configuration of $S\alpha$ locus (Han et al., 2011). Each 80 bp switch repeat unit (with six G-clusters in the DNA sequence shown for the duplex above the diagram or with no G-clusters for the duplex below the diagram) was self-ligated and the products with the desired (legend continued on next page)

Snapper and Finkelman, 1999). The naturally occurring switch regions have an interspersed configuration (i.e., many G-clusters and WGCW sites are mixed together within each repeat). For the synthetic switch regions studied here, we varied the number of 80 bp repeats between 4 and 28 (for a total switch region length of 320–2,240 bp). For the interspersed configurations, we varied the G-clusters number in the repeat units (but not the WGCW number in each repeat unit). For the RIZ/REZ configurations, we varied the G-cluster number in the RIZ and the WGCW number in each repeat of the REZ.

Clusters of G Nucleotides Increase CSR

Biochemical studies showed that G-clusters are able to promote R loop formation (Roy and Lieber, 2009; Roy et al., 2008, 2010). In the genome of activated B cells in culture, we determined that the R loops typically begin near G-clusters and end when the G-density falls below 30% (Huang et al., 2006, 2007). However, a formal study of CSR as a function of G-cluster number has not been done in a cellular assay system. This aspect of the R loop hypothesis is critical for the Ig class switch field to evaluate the extent to which R loop formation is central to the high efficiency of mammalian CSR.

We created an 80 bp repeat unit with six G-clusters, and constructed exchange vectors with increasing numbers of these repeat units. The cellular clones with these sequences of interest were selected as described (Han et al., 2011; Figure S1B). At least five clones were picked for CSR efficiency tests. As a control, a different 80 bp repeat unit with exactly the same nucleotide composition, but no G-clusters, was constructed (Figure 1A).

We found that with nine switch repeats having an interspersed configuration, G-clusters promote CSR by almost 2-fold (Figure 1B). G-clusters appear to slightly inhibit germline transcription (GLT; Figures S3 and S4A). When the repeat number increases from 9 to 14, CSR increases another 1.5-fold (borderline significance, $p = 0.07$), whereas the CSR efficiency in cells without G-clusters remains at a low level.

We wanted to see whether CSR increases further with more G-clusters, so we constructed cellular clones with 28 repeats containing either zero or six G-clusters interspersed within each switch repeat. CSR in cells with G-clusters increases significantly by another 2-fold over the cells without G-clusters for a switch region of 14 repeats, and the increase is even more significant in cells with nine repeats (Figure 1B). Importantly, CSR is inefficient in cells without G-clusters even with 28 switch repeat units.

In all three pairs of genomic configurations, the presence of G-clusters in the synthetic switch repeats significantly increases the CSR ability. Based on this, we conclude that G-clusters are

important for efficient CSR. Without G-clusters, CSR is inefficient, regardless of the number of switch repeats and, hence, regardless of the total switch region length. This clearly indicates that the presence of G-clusters is critical to promote CSR, and the total switch region length alone is insufficient to promote CSR. As we examine in more detail below, but is apparent here, if G-clusters are present, CSR also increases as the number of switch repeats (total switch region length) increases (Figure 1B).

G-Clusters Can Convert Less Efficient Class Switch Regions to More Efficient Ones

We wondered whether we could convert a low-efficiency synthetic switch region to a higher one by adding a short region of G-clustered DNA immediately upstream of a synthetic switch region without G-clusters. We inserted the same 80 bp unit with six G-clusters immediately upstream of the 14 switch repeats without any G-clusters (Figure 1C). We constructed and analyzed the cell clones as described above.

We found that, with just one 80 bp unit containing six G-clusters upstream of the repetitive sequence of 14 repeats lacking G-clusters, the CSR efficiency increases by almost 3-fold (Figure 1C). Therefore, the increase in CSR is not simply due to the longer DNA sequence (insertion of the 80 bp sequence) because the cells with 28 repeats without G-clusters have relatively low CSR (Figure 1B). This suggests that the G-clusters play an important stimulatory role for switch region strength for initiating R loop formation.

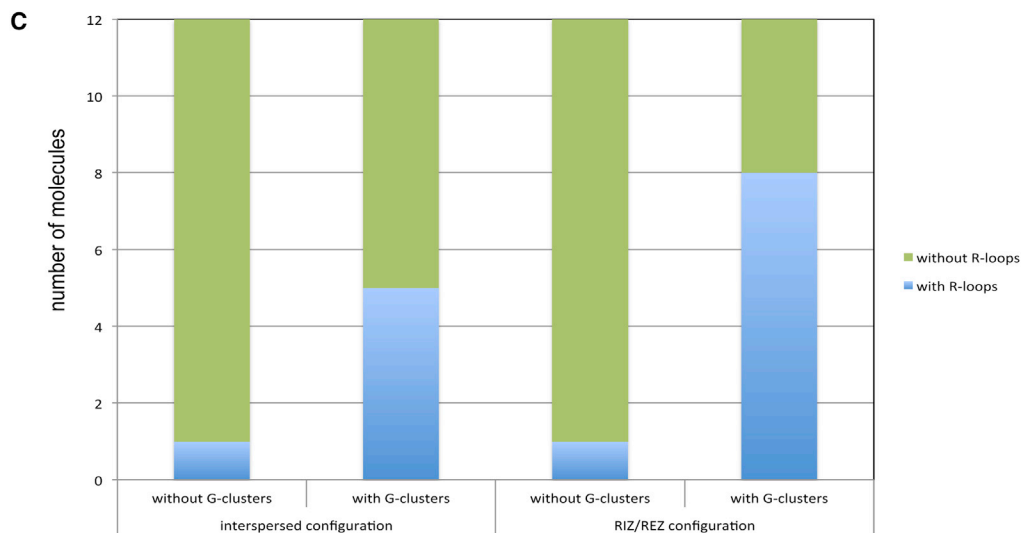
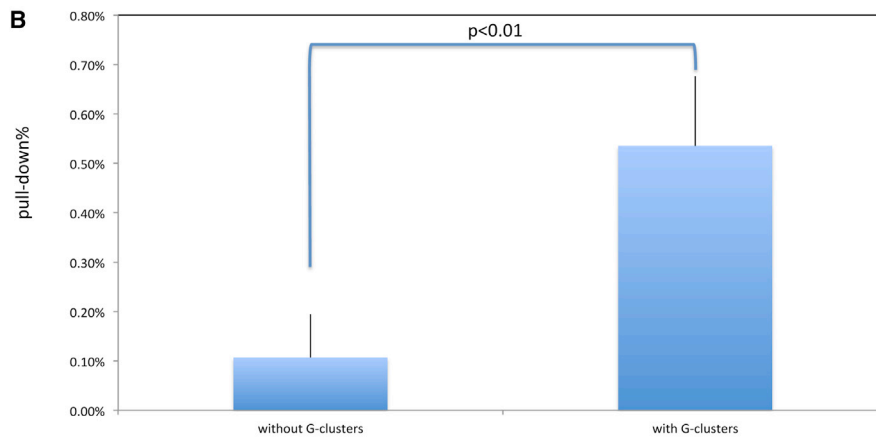
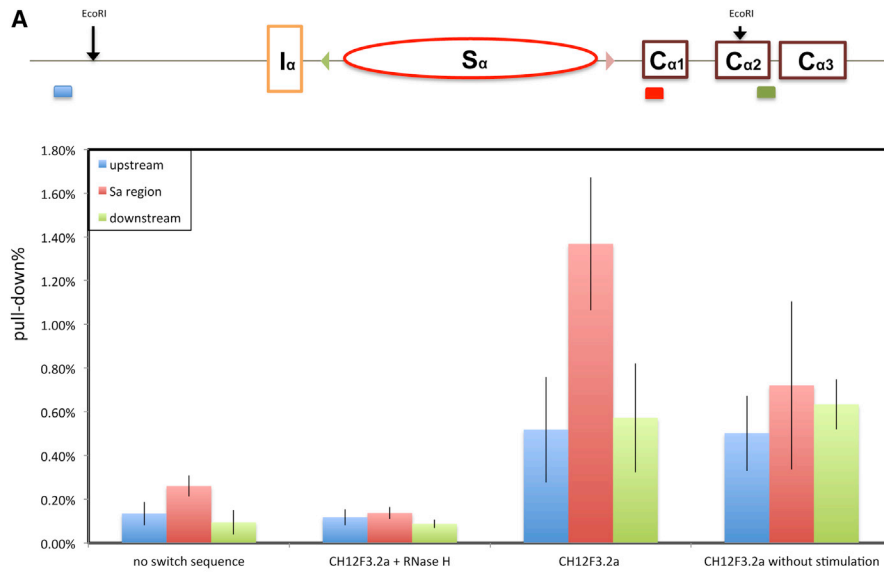
G-Clusters near the Beginning of the Class Switch Region Have a Large Effect on CSR

To investigate how G-clusters affect CSR further, we fully separated G-clusters from the WGCW sites by using an RIZ/REZ configuration (see Introduction). The five G-clusters are only located at the beginning of our synthetic switch regions (the RIZ), whereas the repetitive region (REZ) has nine repeats with four AGCT sites, but no G-clusters. The REZ is also a region with 45% G-density to support R loop elongation, once an R loop is initiated within the RIZ. As a negative RIZ control, a piece of DNA with the same length, but no G-clusters were used (Figure 2A). With G-clusters within the RIZ, CSR increases significantly by almost 2-fold (Figure 2B). This further indicates that G-clusters increase CSR even when G-clusters are only present at the beginning of a switch region.

G-Clusters Increase R Loop Formation within Cells

An antibody has been described that appears to bind RNA:DNA hybrids preferentially over DNA:DNA or RNA:RNA of the same sequence (Hu et al., 2006; Phillips et al., 2013), and we used this antibody to detect R loops (Ginno et al., 2012;

number of switch repeats were inserted into the exchange vector. An exchange vector and a Cre-expression vector were cotransfected into 1F7 cells. Triangles represent different loxP sites, and each oval represents one 80 bp repeat unit. Six to ten independent cellular clones were made for each switch construct. (B) Fluorescence-activated cell sorting (FACS) analysis of CSR is shown. Each data symbol represents an independent clone. The clones are organized in groups of two, based on the number of repeat units shown in the key on the right. The number on the bottom indicates the number of G-clusters within each repeat. The bars represent the mean of each group. The Student's *t* test was used to analyze the effect of G-clusters. (C) One 80 bp repeat with six G-clusters (the red oval in the top sequence) was cloned immediately upstream of the fourteen 80 bp repeats with no G-clusters (purple ovals that have the DNA sequence shown under the diagram at the top). Seven independent cellular clones of each switch construct were analyzed with FACS. The horizontal colored bars represent the mean of each group.



(legend on next page)

units, CSR correlates with total switch region length (Figure S2), as others have found using a murine knockin system (Zarrin et al., 2005).

A longer switch region could affect CSR in either of two ways. Greater length could (1) provide more WGCW sites, or (2) permit longer R loops to form, thereby providing a greater length of ssDNA. In the RIZ/REZ configuration, R loops merely initiate within the RIZ, whereas the REZ, with its high G-density, permits elongation of the R loop to the end of the switch region (Effect 2). In considering the contribution of WGCW sites to switch region strength (Effect 1), there are three related factors: the number of WGCW sites, the density of WGCW sites, and the total switch region length. The total number of WGCW sites equals the number of WGCW sites within each repeat unit (the density of WGCW sites) times the repeat number (total switch region length divided by 80 bp).

Our assay system allows us to determine which of these factors contribute to CSR efficiency. Only one of the three factors can remain constant while the other two vary; therefore, three different scenarios were devised, and switching efficiency was measured.

Constant WGCW Density but Varied Total Switch Region Length and Total WGCW Site Number

We first examined the effect of switch region length and total WGCW site number when the G-clusters are only located at the beginning of the switch region (RIZ/REZ configuration). In the earlier biochemical studies, we showed that a high G-density in the REZ could maintain R loops (Roy and Lieber, 2009; Roy et al., 2008, 2010). If this applies within cells, then with a few G-clusters at the beginning of a switch region in an RIZ and a high overall G-density in the REZ, longer switch regions would be expected to generate higher CSR. We constructed two cell lines with a single 23 bp RIZ containing three G-clusters and different total REZ lengths (four versus nine 80 bp synthetic $S\alpha$ repeats containing four AGCT sites and alternating G-nucleotides in each repeat, but no G-clusters). We found that nine repeats result in significantly higher CSR compared to only four repeats (almost 3-fold; Figure 4A).

To further evaluate this, three constructs with the same REZ repeat unit were created, but with progressively greater numbers of switch repeats: four, six, and nine repeats, each repeat containing six AGCT sites (Figure 4B). As in the case above (Figure 4A), we observe an increase in CSR with increased total switch region length and total WGCW number (Figure 4B).

Therefore, we have shown that with the same repeat unit (i.e., density of WGCW sites), CSR can increase if the total switch

region length and total number of WGCW sites increase, suggesting that either total switch region length or the total numbers of WGCW sites within the region are important for switching.

Constant Total Switch Region Length but Varied WGCW Density and Total WGCW Site Number

Next, we wanted to determine if WGCW number was the most important aspect of Effect 1. We kept the RIZ constant and varied the WGCW number within each 80 bp REZ repeat. Each switch region in this comparison has nine REZ repeats (Figure 5A). Cells were screened and analyzed as previously described. When there are only two WGCW sites within each 80 bp repeat, we find that the average CSR is 4.2% (Figure 5A). If we increase the number of WGCW sites within each repeat to four, then the average CSR doubles to 8.4%. When we increase the number of WGCW sites to six, the average CSR increases to 13.4% (Figure 5A). Hence, we found that with more WGCW sites within each repeat unit, CSR increases significantly. A plot of CSR against the total number of WGCW sites within each 80 bp repeat (or across the entire switch region) shows a strong positive correlation (Figure 5B). Thus, when total switch region length is held constant, CSR increases as a function of the number of WGCW sites, strongly suggesting that the number of WGCW sites is important but total switch region length is not.

Constant Total Number of WGCW Sites but Variable WGCW Density and Switch Region Length

Having ruled out WGCW density and total switch region length, we wanted to further evaluate whether CSR primarily depends on total WGCW number (which, for identical repeats, as in our case, is a multiple of the WGCW site number in each repeat unit). We constructed a pair of cell lines with the same RIZ and the same total number of WGCW sites (Figure 6A). Within each synthetic switch region, we designed six repeats containing six AGCT sites each, for 36 WGCW sites. We compared this to nine repeats containing four AGCT sites each, for 36 WGCW sites. Even though the nine repeat switch region has a longer total switch region length ($9 \times 80 = 720$ bp versus $6 \times 80 = 420$ bp), we found no significant difference in CSR (9.2% versus 8.4%).

To further evaluate this point, we made two additional cell lines with the same total number of WGCW sites as one another, but with different WGCW densities and total switch region lengths. Either four repeats with four WGCW sites (16 total WGCW site number) or eight repeats with two WGCW sites (16 total WGCW site number) were integrated at the $S\alpha$ locus. We found no difference in CSR (Figure 6B). Therefore, the total number of

Figure 3. Effect of G-Clusters on R Loop Formation within Cells

(A) IP with the S9.6 antibody was performed on wild-type CH12F3.2a cells and one mutant with no switch sequences at the endogenous $S\alpha$ locus. The amplicons of qPCR are shown in different colors, corresponding to the bars on the bottom. Background signals from mock samples with no antibody were subtracted. Values were normalized to the total input DNA to calculate the pull-down percentage. Three independent IP experiments were performed for each cell line. Error bars represent SEM.

(B) IP was performed on cellular constructs containing 14 repeats with and without G-clusters (the middle group in Figure 1B) with the S9.6 antibody. Four independent IP experiments were performed for each cellular construct. The pull-down percentage of the mutant with no switch sequence was subtracted as background. The Student's *t* test was used to analyze the effect of G-clusters. Error bars represent SEM.

(C) Bisulfite sequencing analysis was done on the cellular constructs containing 14 repeats with and without G-clusters, or the cellular constructs in Figure 2B with and without G-clusters. The y axis gives the number of IgH alpha locus alleles, out of 12 examined (duplicates were eliminated), that have long (>60 bp) regions of C to T conversion due to bisulfite. Actual patterns of conversion along each allele are shown in Figures S5 and S6.

A WGCW Density X Total Switch Region Length = Total WGCW Site Number

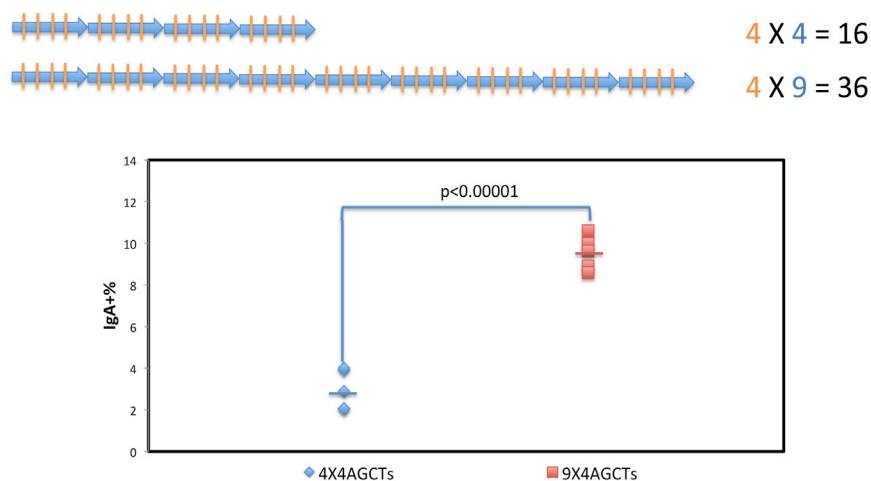
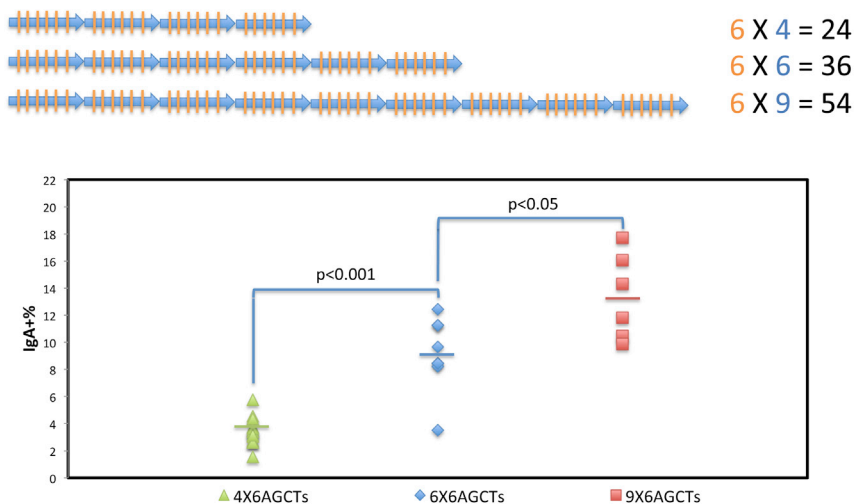


Figure 4. Constant WGCW Density but Varied Total Switch Region Length and Total WGCW Site Number

(A) Each blue arrow and orange bar represents one repeat unit and one AGCT site, respectively. An REZ consisting of four or nine 80 bp repeats with four AGCT sites each was constructed immediately downstream of an RIZ with three G-clusters. Each dot represents an independent clone. Five independent cellular clones of each cellular construct were analyzed with FACS. The bars represent the mean of each group. The Student's t test was used to analyze the effect of switch region length. In the equation at the top, the underlined parameter is kept constant, and the remaining two parameters vary (this also applies to Figures 5 and 6).

(B) An REZ of four, six, or nine 80 bp repeats with six AGCT sites was constructed immediately downstream of an RIZ containing five G-clusters. Six to ten independent cellular clones of each cellular construct were analyzed with FACS.

B WGCW Density X Total Switch Region Length = Total WGCW Site Number



C deamination that begins the DNA breakage process. But these WGCW sites must be within ssDNA for AID to recognize them as substrates (Branstetter et al., 2003; Pham et al., 2011; Yu et al., 2004). For SHM, transcription might inefficiently provide ssDNA to permit the relatively slow process of SHM to occur over weeks during an immune response. But CSR occurs typically within a few days in mammals. We have proposed that some of the more rapid AID action in CSR in mammals is due to the high G-clustering and G-density that is the distinctive feature of all mammalian IgH switch regions (Yu et al., 2003; Yu and Lieber, 2003) and not present in organisms that evolved CSR earlier, such as avians and amphibians (Hackney et al., 2009; Zhu et al.,

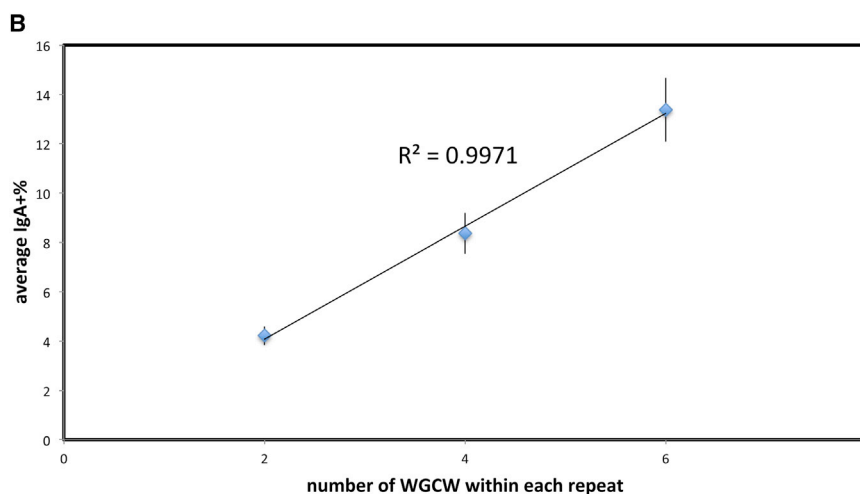
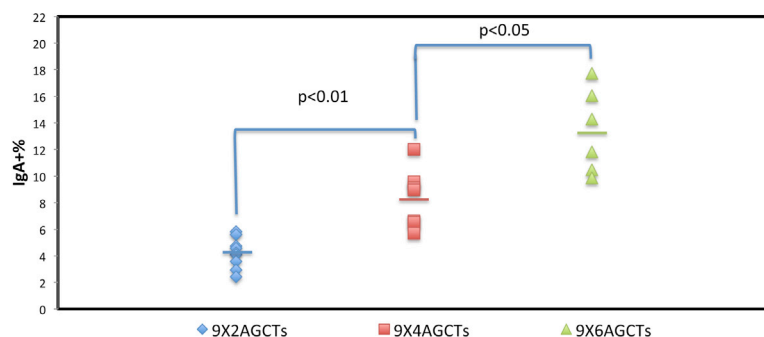
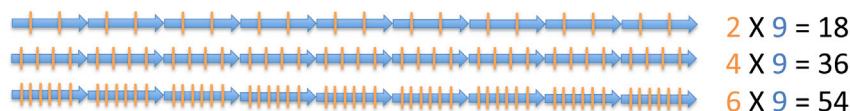
2012). These considerations lead to the following questions. First, are the G-clusters that we define in our biochemical studies as RIZs also important for cellular CSR? Second, can the G-clusters of the RIZ be separated from the G-dense REZ, or must these be interspersed, as they have evolved in mammals? Third, can the RIZs be separated from the WGCW sites, which are located in the REZ in the RIZ/REZ configuration? Fourth, is total length of the switch region more important to provide a longer R loop (i.e., ssDNA) or to provide a larger number of WGCW sites? Fifth, is it the total number of WGCW sites that is important; or is it the density of these sites or the length over which the WGCW sites are distributed that is the key determinant? Sixth, how do the answers to these questions relate to the function of IgH switch regions, to their evolution, and to earlier vertebrate switch regions that do not have G-clusters or do not have a high G-density?

WGCW sites, rather than their density or the total switch region length determines CSR (Effect 1).
Regarding R loop length (Effect 2), whereas R loop initiation (RIZ) is important, the total length of the REZ (Figures 5 and 6), and hence, the total switch region length, are not important. We found that with the same RIZ, the R loop distribution is similar among REZs that we have tested (Table S5). Hence, within the total switch region lengths studied here, as long as numerous WGCW sites are present within an R loop zone, the length over which that R loop is distributed is not critical (Figure 7).

DISCUSSION

Here we define the major factors that determine switch region strength. AID requires WGCW sites to efficiently create the

A $\text{WGCW Density} \times \text{Total Switch Region Length} = \text{Total WGCW Site Number}$



G-Clusters and Switch Region Strength

Biochemical data had shown that G-clusters are extremely important to initiate the thread-back of the RNA onto the DNA template, which thereby initiates R loop formation (Roy and Lieber, 2009; Roy et al., 2008, 2010). Here, we have demonstrated that G-clusters are a very important determinant of whether a switch region leads to CSR, predominantly by causing R loop formation. The G-clusters can be located at the leading edge of the switch region, thereby leading to R loop formation that can allow downstream WGCW sites to be within a zone of ssDNA (and therefore be substrates for AID). On the other hand, the G-clusters can be interspersed among the WGCW sites, as they are with all mammalian switch regions. We find that both of these configurations can support CSR.

We have not extensively tested for variation in the number of Gs in each G-cluster. However, most of our G-clusters contain three Gs. We have noted previously that any clustering of Gs appears to be sufficient, including GGG and GGGG. GGG-clusters are not likely to form stable G-quartets (Sinden, 1994). We

Figure 5. Constant Total Switch Region Length but Varied WGCW Density and Total WGCW Site Number

(A) Each blue arrow and orange bar represents one repeat unit and one AGCT site, respectively. REZ with the same number of 80 bp repeat units with different numbers of AGCT sites in the repeats were constructed immediately downstream of an RIZ with five G-clusters. Each data symbol represents an independent clone. Six to nine independent cellular clones of each cellular construct were analyzed with FACS. The bars represent the mean of each group. The Student's t test was used to analyze the effect of WGCW density.

(B) Average CSR values of each group were plotted against their WGCW density. Error bars represent SEM.

Each blue arrow and orange bar represents one repeat unit and one AGCT site, respectively.

have previously noted that G-quartet formation is not necessary for R loop formation in biochemical systems, where we have formally ruled them out in R loop formation (Roy and Lieber, 2009; Roy et al., 2008, 2010). G-quartet formation at switch regions also is unlikely within B cells for additional reasons that we have described elsewhere (Huang et al., 2007). We acknowledge that the G-clusters may contribute to Ig SCR in additional ways other than R loop formation.

Number of WGCW Sites as a Determinant of Switch Region Strength

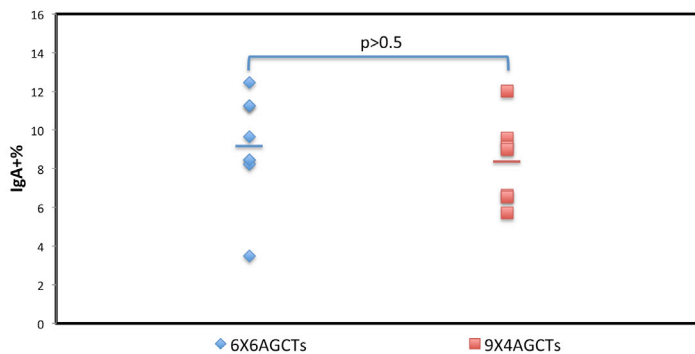
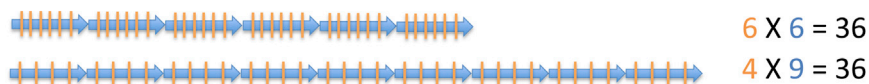
Previous work using this system showed that WGCW is the key DNA sequence

required for CSR (Han et al., 2011). Building on this observation, we have varied the number and density of WGCW sites and have found that the number of WGCW sites is more critical than the density of these sites, at least for total switch region lengths of less than 1 kb.

Mammalian switch regions have evolved to be long. The longer switch regions typically provide more WGCW sites and more G-clusters, but the total switch region length itself (independent of WGCW sites and G-clusters) does not appear to be critical for CSR efficiency.

For SHM, WGCW sites must be in a ssDNA state, at least transiently, to be a substrate for AID. Transcription through random DNA (containing few or no G-clusters) appears to be able to induce such transient ssDNA states, though inefficiently. Once AID causes a U:G mismatch at a WGCW site, this is converted to a nick via UNG and APE (Di Noia and Neuberger, 2007; Masani et al., 2013). In biochemical systems, R loops can form when a nick is present in the nontemplate strand of the DNA, even when there are no G-clusters

A WGCW Density X Total Switch Region Length = Total WGCW Site Number



B WGCW Density X Total Switch Region Length = Total WGCW Site Number

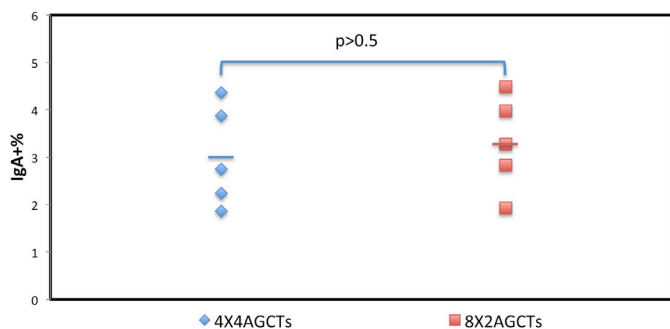


Figure 6. Constant Total Number of WGCW Sites but Variable WGCW Density and Switch Region Length

(A) Each blue arrow and orange bar represents one repeat unit and one AGCT site, respectively. FACS analysis of CSR is shown. The same total number of WGCW sites was constructed immediately downstream of an RIZ containing five G-clusters, as shown on the bottom left. Both switch region length and WGCW density are different. Six repeats of 80 bp oligonucleotides with six AGCT sites or nine repeats of 80 bp oligonucleotides with four AGCT sites were constructed downstream of an RIZ containing five G-clusters.

(B) Four repeats of 80 bp oligonucleotides with four AGCT sites or eight repeats of 80 bp oligonucleotides with two AGCT sites were constructed downstream of an RIZ containing six G-clusters. Each data symbol represents an independent clone. Five to eight independent cellular clones are shown in each group. The color bars represent the mean of each group. The Student's t test was used to analyze the effect of WGCW number.

Evolution of G-Clusters for CSR Efficiency

Given that both interspersed G-clusters and a separated zone of G-clusters (RIZ) are sufficient, why did the interspersed configuration evolve for mammalian switch regions? This may simply be a consequence of the mechanism by which switch regions became repetitive. DNA expands over evolutionary time via unequal sister chromatid exchange (Smith, 1974, 1976; Strachen and Read, 1999). Such an expansion requires that each repeat unit be nearly identical. This means that each 25 to 80 bp repeat unit in a mammalian switch region would harbor all of the key elements for CSR: WGCW sites, G-clusters, and an overall high G-density to permit R loop elongation since the first

repeat unit had to contain all of these elements. If each repeat unit carries its own G-clusters, then an interspersed configuration would be the result, once all of the repeats are assembled end to end. After sufficient evolutionary expansion of the switch regions, sequence divergence of the repeats from one another may then be selected in order to minimize deletion of the repeats by homologous recombination within a switch region. Therefore, the repeats became not quite so identical over evolutionary time, since they only need to retain the three key mechanistic features: a high number of WGCW sites, G-clusters, and a high G-density.

and the G-density is low (Roy et al., 2010). This may be the basis for how AID can act during SHM, despite a lack of G-clusters. AID acts more rapidly in CSR than in SHM. For CSR studied here, the total number of WGCW sites is critical, regardless of the length over which the R loop extends. Hence, a given number of WGCW sites yields a certain level of CSR, regardless of whether the WGCW sites are spread over twice the length of DNA in the switch region. This indicates that the number of WGCW lesion sites is the rate-limiting factor for CSR, as long as they are located within the R loop zone. The distance between WGCW sites (WGCW density) does not seem to be critical, and increased total length of the switch region is important only to the extent that it provides a greater number of WGCW sites.

Transcription through R Loop Regions
We find that G-clusters increase CSR but typically reduce transcription. This is precisely what we see when studying

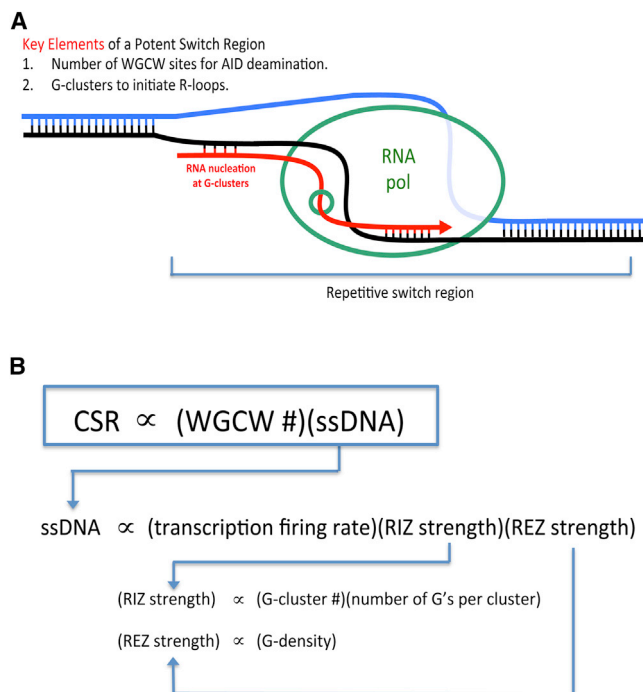


Figure 7. Model of CSR

(A) Features of typical Ig class switch recombination regions. AID is the key enzyme that initiates CSR. WGCW sites are the preferred DNA sequence motif targets of AID, but AID only recognizes cytosine deamination sites within ssDNA. The G-clusters initiate R loops in mammalian switch regions and are proposed to provide the ssDNA for AID action. We have shown that G-clusters are important for efficient R loop initiation within cells.

(B) Proposed relationships in CSR. CSR efficiency is in direct proportion to the total number of WGCW sites and the amount of ssDNA. The amount of ssDNA is proportional to the transcription firing rate, RIZ strength, and REZ strength. RIZ strength is a function of the total G-cluster number and the G nucleotide number within each G-cluster. REZ strength is dependent on overall G-density, which determines the length of the R loop. Naturally occurring switch regions have interspersed G-clusters, WGCW sites, and a high G-density. Hence, for naturally occurring R loops in cells, the ssDNA length (R loop length) varies depending at which G-cluster the R loop initiates (Roy and Lieber, 2009; Roy et al., 2008, 2010).

R loop formation in a biochemical system using T7 RNA polymerase and naked DNA templates (Daniels and Lieber, 1995). With each round of transcription, a new RNA polymerase must displace the RNA from the existing R loop to progress through that region. Movement of the RNA polymerase through an R loop would be expected to be slower due to the increased difficulty of separating three nucleic acid strands rather than only two.

We also observe that with the same repeat unit, a greater number of such repeats, namely, a longer switch region, causes a decrease in transcription (Figures S3 and S4). This is likely because the RNA polymerase movement through R loops is slower than through duplex DNA, and longer R loops are more difficult than shorter ones for the RNA polymerase.

When we plotted the transcription values against the CSR efficiency of individual cellular clones, the distribution is random (Figure S4). This suggests that above a certain level of tran-

scription, there is no direct correlation between CSR and GLT, supporting previous observations (Dunnick et al., 2004).

Conclusions

Our studies show that Ig switch region strength depends on R loop formation and on the number of WGCW sites within the R loop region. The R loop formation requires transcription and a switch region with a nontemplate DNA strand that is rich in clusters of G-nucleotides and an overall high G-density. The G-clusters can be interspersed with the WGCW sites, just as they are in naturally occurring switch regions; or the G-clusters can be limited to the upstream edge of the WGCW-rich zone, as long as this WGCW-rich zone also has a high G-density to support elongation of the R loop. Therefore, while the total number of WGCW sites is important in the context of our experiments, this number of WGCW sites must be located within an R loop forming region.

EXPERIMENTAL PROCEDURES

Cell Culture and CSR Assay

CH12F3.2a and its derivative cells were cultured in 10% fetal calf serum RPMI medium supplemented with 50 μM β -mercaptoethanol. As for CSR assay, healthy cells in log phase were seeded at 5×10^4 cells/ml in medium with 1 $\mu\text{g/ml}$ anti-CD40 (eBioscience, #16-0404-86), 5 ng/ml interleukin-4 (IL-4; R&D Systems, #404-ML-010), and 0.5 ng/ml transforming growth factor β 1 (TGF- β 1; R&D, #240-B-002), and grown for 72 hr. Cells were stained with fluorescein isothiocyanate-conjugated mouse IgA antibody (BD Biosciences, #559354) and analyzed with flow cytometry. CSR efficiency was determined by the percentage of IgA⁺ cells.

Plasmid Construction

The 80 nt oligonucleotides were 5'-phosphorylated and purified on a denaturing polyacrylamide gel. A pair of complementary oligonucleotides was annealed and self-ligated at a final concentration of 4 μM . Ligation products were resolved on agarose gel, and DNA with expected length was recovered, blunted with Klenow, and cloned into the exchange vector backbone. The sequence was confirmed by DNA sequencing.

Cellular Targeting and Screening

Five micrograms exchange vector and 1 μg Cre expression vector were co-transfected into 1F7 cells by electroporation (Lonza). Transfected cells were serially diluted and seeded in 96-well plates. After 72 hr, ganciclovir (Sigma-Aldrich, #G2536-100MG) was added at a final concentration of 2 $\mu\text{g/ml}$. At 7 days after transfection, single clones were picked for puromycin sensitivity test at a final concentration of 1 $\mu\text{g/ml}$. Puromycin-sensitive clones were screened by PCR around upstream and downstream boundaries of LoxP sites and across the entire switch region. A subset of clones were also examined by Southern blot. At least five clones were used for CSR assay.

Germ-Line Transcript Quantification

Two million healthy cells at a density around 1×10^6 cells/ml were supplemented with anti-CD40, IL-4, and TGF- β 1 for 6 hr, and total RNA was extracted with GenElute Mammalian Total RNA Kit (Sigma-Aldrich, #RTN350). Ten percent of RNA was reverse transcribed into cDNA with M-MuLV Reverse transcriptase (New England Biolabs, M0253S), and 10% of the reverse transcript products were analyzed with real-time PCR. β -Actin was used as an internal control. Each sample was done in duplicate, and at least three independent cellular clones were analyzed for each construct.

Purification

ATCC HB-8730 hybridoma line (generously provided by Bradley Cairns) was cultured in a CELLLine 1000 bioreactor (Satorius Biotech) according to manufacturer's instructions. Harvested antibody (culture supernatant) was purified

on a column packed with Protein G Sepharose 4 Fast Flow (GE Healthcare) equilibrated with 1× PBS.

Immunoprecipitation

Healthy cells in log phase were seeded at 3×10^5 cells/ml in medium with anti-CD40, IL-4, and TGF- β 1 and grown for 24 hr. Genomic DNA was prepared by overnight proteinase K digestion, phenol-chloroform extraction, and ethanol precipitation. Genomic DNA was fragmented with EcoRI; importantly, RNase A was added at this step to prevent S9.6 antibody binding to RNA species in subsequent steps (Phillips et al., 2013). Five micrograms of fragmented genomic DNA was incubated with 5 μ g S9.6 antibody in 400 μ l IP buffer (10 mM sodium phosphate [pH 7.0], 0.195 ml 1M NaH₂PO₄, 0.61 ml 0.5 M Na₂HPO₄/50 ml, 140 mM NaCl [1.4 ml NaCl/50 ml], 0.1% Tween 20 [250 μ l 20%/50 ml]) for 2 hr at 4°C. Ten microliter preblocked Dynabeads (Invitrogen, 10004D) were added into the mixture and gently rotated at 4°C. After 2 hr, beads were washed with IP buffer three times, and treated with proteinase K overnight. DNA bound to the beads was recovered by phenol-chloroform extraction, and quantified by real-time PCR.

Bisulfite Sequencing

One million healthy cells at a density around 1×10^5 cells/ml were supplemented with anti-CD40, IL-4, and TGF- β 1 for 48 hr, and genomic DNA was extracted. Five micrograms genomic DNA was incubated with the bisulfite solution (Lightning Conversion Reagent by Zymo Research, D5030) for 16 hr at 37°C, and recovered as described by the manufacturer. The switch regions were amplified with one converted primer and one native primer by PCR, and subcloned into the pGEM-T Easy vector (Promega, A1360). At least 12 clones from each switch region were randomly chosen and analyzed.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Discussion, six figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2014.06.021>.

ACKNOWLEDGMENTS

We thank former lab member Dr. Deepankar Roy (Genentech) for initial advice and guidance and Dr. Ray Mosteller for comments on the manuscript. This work was supported by NIH grants to M.R.L. and K.Y. Z.Z.Z. was supported initially by a USC Provost's Fellowship. N.R.P. is supported in part by The John H. Richardson Fellowship through the ARCS Foundation.

Received: March 28, 2014

Revised: May 20, 2014

Accepted: June 12, 2014

Published: July 10, 2014

REFERENCES

- Baaklini, I., Usongo, V., Nolent, F., Sanscartier, P., Hraiky, C., Drlica, K., and Drolet, M. (2008). Hypernegative supercoiling inhibits growth by causing RNA degradation. *J. Bacteriol.* **190**, 7346–7356.
- Bailey, J.A., and Eichler, E.E. (2006). Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* **7**, 552–564.
- Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., and Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nat. Struct. Mol. Biol.* **20**, 1199–1205.
- Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. USA* **100**, 4102–4107.
- Broccoli, S., Rallu, F., Sanscartier, P., Cerritelli, S.M., Crouch, R.J., and Drolet, M. (2004). Effects of RNA polymerase modifications on transcription-induced negative supercoiling and associated R-loop formation. *Mol. Microbiol.* **52**, 1769–1779.
- Chaudhuri, J., Basu, U., Zarrin, A., Yan, C., Franco, S., Perlot, T., Vuong, B., Wang, J., Phan, R.T., Datta, A., et al. (2007). Evolution of the immunoglobulin heavy chain class switch recombination mechanism. *Adv. Immunol.* **94**, 157–214.
- Daniels, G.A., and Lieber, M.R. (1995). RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Res.* **23**, 5006–5011.
- Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22.
- Drolet, M., Broccoli, S., Rallu, F., Hraiky, C., Fortin, C., Massé, E., and Baaklini, I. (2003). The problem of hypernegative supercoiling and R-loop formation in transcription. *Front. Biosci.* **8**, d210–d221.
- Dunnick, W., Hertz, G.Z., Scappino, L., and Gritzmacher, C. (1993). DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Res.* **21**, 365–372.
- Dunnick, W.A., Shi, J., Graves, K.A., and Collins, J.T. (2004). Germline transcription and switch recombination of a transgene containing the entire H chain constant region locus: effect of a mutation in a STAT6 binding site in the gamma 1 promoter. *J. Immunol.* **173**, 5531–5539.
- El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* **24**, 1546–1558.
- Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825.
- Hackney, J.A., Misaghi, S., Senger, K., Garris, C., Sun, Y., Lorenzo, M.N., and Zarrin, A.A. (2009). DNA targets of AID evolutionary link between antibody somatic hypermutation and class switch recombination. *Adv. Immunol.* **101**, 163–189.
- Han, L., Masani, S., and Yu, K. (2011). Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination. *Proc. Natl. Acad. Sci. USA* **108**, 11584–11589.
- Hu, Z., Zhang, A., Storz, G., Gottesman, S., and Leppla, S.H. (2006). An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res.* **34**, e52.
- Huang, F.-T., Yu, K., Hsieh, C.-L., and Lieber, M.R. (2006). Downstream boundary of chromosomal R-loops at murine switch regions: implications for the mechanism of class switch recombination. *Proc. Natl. Acad. Sci. USA* **103**, 5030–5035.
- Huang, F.-T., Yu, K., Balter, B.B., Selsing, E., Oruc, Z., Khamlichi, A.A., Hsieh, C.-L., and Lieber, M.R. (2007). Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region. *Mol. Cell. Biol.* **27**, 5921–5932.
- Kouzine, F., Wojtowicz, D., Yamane, A., Resch, W., Kieffer-Kwon, K.R., Bandle, R., Nelson, S., Nakahashi, H., Awasthi, P., Feigenbaum, L., et al. (2013). Global regulation of promoter melting in naive lymphocytes. *Cell* **153**, 988–999.
- Li, X., Wang, J., and Manley, J.L. (2005). Loss of splicing factor ASF/SF2 induces G2 cell cycle arrest and apoptosis, but inhibits internucleosomal DNA fragmentation. *Genes Dev.* **19**, 2705–2714.
- Masani, S., Han, L., and Yu, K. (2013). Apurinic/apryrimidinic endonuclease 1 is the essential nuclease during immunoglobulin class switch recombination. *Mol. Cell. Biol.* **33**, 1468–1473.
- Masukata, H., and Tomizawa, J. (1990). A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. *Cell* **62**, 331–338.
- Mischo, H.E., Gómez-González, B., Grzechnik, P., Rondón, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol. Cell* **41**, 21–32.
- Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* **274**, 18470–18476.

- Muramatsu, M., Nagaoka, H., Shinkura, R., Begum, N.A., and Honjo, T. (2007). Discovery of activation-induced cytidine deaminase, the engraver of antibody memory. *Adv. Immunol.* 94, 1–36.
- Murphy, K.M. (2011). *Janeway's Immunobiology, Eighth Edition* (New York: Garland Science).
- Nakama, M., Kawakami, K., Kajitani, T., Urano, T., and Murakami, Y. (2012). DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. *Genes to cells: devoted to molecular & cellular mechanisms* 17, 218–233.
- Pfeiffer, V., Crittin, J., Grolimund, L., and Lingner, J. (2013). The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *EMBO J.* 32, 2861–2871.
- Pham, P., Calabrese, P., Park, S.J., and Goodman, M.F. (2011). Analysis of a single-stranded DNA-scanning process in which activation-induced deoxycytidine deaminase (AID) deaminates C to U haphazardly and inefficiently to ensure mutational diversity. *J. Biol. Chem.* 286, 24931–24942.
- Phillips, D.D., Garboczi, D.N., Singh, K., Hu, Z., Leppla, S.H., and Leysath, C.E. (2013). The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. *J. Mol. Recognit.* 26, 376–381.
- Roy, D., and Lieber, M.R. (2009). G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol. Cell Biol.* 29, 3124–3133.
- Roy, D., Yu, K., and Lieber, M.R. (2008). Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol. Cell Biol.* 28, 50–60.
- Roy, D., Zhang, Z., Lu, Z., Hsieh, C.L., and Lieber, M.R. (2010). Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol. Cell Biol.* 30, 146–159.
- Sinden, R.R. (1994). *DNA Structure and Function* (San Diego: Academic Press).
- Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* 42, 794–805.
- Smith, G.P. (1974). Unequal crossover and the evolution of multigene families. *Cold Spring Harb. Symp. Quant. Biol.* 38, 507–513.
- Smith, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* 191, 528–535.
- Snapper, C.M., and Finkelman, F.D. (1999). Immunoglobulin Class Switching. In *Fundamental Immunology*, W.E. Paul, ed. (Philadelphia: Lippincott-Raven), pp. 831–861.
- Stavnezer, J., Guikema, J.E., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* 26, 261–292.
- Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev.* 26, 163–175.
- Strachen, T., and Read, A.P. (1999). *Human Molecular Genetics, Second Edition* (New York: Wiley-Liss).
- Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science* 340, 619–621.
- Wabl, M., and Steinberg, C. (1996). Affinity maturation and class switching. *Curr. Opin. Immunol.* 8, 89–92.
- Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell* 44, 978–988.
- Wongsurawat, T., Jenjaroenpun, P., Kwok, C.K., and Kuznetsov, V. (2012). Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res.* 40, e16.
- Yang, Y., McBride, K.M., Hensley, S., Lu, Y., Chedin, F., and Bedford, M.T. (2014). Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol. Cell* 53, 484–497.
- Yu, K., and Lieber, M.R. (2003). Nucleic acid structures and enzymes in the immunoglobulin class switch recombination mechanism. *DNA Repair (Amst.)* 2, 1163–1174.
- Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* 4, 442–451.
- Yu, K., Huang, F.T., and Lieber, M.R. (2004). DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *J. Biol. Chem.* 279, 6496–6500.
- Yu, K., Roy, D., Bayramyan, M., Haworth, I.S., and Lieber, M.R. (2005). Fine-structure analysis of activation-induced deaminase accessibility to class switch region R-loops. *Mol. Cell Biol.* 25, 1730–1736.
- Zarrin, A.A., Tian, M., Wang, J., Borjeson, T., and Alt, F.W. (2005). Influence of switch region length on immunoglobulin class switch recombination. *Proc. Natl. Acad. Sci. USA* 102, 2466–2470.
- Zhu, C., Lee, V., Finn, A., Senger, K., Zarrin, A.A., Du Pasquier, L., and Hsu, E. (2012). Origin of immunoglobulin isotype switching. *Curr. Biol.* 22, 872–880.