R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells

Kefei Yu, Frederic Chedin, Chih-Lin Hsieh, Thomas E. Wilson and Michael R. Lieber

Published online 7 April 2003; doi:10.1038/ni919

The mechanism responsible for immunoglobulin class switch recombination is unknown. Previous work has shown that class switch sequences have the unusual property of forming RNA-DNA hybrids when transcribed *in vitro*. Here we show that the RNA-DNA hybrid structure that forms *in vitro* is an R-loop with a displaced guanine (G)-rich strand that is single-stranded. This R-loop structure exists *in vivo* in B cells that have been stimulated to transcribe the γ3 or the γ2b switch region. The length of the R-loops can exceed 1 kilobase. We propose that this distinctive DNA structure is important in the class switch recombination mechanism

V(D)J and immunoglobulin class switch recombination (CSR) are two site-directed gene rearrangement processes that occur in the immune systems of vertebrates[1,2]. V(D)J recombination assembles the gene segments that comprise the exon encoding the antigen receptor–binding pockets for immunoglobulins and T cell receptors[3,4]. CSR exchanges the isotype of the immunoglobulin constant region from μ (which generates IgM) to any of the downstream isotypes (γ, α or ε) to generate IgG1, IgG2a, IgG2b, IgG3, IgA or IgE[5–7]. Only the heavy chain constant region (and thus the functional properties) of the immunoglobulin changes as a result of this, leaving the antigen specificity unchanged. CSR is triggered by antigen contact and occurs in B cells in the germinal centers of lymph nodes, spleen and the gut-associated lymphoid tissue[8].

Class switch recombination is distinct from all other known DNA recombination processes[1,2]. Switch regions for immunoglobulin-γ, α and ε isotypes (S_γ, S_α and S_ε, respectively) precede the constant region of each isotype (except IgD) and comprise DNA repeats of 25–80 base pairs (bp), which are reiterated[9] over several kilobases (kb). These repeated regions, although G-rich in sequence on the nontemplate DNA strand in mammals, are not homologous to each other, and no consensus sequence has been identified at the junctions of recombined DNA fragments[5]. This raises the possibility that the CSR machinery is targeted to the S regions not by a common sequence motif, but by a common structure formed at these loci[10]. Each S region is located downstream of cytokine-inducible promoters that direct the production of transcripts that do not encode protein (called sterile transcripts)[11]. Without this transcription, class switching is markedly reduced[12–14]. Splicing of the sterile transcript seems to be required for CSR to occur, although the mechanistic role of this is unclear[15,16]. Similarly, the role of both the S region repeats and transcription is not clear.

Several DNA structural alterations have been proposed to occur at S regions. Short DNA stem-loop structures have been suggested to form transiently during transcription[17], although so far no experimental data support this claim. We regard this possibility as unlikely because the transcription bubble, where these structures would form, is short (10–20 nucleotides) and transient (it moves with the RNA polymerase), which would impose strong spatial and temporal constraints on the formation of such DNA structures[18]. The formation of higher order DNA structures, such as guanine quartets (G4), has also been examined for S regions[19,20]. G-rich sequences can indeed form this structure *in vitro*[19]; however, there are no known examples of G4 DNA forming *in vivo* in prokaryotes or eukaryotes.

RNA-DNA hybrid structures have been shown to form on transcription of S regions *in vitro*. A homopurine-homopyrimidine region of 140 bp immediately upstream of S_α forms a stable RNA-DNA hybrid on transcription in the physiological but not the nonphysiological direction[21,22]. In these *in vitro* structures, the newly synthesized RNA molecule remains associated with the S region DNA template strand and does not dissociate, which would generally occur with any random DNA sequence. Such RNA-DNA hybrids form at all tested full-length S regions, thereby generalizing the formation of RNA-DNA hybrids beyond the 140-bp region located upstream of S_α[10].

Although the formation of the RNA-DNA hybrid *in vitro* is clear, the precise base-pairing between the RNA strand and the two DNA strands is not. An intermolecular triplex, in which RNA is the third strand, forms the basis of one proposed model[22]. The R-loop has been suggested as another *in vitro* structure on the basis of the P1 sensitivity at the edge of the S region[23]. The determination of such a structure would require the demonstration of single strands internal to the edges of the S region. This has not been tested, however, and a wide range of *in vitro* structures remain possible[22]. In other words, there are no data that implicate any precise configuration of the structure in the RNA-DNA hybrids formed at S regions *in vitro*. *In vivo*, there are no reports of any considerable length of unusual DNA at S regions or at any other locations.

USC Norris Comprehensive Cancer Center, Room 5428, University of Southern California Keck School of Medicine, 1441 Eastlake Avenue, MC 9176, Los Angeles, CA 90033, USA. Correspondence should be addressed to M.R.L. (lieber@usc.edu).

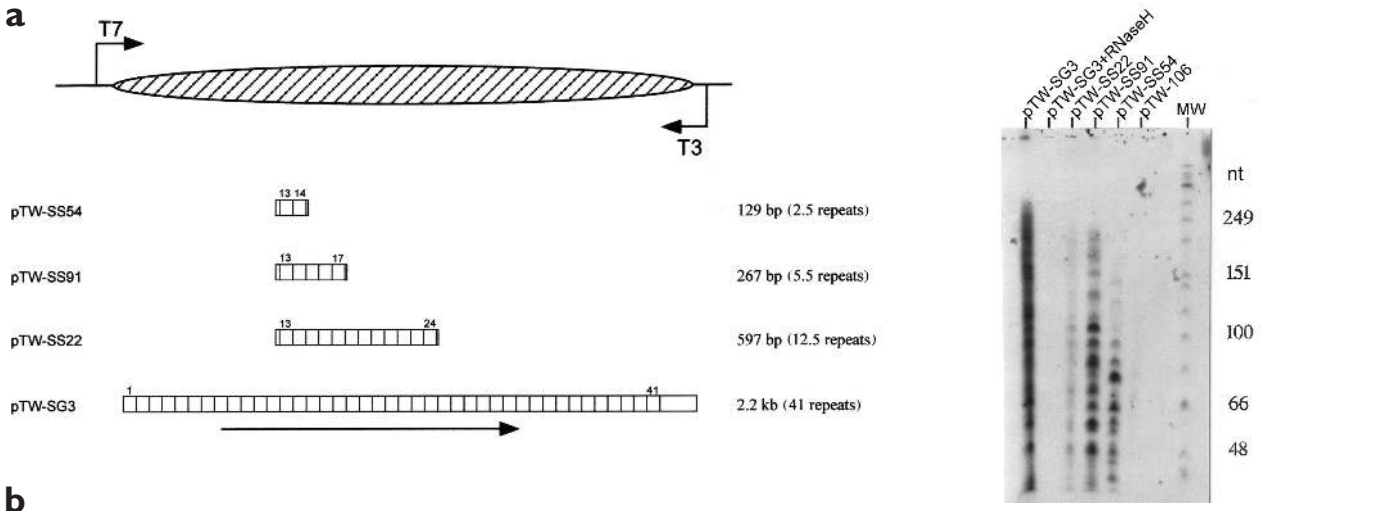

Figure 1. *In vitro* transcription of class switch sequences. (**a**) $S_{\gamma 3}$ internal deletion constructs. The oval represents the murine $S_{\gamma 3}$ region on the plasmid. Each open box unit represents one 49-bp $S_{\gamma 3}$ repeat. pTW-SG3 contains the whole $S_{\gamma 3}$ region with 41 repeats, whereas other plasmids contain only part of the S region, as indicated. The large arrow under pTW-SG3 indicates the physiological direction of transcription. (**b**) Mobility of transcribed plasmids. Each plasmid was transcribed with the indicated RNA polymerase or a mock control (–), treated with RNase A and resolved on an agarose gel. 'OC' and 'SC' indicate the positions of the open circular and supercoiled forms of pTW-SG3, respectively. (**c**) Length of $S_{\gamma 3}$ RNA in the RNA-DNA hybrid. Each plasmid template was transcribed with [α-^{32}P]UTP, treated with RNase A and, where indicated, treated with RNase H. Transcribed plasmids were resolved on an agarose gel and the retarded DNA-bound RNA species were purified and resolved by denaturing PAGE.

Here we have examined the RNA-DNA hybrid structure that forms at S regions *in vitro*. We show that this structure is an R-loop with a displaced G-rich strand that is single-stranded. We also show that this R-loop structure exists *in vivo* in B cells that have been stimulated to transcribe either $S_{\gamma 3}$ or $S_{\gamma 2b}$. The length of these R-loops can exceed 1 kb.

Results

RNA length in *in vitro*–transcribed switch sequences

The murine $S_{\gamma 3}$ region consists of 41 repeats of 49 bp that are similar but not identical to each other[24]. To facilitate the *in vitro* analysis of the structure of S regions and to allow their *in vitro* transcription, we created plasmids containing 2.5, 5.5 and 12.5 repeats of $S_{\gamma 3}$ (called pTW-SS54, pTW-SS91 and pTW-SS22, respectively), cloned such that transcription from the T7 promoter generated a G-rich RNA transcript in the same direction as physiological transcription (**Fig. 1a**). Transcription from the T3 promoter generated a cytosine (C)-rich transcript in the antisense direction. The 2.3-kb fragment from *Hin*dIII-digested λ phage DNA was cloned into the same vector backbone to generate a control plasmid (pTW106).

As reported for a wide range of S regions[10], *in vitro* transcription with the T7 RNA polymerase through the full-length (pTW-SG3) or truncated $S_{\gamma 3}$ regions resulted in a mobility shift of the DNA substrate from the supercoiled position to the nicked circular position (**Fig. 1b**). These shifts were not observed on transcription in the nonphysiological orientation of transcription with T3 RNA polymerase[10,21,22] (**Fig. 1b**). The aberrant migration pattern of this complex was reversed on treatment with ribonuclease H (RNase H), which destroys only RNA that is base-paired with DNA and does not degrade single-stranded RNA (**Fig. 1b**). When [α-^{32}P]rUTP was included in the transcription reactions, the labeled RNA showed the same mobility as the shifted plasmid species, suggesting that the RNA was base-paired with the DNA template[10,21,22]. These data show that S regions, and in particular the $S_{\gamma 3}$ region and its shorter derivatives, can form stable RNA-DNA hybrid structures when transcribed in the physiological orientation.

Once formed, these RNA-DNA hybrids were very stable: they were resistant to treatment with RNase A or phenol, ethanol precipitation and prolonged storage. The RNA-DNA hybrids formed on a circular template were completely stable on linearization (data not shown). Formation of an RNA-DNA hybrid was observed only in *cis*. Incubating supercoiled template with G-rich RNA did not lead to the formation of RNA-DNA hybrids (data not shown).

Including [α-^{32}P]rUTP in the transcription step enabled us to determine the length of the RNA. The RNA species present in RNA-DNA hybrids were distinctly heterogeneous and ranged from fewer than 50 to up to 250 nucleotides (**Fig. 1c**). We did not observe RNA species with more than 250 nucleotides, even when the full-length, 2.2-kb $S_{\gamma 3}$ region was transcribed. Formation of RNA-DNA hybrids on a circular template creates topological constraints that cannot be relieved[22,25,26]. DNA gyrase relieved the torsional stress, resulting in longer RNA species (data not shown).

Single-stranded G-rich strand in RNA-DNA hybrids

Because the G-rich RNA was base-paired to its template DNA strand (thereby accounting for its sensitivity to RNase H), we considered that the nontemplate, G-rich DNA strand might be unpaired and displaced. To test this, we carried out primer annealing and extension reactions on pTW-SS91 in the absence of a denaturation step. Annealing of the primer to pTW-SS91 in the absence of transcription resulted in a weak reaction, probably owing to the natural 'breathing' of DNA. By contrast, annealing of the primer to T7-transcribed pTW-SS91 led to a strong primer extension reaction, which was characterized by the production of many larger species (data not shown). When transcribed pTW-SS91 was treated with RNase H before priming, very little extension was detected, indicating that the single-strandedness of the G-rich DNA strand was dependent on the presence of an RNA-DNA hybrid.

On the basis of these initial findings, we coupled the primer extension step with a subsequent polymerase chain reaction (PCR) assay

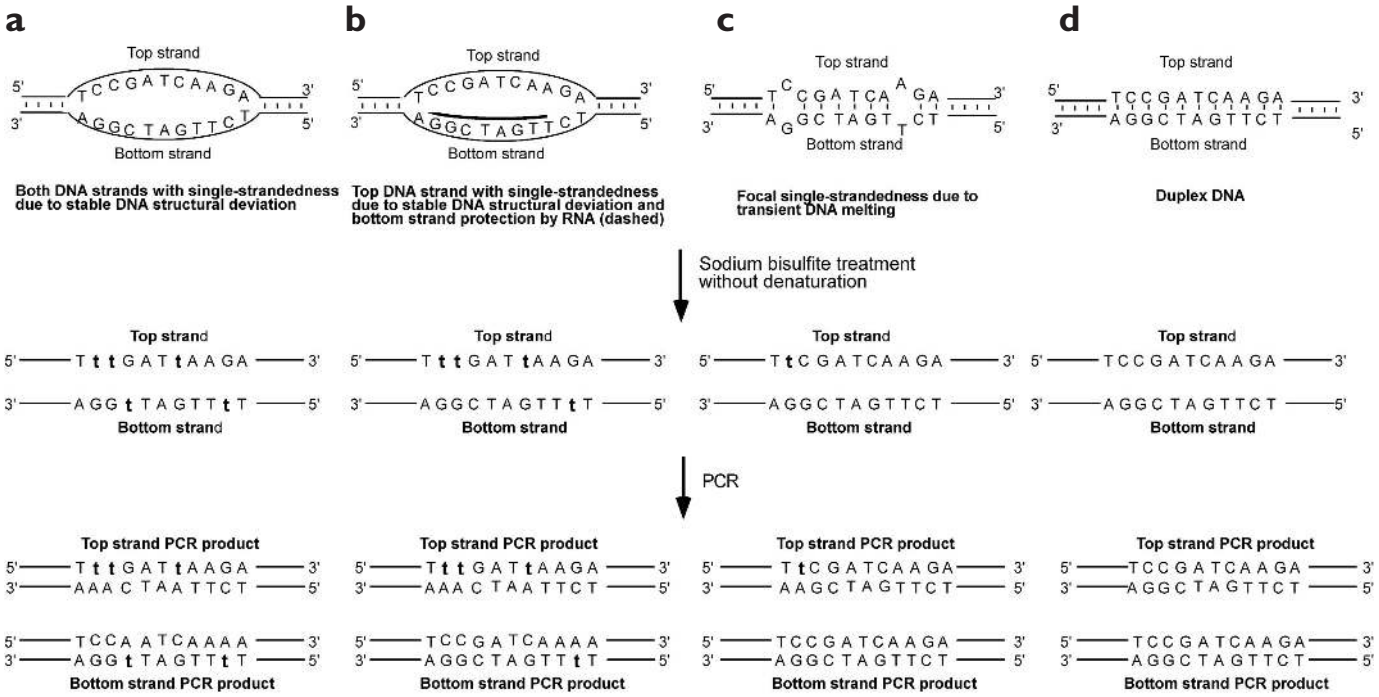

Figure 2. Detection of single-stranded regions with bisulfite. DNA molecules containing regions in a non-B conformation (**a**–**c**) are subjected to bisulfite conversion of unpaired C nucleotides, whereas standard duplex DNA (**d**), which is B-form DNA, does not react with bisulfite. After PCR, converted C nucleotides become T nucleotides (indicated in bold lower case) in the final PCR product. Bisulfite-converted top strands and bottom strands yield distinct PCR products (**a**–**c**). If there is no conversion (**d**), the top and bottom PCR products are identical.

(see **Supplementary Fig. 1** online). Because of the repetitive nature of $S_{\gamma 3}$, the extension oligonucleotide could prime not only at its correct location but also at two other repeats. When T7-transcribed pTW-SS91 was used, strong amplification was observed (**Supplementary Fig. 1**) and two principal species, corresponding to the two most optimal priming sites, were produced. Quantification of the gel (**Supplementary Fig. 1**) showed that transcription in the physiological orientation resulted in a 10- to 50-fold increase in the production of molecules primed on the G-rich DNA strand and that transcription extended up to the site of the T7 primer.

These data show that the G-rich DNA strand contained in the RNA-DNA hybrid region is available for priming with a DNA oligonucleotide at several locations throughout the switch sequence, thereby implying that it is predominantly single-stranded.

Structure of the *in vitro*–transcribed RNA-DNA

Bisulfite modification of DNA followed by DNA sequencing (hereafter referred to as bisulfite modification) is used commonly to detect methylated C nucleotides in DNA methylation studies[27], but sodium bisulfite can also be used as a structural probe to distinguish single- from double-stranded C nucleotides in DNA[28]. Bisulfite converts unpaired C nucleotides to uracil (U) on deamination. Such conversion can be detected by PCR followed by DNA sequencing. During PCR, the polymerase will incorporate A opposite U, resulting in a transition of C•G to T•A. C residues that are paired with G residues in B-form DNA (or C nucleotides that are methylated) do not react with bisulfite and remain as C nucleotides in the readout. For a duplex DNA molecule that is fully melted, all C nucleotides will be converted to T on both strands (after PCR). The two strands will no longer be complementary at a substantial fraction of positions. For a duplex that has a region of single-strandedness, the C conversions will be limited to that single-stranded region. The outcomes for various types of single-strandedness can be distinguished (**Fig. 2**).

We therefore used the reactivity of bisulfite to determine the base-pairing status of the two DNA strands involved in the RNA-DNA hybrid structure. We treated pTW-SS22 transcribed by T7 RNA polymerase with sodium bisulfite and then used the bisulfite-treated DNA as a PCR template to amplify a region containing 12 repeats of $S_{\gamma 3}$. In contrast to conventional PCR, where two complementary strands of the template generate the same PCR product, one can distinguish the initial template strand in the first round of the PCR in bisulfite-treated DNA, based on the sequence of the final PCR product (**Fig. 2**). There will be two types of molecular clones: one reveals bisulfite conversions, and hence single-strandedness on the G-rich strand; the other reveals conversions (single-strandedness) on the C-rich strand.

All cloned PCR products showed at least one conversion, even with bisulfite-treated untranscribed plasmid (see below). This suggested that there was a low background of reactivity even for double-stranded DNA, possibly owing to transient DNA breathing at preferred locations. When we aligned the sequences of all of the clones derived from the G-rich strand, we found that 8 of 12 (67%) molecules showed long stretches of conversion covering at least 170 bp (**Fig. 3a**, top eight molecules). We defined a 'long stretch' as at least 100 bp and a 'short stretch' as between 30 and 100 bp; we did not score lengths shorter than 30 bp. A given length of conversion was defined conservatively as the distance from the first unpaired C to the last unpaired C. In a stretch of conversion, every C nucleotide was converted except for C nucleotides located in CCAGG sequences in which the second C had been methylated by the *Escherichia coli* Dcm methylation system and thus was resistant to bisulfite modification. The percentage of molecules (67%)

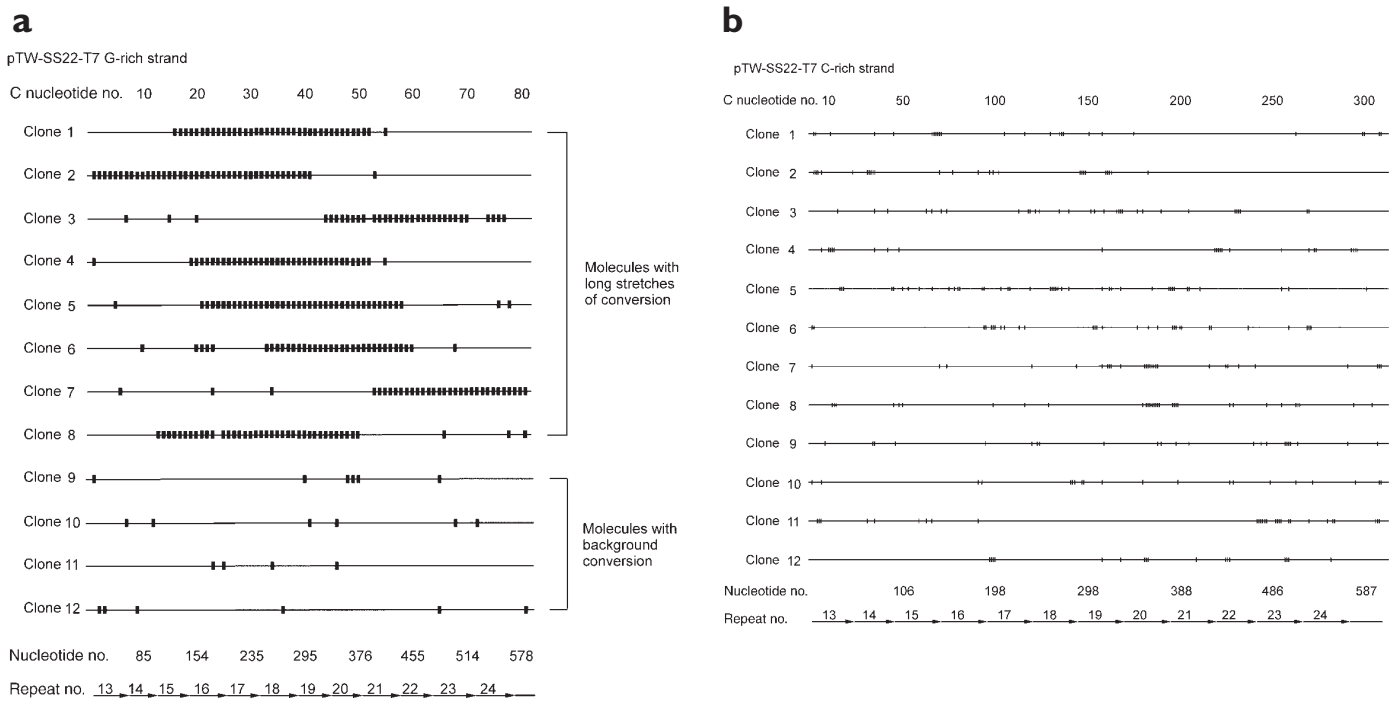

Figure 3. *In vitro*–transcribed $S_{\gamma 3}$ sequences analyzed with bisulfite. (**a**) Single-strandedness of the G-rich nontemplate strand after *in vitro* transcription. Plasmid pTW-SS22 was transcribed with T7 RNA polymerase *in vitro* to form a stable RNA-DNA hybrid, which was treated with bisulfite and then amplified by PCR with a pair of primers located outside the S region. The PCR products were cloned and sequenced in full. Each long horizontal line represents a single clone. Eighty-one C nucleotides in the S region are displayed evenly along the horizontal line; their numbers are indicated at the top in increments of 10. Actual nucleotide positions are indicated below the line (not to scale). Short bold vertical marks indicate the C residues that converted to U. Each horizontal arrow on the bottom represents a 49-bp $S_{\gamma 3}$ repeat (repeats 13–24 of the 41 repeats in $S_{\gamma 3}$). (**b**) The C-rich strand of the *in vitro* R-loop is protected against bisulfite treatment. Twelve molecules derived from the C-rich strand are shown. The symbols are the same as in **a**.

that showed long stretches of conversion on the G-rich strand correlated well with the percentage of the shifted species (80%) on ethidium bromide–stained agarose gels (data not shown).

The single-stranded portion of the G-rich strand occurred at different positions in the S region and extended to different lengths. The longest single-stranded region observed covered about 300 bp (**Fig. 3a**, molecule 2), which was in agreement with the longest RNA molecules detected in the RNA-DNA hybrid (**Fig. 1c**). Molecules 3 and 8 showed a skip of one C in a long stretch of converted C nucleotides, possibly corresponding to a T-to-C mutation introduced by PCR. In our assays, T-to-C mutations were observed more frequently than were other PCR-generated mutations (data not shown). Some molecules showed sporadic background conversions and most probably corresponded to the small percentage of molecules that were not involved in the RNA-DNA structure (**Fig. 3a**, last four molecules).

We also obtained 12 molecules originating from the C-rich strand. None of these molecules showed any long stretches of conversion (**Fig. 3b**). Although small clusters of conversion covering 20–30 bp were detected, the C-rich strand was predominantly protected from bisulfite modification (**Fig. 3b**). This protection was most probably due to the base-pairing between the G-rich RNA and the C-rich DNA strand. These results strongly support the conclusion that the RNA-DNA hybrids formed on transcription of $S_{\gamma 3}$ correspond to R-loop structures of varying lengths.

We next examined whether the formation of the R-loop structures was dependent on transcription. When untranscribed supercoiled pTW-SS22 was treated with bisulfite, none of the 11 molecules that we analyzed from the G-rich strand showed long stretches of conversion (see **Supplementary Fig. 2** online). Similarly, when pTW-SS22 was transcribed with T3 RNA polymerase in the nonphysiological orientation, no long stretches of single-strandedness on the G-rich strand were observed (**Supplementary Fig. 2**). In both cases, the C-rich strand was also protected from bisulfite modification (data not shown). Thus, bisulfite treatment of untranscribed DNA, or of DNA transcribed in the nonphysiological orientation, did not result in the formation of long single-stranded regions. The only modifications observed for these molecules reflected spontaneous breathing sites and were very localized and sporadic.

To assess whether the ability to form R-loops was specific to the $S_{\gamma 3}$ sequence, we analyzed the gene encoding β-lactamase on the plasmid backbone. We used T7-transcribed pTW-SS22 DNA, for which long stretches of bisulfite conversion were observed at the $S_{\gamma 3}$ region (**Fig. 3**), and amplified a portion of the β-lactamase coding region by PCR. Only a few sporadic background conversions on both DNA strands were detected (see **Supplementary Fig. 2** online). R-loop formation is therefore specific to the S region.

To test whether RNA-DNA hybrid formation was necessary to detect long stretches of conversion on the G-rich strand, we treated T7-transcribed pTW-SS22 with RNase H to remove the RNA from the RNA-DNA hybrid before bisulfite treatment. The S region was then amplified by PCR and sequenced. We obtained six molecules from the G-rich strand. None of these contained any long stretch of single-strandedness, although a few showed some extent of single-strandedness. Molecule 3 showed bisulfite conversion for a region of about 95 bp and molecule 1 showed a similar stretch for about 70 bp (see **Supplementary Fig. 3** online).

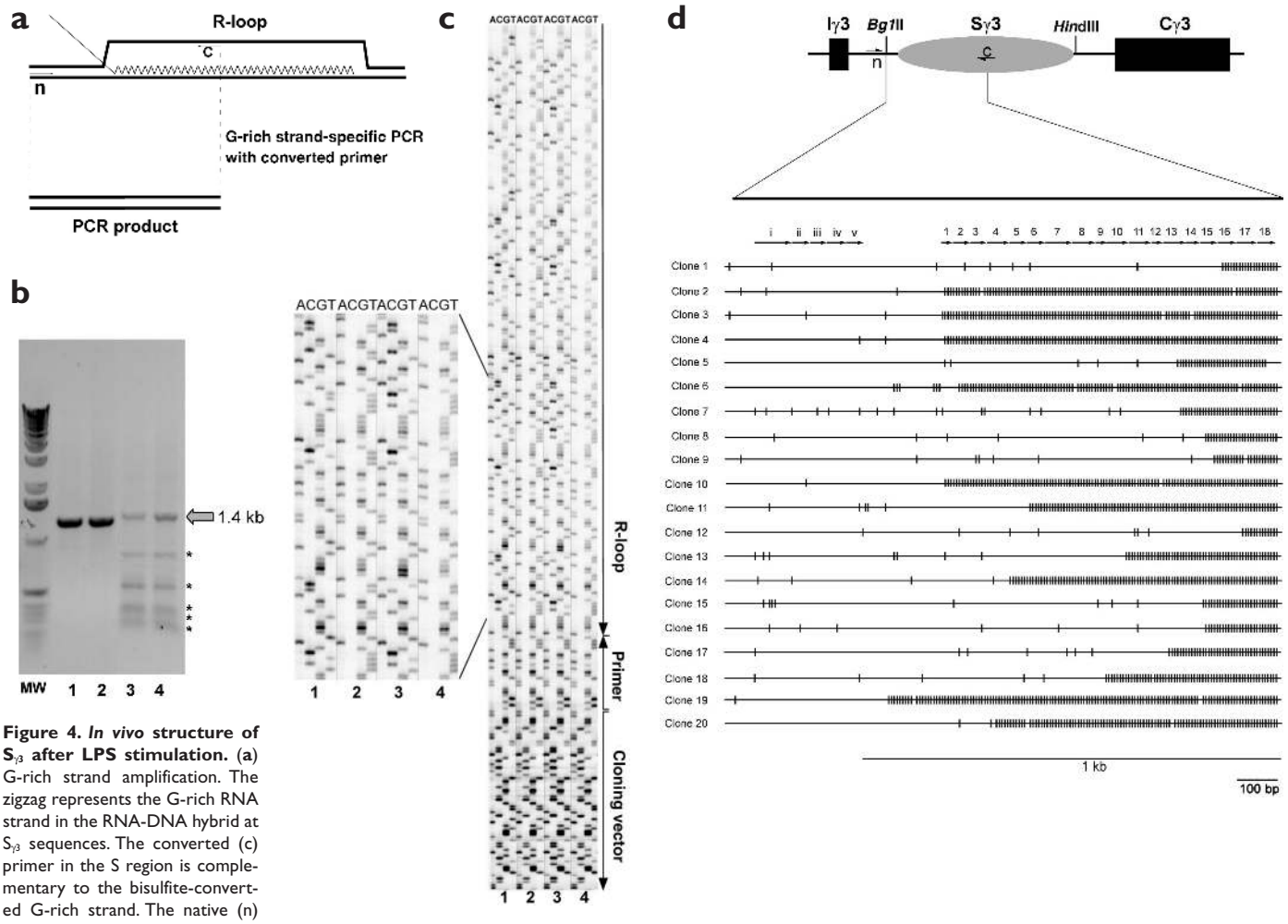

Figure 4. *In vivo* structure of $S_{\gamma 3}$ after LPS stimulation. (**a**) G-rich strand amplification. The zigzag represents the G-rich RNA strand in the RNA-DNA hybrid at $S_{\gamma 3}$ sequences. The converted (c) primer in the S region is complementary to the bisulfite-converted G-rich strand. The native (n) primer outside the S region is complementary to the unconverted C-rich strand. Only single-stranded templates accessible to bisulfite can anneal to converted primers and undergo amplification. (**b**) PCR products from the converted and native primers. Bisulfite-treated DNA from mouse spleen B cells was amplified, and the PCR product of the correct size was cloned and sequenced. To control for the overall template quality, a region upstream of $S_{\gamma 3}$ including the $I_{\gamma 3}$ exon was also amplified. Lanes 1 and 2, control PCR with unstimulated and stimulated B cell templates, respectively; lanes 3 and 4, $S_{\gamma 3}$ G-rich strand–specific PCR with unstimulated and stimulated B cell templates, respectively. Arrow indicates the expected PCR product; asterisks indicate PCR by-products formed through mispriming to the most similar repeats of the S region. (**c**) Extensive continuous C-to-U conversion on the G-rich strand of $S_{\gamma 3}$ in stimulated mouse B cells. The sequencing gel shows long stretches of C-to-U (T) conversion on the G-rich strand of the S region in stimulated B cells. Each sequencing lane was labeled with the complementary nucleotide to facilitate the reading of the G-rich strand sequence. Four sequenced samples are shown. (**d**) Extensive single-strandedness of the G-rich strand in the $S_{\gamma 3}$ region of stimulated B cells. All 20 molecules with long stretches of single-strandedness on the G-rich strand (**Table 1**) are shown. Arrows represent $S_{\gamma 3}$ repeats[24]. Degenerate repeats are numbered i–v; canonical repeats are numbered 1–19. Arrows indicate the positions of the PCR primers (n, native; c, converted). Other symbols are defined in **Fig. 3**.

Similar results were obtained from analysis of the nine molecules that derived from the C-rich strand. Most of these molecules contained single-stranded conversion stretches that were much longer than those observed on the C-rich strand in RNA-DNA hybrids not treated with RNase H (compare **Fig. 3b** with **Supplementary Fig. 3** online). For example, molecule 8 showed a long stretch of single-strandedness covering 160 bp.

We propose that the stretches remaining in the RNase H–treated samples are due to the misalignment of the repeats as the top and bottom DNA strands reanneal on degradation of the RNA by RNase H. Misalignment of repeats after RNase H treatment would yield a final structure of two heterologous loops: one on the top strand and one on the bottom strand.

RNA-DNA formation on a human minichromosome

Given our observations with *in vitro* substrates, we were interested in whether R-loops could form at S regions *in vivo*. We therefore cloned parts of the $S_{\gamma 3}$ region into plasmids that could be maintained stably as episomes in the nucleus of human cell lines. Plasmids bearing the Epstein-Barr viral replication origin sequence (OriP) replicate once every cell cycle when the EBV nuclear antigen 1 (EBNA1) protein is present. We use the human epithelial carcinoma cell line 293, which has a chromosomal integration of the *EBNA1* gene (293/EBNA1), as the host cell line to maintain such minichromosomes stably. We cloned 12 $S_{\gamma 3}$ repeats from pTW-SS22 into pREP4 downstream of the RSV promoter in both orientations, generating pKY81 and pKY82. Transcription through the S region repeats in pKY81 occurred in the physiological direction and generated a G-rich RNA transcript, whereas transcription ran in the nonphysiological direction in pKY82 and generated a C-rich RNA transcript (see **Supplementary Fig. 4** online).

The pKY81 and pKY82 minichromosomes were obtained from 293/EBNA1 cells[29] and digested with restriction endonucleases to release the switch repeats. Digested DNA was analyzed by Southern

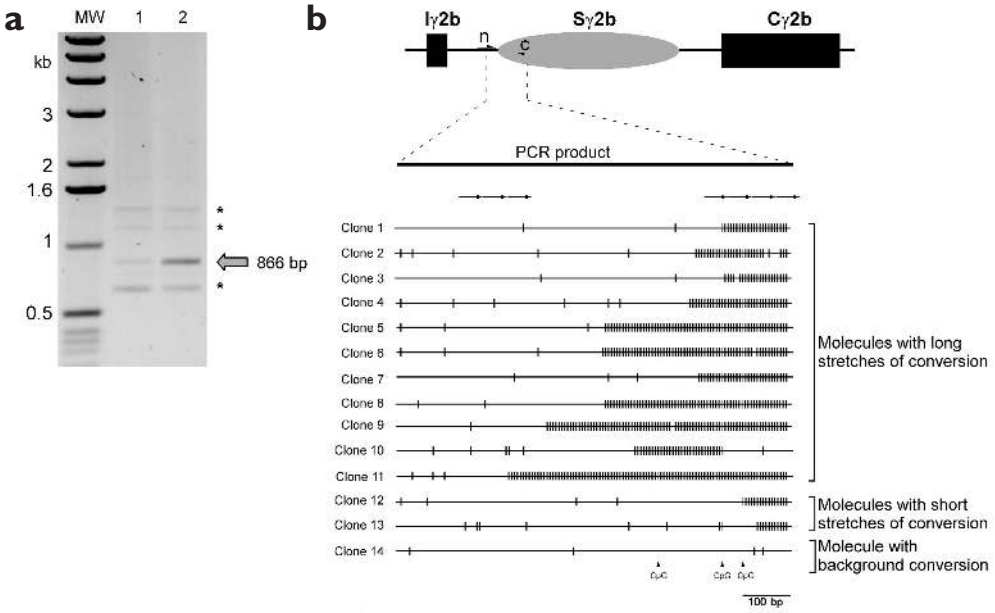

Figure 5. Evidence of R-loops at $S_{\gamma 2b}$ in LPS-stimulated B cells. (**a**) PCR product of $S_{\gamma 2b}$ amplification. The DNA template in **Fig. 4** was used to amplify an 866-bp fragment containing the first portion of the $S_{\gamma 2b}$ region. Arrow indicates the expected PCR product; asterisks indicate the PCR by-products. (**b**) Extensive single-strandedness of the G-rich strand of the $S_{\gamma 2b}$ region in stimulated B cells. The seven arrows represent $S_{\gamma 2b}$ repeats on the upstream side of this S region, according to GenBank entry D78344. Short vertical gray marks indicate an unconverted C in a CpG context; these C nucleotides are most probably unconverted because of mammalian CpG methylation. Other symbols are defined in **Figs. 3** and **4**.

blotting and probed with an $S_{\gamma 3}$ probe. When the $S_{\gamma 3}$ region was transcribed in the physiological orientation (pKY81), a diffuse band corresponding to species with slower mobility was observed above the normal position of the double-stranded 12-repeat fragment (see **Supplementary Fig. 4** online). This slightly retarded migration pattern corresponded to the pattern expected for RNA-DNA hybrids present on linear molecules (data not shown). Indeed, RNase H treatment completely removed these species (**Supplementary Fig. 4**), confirming that they corresponded to RNA-DNA hybrids. In several repetitions of this experiment, the population of RNA-DNA hybrid species accounted for about 2–10% of the radioactivity of the lane.

Transcription through the S region repeats in the nonphysiological orientation (pKY82) did not generate retarded bands (**Supplementary Fig. 4**), as expected from our *in vitro* studies described above and from earlier work[10]. This indicates that formation of RNA-DNA hybrids is not restricted to transcription initiated by bacterial phage RNA polymerases. Transcription of S regions by eukaryotic RNA polymerase II inside mammalian cells can also result, albeit at a reduced efficiency, in the formation of stable RNA-DNA hybrids.

The structure of these RNA-DNA hybrids most probably corresponds to R-loops in which the G-rich strand is unpaired. Indeed, the properties (slower migration, orientation dependence and sensitivity to RNase H) of the *in vivo* RNA-DNA hybrids were similar to those of the *in vitro* RNA-DNA hybrids. In addition, bisulfite modification analysis of pKY81 obtained from mammalian cells also detected long stretches of single-stranded character on the G-rich strand of the S region repeats but indicated that the C-rich strand was fully protected (data not shown).

R-loops in mouse B lymphocyte chromosomes

The bisulfite modification assay provided us with a specific and sensitive way to detect R-loops and to elucidate their fine structure. As R-loops formed at the $S_{\gamma 3}$ region carried by episomes in mammalian cells *in vivo*, we next assessed whether these structures could form at their endogenous loci in the chromosomes of class switching B cells.

We first adapted the original bisulfite method to increase its sensitivity because only a small population of B cells undergo CSR at any given time. For example, after 5 d of stimulation with lipopolysaccharide (LPS), only 3–5% of the mouse spleen naive B cells become $IgG3^+$ cells[30]. To increase the sensitivity of the method, we designed a primer that was complementary to the bisulfite-converted G-rich strand (converted primer). Such converted primers should anneal optimally only to the converted G-rich strand, thereby increasing the chance of recovering molecules with single-stranded character at the region of priming (**Fig. 4a**). Normal genomic DNA will be protected against bisulfite modification and will be disfavored in the PCR reaction.

This enrichment method has two limitations: the proportion of molecules that show long R-loops will not reflect the actual frequency of the formation of this structure in the genome, and the size of the R-loops recovered after this enrichment will be underestimated because the extent of detected R-loops is limited to the region between the two priming sites (**Fig. 4a**). No information can be gained about the portion of $S_{\gamma 3}$ outside the amplification region, and thus the sizes of the structures recovered from the genome will be underestimated. Similarly, any R-loop that is not positioned to encompass the converted primer site will not be detected by this method, even if it is located in the amplified region. Despite these limitations, the recovery of any molecule with an R-loop structure by this method would be clear evidence that such structures could form in the chromosome of B cells.

To obtain B cells that could undergo class switching, we obtained spleens from C57BL/6 mice aged 8–12 weeks and isolated small resting B cells by magnetic cell sorting. Purified $CD43^-$ B cells were cultured *in vitro* with 20 μg/ml of LPS. Cells were collected after 2 d of stimulation and genomic DNA was obtained. Purified genomic DNA was then fragmented with a restriction endonuclease (*Eco*RI) that cuts far outside the murine $S_{\gamma 3}$ locus to reduce the size of the DNA. We then treated the *Eco*RI-fragmented DNA with bisulfite, as described above. PCR was carried out with a native primer located about 500 bp upstream of the $S_{\gamma 3}$ region and a converted primer located at the nineteenth repeat of $S_{\gamma 3}$ (equivalent to the twenty-third repeat for the BALB/c strain). We expected a PCR product to be produced only if the priming region for the converted oligonucleotide was single-stranded and was therefore converted by bisulfite. This could happen if the target molecule was in an R-loop conformation, although it also could happen if there were substantial background conversions at this priming site that enabled the converted primer to anneal.

Table 1. R-loops in the chromosomes of stimulated B cells

B cells	Total molecules sequenced	Molecules with long stretches of conversion	Molecules with short stretches of conversion
Unstimulated	58	0	1
Stimulated	57	20	4
Stimulated + RNase H	26	1	2

Genomic DNA was harvested from mouse spleen B cells that were naive (unstimulated) and LPS-stimulated B (stimulated). A fraction of the stimulated B cell DNA was treated with an excess of RNase H at 37 °C overnight. PCR amplification of the G-rich strand of the mouse $S_{\gamma 3}$ region was carried out with a native primer annealed upstream of $S_{\gamma 3}$ and a converted primer annealed to the converted nineteenth $S_{\gamma 3}$ repeat (Methods). PCR products were cloned and sequenced. Long and short stretches of conversion are shown in **Fig. 3**. A more detailed breakdown of these data from multiple independent experiments is listed in **Supplementary Table 1**.

Bisulfite-modified DNA from unstimulated or stimulated B cells efficiently yielded PCR products when a region upstream of $S_{\gamma 3}$ was amplified (**Fig. 4b**, lanes 1 and 2), suggesting that both were good templates for amplifying a region outside the S region. When we amplified the G-rich strand of $S_{\gamma 3}$ with the converted primer, we consistently found that DNA from stimulated B cells was a slightly more efficient template (**Fig. 4b**, lanes 3 and 4). Although the expected PCR product size was 1.4 kb, smaller PCR by-products could also be detected (**Fig. 4b**). This was most probably caused by the converted primer annealing at suboptimal priming sites owing to the high degree of repetitiveness in the $S_{\gamma 3}$ region. For example, the ninth repeat matched very well with the converted primer, which resulted in a fragment of 900 bp. To avoid confusion arising from the generation of multiple PCR fragments, we cloned and sequenced only the 1.4-kb PCR product isolated from an agarose gel (**Fig. 4b,c**).

We sequenced 58 molecular clones obtained from unstimulated B cell DNA and identified only one molecule with a short stretch of conversion starting from the converted primer (**Table 1**). The other molecules had sporadic conversions on the G-rich strand, which most probably resulted from the converted primer binding at its cognate site owing to background conversion. The clone that showed a short stretch of conversion was probably generated from an authentic R-loop. The presence of R-loops in unstimulated B cells is not surprising because unstimulated B cells have some basal transcription through the $S_{\gamma 3}$ region[30]. A few B cells in the process of activation might also have been mixed with the unstimulated B cells.

We sequenced 57 molecular clones obtained from stimulated B cell DNA and found that 20 of them contained long stretches of conversion starting from the primer site and 4 of them contained short stretches of conversion (**Table 1** and **Fig. 4d**). The long stretches of conversion were distinctive on a sequencing gel (**Fig. 4c**), in which the C lane was completely devoid of bands from the primer site because all of the C nucleotides had been converted to T (**Fig. 4c**, compare the C and T lanes from samples 1 and 3 to those from samples 2 and 4). The observation of numerous molecules with long regions of single-stranded character accessible to bisulfite modification indicates that R-loop structures are likely to exist in the chromosomes of stimulated B cells.

The positions at which the various R-loops stopped were variable and had no obvious consensus termination signal, although a higher than random percentage of them ended at or very near the first partial repeat of the conventionally numbered $S_{\gamma 3}$ region defined by GenBank entry D78343 (**Fig. 4d**). Although we cannot determine the percentage of cellular alleles that contain R-loops, it is clear that the R-loop structure is more abundant in stimulated B cells, suggesting that transcription of the S region in stimulated B cells is required for R-loop formation. Although PCR with a converted primer targeting the G-rich strand is a sensitive way to detect infrequent RNA-DNA hybrid structures among a vast excess of double-stranded target DNA, it does not provide information on the base-pairing status of the C-rich strand. Our initial attempts to detect the endogenous RNA-DNA hybrid structure with native primers located on each side of the $S_{\gamma 3}$ region had a large bias toward amplifying the C-rich strand.

To study further the role of transcription in chromosomal R-loop formation, we treated genomic DNA obtained from stimulated B cells with RNase H before the modification assay. Of 27 molecular clones obtained, we detected 1 clone with a long stretch of conversion of about 100 bp and 2 clones with short stretches (**Table 1**). Similar to the *in vitro* R-loops, these infrequent tracts of conversion might be due to misalignment of the two DNA strands on the removal of the RNA by RNase H.

Murine B cells switch to $S_{\gamma 2b}$ on stimulation by LPS[31]. We therefore examined the $S_{\gamma 2b}$ region from the same LPS-stimulated B cells that were used for the $S_{\gamma 3}$ analysis above. We positioned a native primer upstream of $S_{\gamma 2b}$ and a converted primer about 200 bp into the $S_{\gamma 2b}$ core repeats; these primers were expected to produce a PCR product of 866 bp. On PCR, the bisulfite-treated stimulated B cell DNA did indeed produce a prominent band at 866 bp, whereas the unstimulated sample showed a very weak band at this position (**Fig. 5a**).

We isolated the 866-bp band from both the stimulated and unstimulated sample lanes and sequenced the clones as described above. In the stimulated B cell DNA, we found extensive R-loops in 11 of 14 clones, with the longest one up to 600 bp (**Fig. 5b**). In the unstimulated B cells, most clones had sequences of unknown origin, consistent with the very weak 866 bp PCR product on the agarose gel. Of the few clones that were from the $S_{\gamma 2b}$ region, long R-loops were present in 4 of 13 clones. In contrast to the R-loops in stimulated B cells, those from unstimulated B cells were all shorter than 200 bp (data not shown).

We conclude that R-loops formed at the $S_{\gamma 2b}$ locus are more abundant in stimulated B cells than in unstimulated cells, as they were at the $S_{\gamma 3}$ region. The R-loops did not extend into the region upstream of the S region (**Figs. 4d** and **5b**). The relative precision of this boundary argues that the *in vivo* R-loops, like the *in vitro* ones, are restricted to the S regions.

Discussion

We have shown that a long (>1 kb), stable R-loop can exist on chromosomes in mammalian cells at specialized locations that undergo CSR. The existence of R-loops in prokaryotes, including stable R-loops at the origins of DNA replication for the *colE1* origin and the T4 phage origin, is well documented[32,33]. An *in vivo* R-loop at the ribosomal RNA genes in *E. coli* has been also reported[25,26]. In eukaryotes, mitochondrial replication origins have stable R-loops[34,35], and R-loops at the *E. coli* origin of replication[36] and D-loops at eukaryotic telomeres[37] have been identified *in vitro* but not *in vivo*. Despite these well-documented examples, previously there has been no *in vivo* evidence for R-loops in the eukaryotic nucleus, where the balance of nuclear enzymes (helicases, nucleases and topoisomerases) can differ considerably.

The R-loops at replication origins (typically <40 bp) are considerably shorter than the ones at S regions identified here. The observation of long stable R-loops in mammalian chromosomes indicates that deviations from normal B-form DNA should be added to the ways in which the information content of DNA can be conveyed. That is, the structure

of the DNA can be as important, or even more important, than the precise sequence *per se*.

In contrast to other DNA structural analysis methods, such as P1 nuclease sensitivity and chemical cleavage methods with potassium permanganate and osmium tetroxide, the bisulfite modification assay provides nucleotide resolution of the single-stranded regions of DNA structure. In addition, a complete picture of the DNA structure on a given strand can be obtained for each sequenced clone, whereas cleavage-based assays show only the structural information for a population of molecules across a strand of DNA. Thus, the compilation of information from several bisulfite-sequenced molecules permits a useful molecule-to-molecule structural comparison.

It is known that a purine-rich RNA strand pairs with a pyrimidine-rich DNA strand more stably than a pyrimidine-rich RNA pairs with a purine-rich DNA[38,39], which explains the directional requirement for R-loop formation observed in both the *in vitro* studies and the *in vivo* minichromosome studies. It also explains why class switch sequences have evolved to have a purine-rich nontemplate strand rather than a pyrimidine-rich one.

G-rich DNA can form G4 DNA structures *in vitro*[19]. Although we cannot definitively rule out G4 DNA formation by the single-stranded G-rich DNA strand at class switch R-loops, our observations do not support such formation. First, the amount of the RNA-DNA hybrid generated in *in vitro* transcription reactions did not vary as a function of the monovalent (Na^+, K^+ or Li^+) or divalent (Mg^{2+} or Ca^{2+}) ions (ref. 10; and K.Y. and M.R.L., unpublished data). Second, we could anneal a primer to the displaced G-rich strand and extend it, indicating that the displaced strand had single-stranded character. Last, our bisulfite modification assay data implied the single-stranded character of the displaced G-rich strand. These observations suggest that the G-rich DNA strand may not include any G4 conformation, although they do not rule out the presence of very small amounts of G4 structure.

It is important to note the marked molecule-to-molecule variation in R-loop formation. We could achieve a mobility shift of nearly all of the molecules *in vitro*. Among the shifted molecules, however, the precise start and stop positions of the R-loop varied widely, as determined by the bisulfite modification assay. In fact, from a technical standpoint, the heterogeneity of the structure necessitates a method such as the bisulfite modification assay because such heterogeneity makes evaluation by other methods difficult or impossible.

The heterogeneity of the start and stop positions of the R-loop relative to the repeats may reflect the fact that any one repeat is inefficient at forming an R-loop. This would explain why so many repeats are needed for CSR. As stated above, class switch regions are the only DNA recombination targets that are not highly conserved in sequence between different S regions and different vertebrates. Thus, their sequence variation raises the possibility that their composition and configuration affect DNA structure. A second reason for heterogeneity of the R-loop location may be that each round of action by RNA polymerase may displace the previous RNA molecule and lay in a new one.

In vitro data show that the RNA-DNA hybrid is a structure that is common to all S regions that can function as a target of the class switch machinery[10,23]. If the *in vivo* R-loop identified here is the target for a CSR enzyme, it would explain why CSR is a region-specific process and not a signal sequence–specific process. It would also explain the role of transcription (which produces sterile transcripts), which would be necessary to position the RNA in the RNA-DNA structure. Highly relevant to elucidating the *in vivo* role of S region orientation and possible R-loop formation, murine gene replacement studies reported in this issue have shown that inversion of the $S_{\gamma 1}$ region markedly decreases CSR at this locus[40]. The directional dependence of class switching also has been seen for the S_μ and $S_{\gamma 3}$ regions by using extrachromosomal class switch substrates[41].

We found that RNase H treatment of *in vitro* R-loops enabled single-stranded regions to remain on both strands of the S region. In addition, RNase H treatment of chromosomal DNA from stimulated B cells did not completely remove the single-stranded region on the G-rich strand of the S region, even though we estimate that we used a 10- to 100-fold excess of enzyme. In an attempt to determined whether RNase H treatment could completely remove clustered C-to-U conversion on the C-rich strand, we found a few molecules that had long stretches of single-strandedness on the C-rich strand (data not shown). These molecules were not found in samples that were not treated with RNase H. This unexpected finding, together with the evidence from RNase H–treated *in vitro* R-loops, raises the possibility that removal of the RNA from the RNA-DNA hybrid (by mammalian RNase H enzyme) results in a collapsed structure, in which misaligned repeats are exposed as single-stranded regions on both strands of the switch sequence; for example, the top strand of repeat 18 could anneal with the bottom strand of repeat 24.

So far we have not been able to determine whether it is the RNA-DNA hybrid structure or the collapsed, misaligned switch repeats that constitute the direct target of CSR. It has been reported that overexpression of *E. coli* RNase H does not affect class switch efficiency in a cultured cell line[42]. In that study, the expressed prokaryotic RNase H may not have been abundant enough to destroy the RNA-DNA hybrid, or the RNA-DNA hybrid in the chromosome may not have been accessible to the expressed RNase H activity. But if these are not the reasons why RNase H did not affect class switching, it may be that the misaligned repeats (secondary DNA structure) are in fact the target for CSR. If so, then overexpression of RNase H might facilitate, rather than inhibit, CSR in those cells. If the endogenous RNase H activity in mammalian cells is already sufficient to remove the RNA in the R-loop with some efficiency, exogenous expression may not accelerate the class switch process any further. It will be interesting to characterize the RNase H activities in class switching B cells and to determine whether a deficiency in any of these activities will hinder CSR.

What enzyme activity might cleave class switch regions? Activation-induced deaminase (AID) is a protein of 24 kDa that is essential for both CSR and somatic hypermutation[43,44]. AID deaminates DNA[45], and it has been suggested that, in conjunction with uracil-DNA glycosylase[46], AID may result in the cleavage of DNA. If so, there would need to be structural distinction among class switch regions that would make them the target of AID. Perhaps the R-loop or the collapsed misaligned R-loop identified here might function as such a target. Irrespective of the possible link between AID and the R-loop structures in the chromosomes of activated B cells, the generation of a long R-loop conformation in the eukaryotic nucleus is an unusual configuration that is likely to be important in class switching.

Methods

Enzymes, chemicals and reagents. We purchased RNase H from Promega (Madison, WI) and RNase A from Sigma (St. Louis, MO). All restriction endonucleases were from New England Biolabs (Beverly, MA) and sodium bisulfite and other chemicals were from Sigma. Radioisotope-labeled nucleotides were purchased from NEN (Boston, MA). Plasmid pREP4 was purchased from Invitrogen (Carlsbad, CA). The University of Southern California Institutional Animal Care and Use Committee approved all experiments on live animals. We purchased C57BL/6 inbred mice from the Jackson Laboratories (Bar Harbor, Maine).

Cell culture and mouse spleen B cells. The human epithelial cell line 293/EBNA1 was cultured in complete Dulbecco's Modified Eagle's medium (high glucose; Irvine Scientific, Santa Ana, CA) supplemented with 10% fetal bovine serum (Life Technology, Grand Island, NY), 100 μg/ml of penicillin and 100 μg/ml of streptomycin (Irvine Scientific). We transfected human minichromosomes by calcium phosphate precipitation. Cells stably

maintaining human minichromosomes were selected by growing cells in medium containing 200 μg/ml of hygromycin.

We purified mouse spleen B cells by magnetic cell sorting (Miltenyi, Auburn, CA) according to the manufacturer's instructions. In brief, a single-cell suspension was prepared from the spleen of C57BL/6 mice aged 8–12 weeks. Cells were resuspended in an ammonium chloride red cell lysis buffer and incubated 5 min at 22 °C to lyse the red blood cells. Intact white blood cells were washed twice in PBS containing 1% bovine serum albumin (BSA). Cells were resuspended in 900 μl of PBS plus 1% BSA and mixed with 100 μl of anti-CD43–conjugated microbeads (Miltenyi) for 30 min at 4 °C. We loaded the mixture onto an LS25 column attached to a magnetic stand. The column was washed three times with 3 ml of PBS plus 1% BSA and the flow-through ($CD43^-$) was collected. We determined cell numbers with a hemacytometer and collected a fraction of the unstimulated B cells. The remaining $CD43^-$ cells were resuspended at 2×10^5 cells/ml in complete RPMI 1640 medium supplemented with 10% fetal bovine serum, 100 μg/ml of penicillin, 100 μg/ml of streptomycin and 20 μg/ml of LPS and cultured in a humidified cell incubator at 37 °C and 5% CO_2 for 2 d.

Genomic DNA from mouse B cells was obtained by lysing the cells in 10 mM Tris and 1 mM EDTA (TE) buffer (pH 8.0) containing 0.5% sodium dodecyl sulfate, and then incubating them with proteinase K overnight at 37 °C. We extracted the genomic DNA with phenol and chloroform and precipitated it in ethanol. Air-dried genomic DNA pellets were dissolved at about 1 mg/ml in TE (pH 8.0). Before bisulfite treatment, genomic DNA was digested with *Eco*RI to reduce its size. Digested genomic DNA was extracted again with phenol and chloroform and precipitated in ethanol.

Oligonucleotides. All oligonucleotides were purchased from Qiagen/Operon Techologies (Richmond, CA). The extension primer (see **Supplementary Fig. 1** online) was FC34 (5′-GGGCAAGTCTTACTAGGCTCCCCACCTACCCTAGTTCCCCCAAG-3′) and the tail primer used in the subsequent PCR amplification was FC33 (5′- GGGCAAGTCTTACTAGGC-3′). The native primers (PCR primers that matched unconverted template sequence outside the S region) for amplification of the 12-switch repeat insert of murine $S_{\gamma 3}$ in pTW-SS22 were KY253 (5′-TCTGAATTCAGATCTATGAAAGGTC-3′) and KY254 (5′-CATGCTCTAGAATTAACCCTCAC-3′). KY253 and KY254 were located on the backbone of pTW-SS22 flanking the Sγ3 sequence. The native primers for amplification of a portion of the β-lactamase coding region were KY260 (5′-GAATTATGCAGTGCTGCC-3′) and KY261 (5′-GCGCAGAAGTGGTCCTGC-3′). To detect the RNA-DNA hybrid structure in mouse chromosomes, we used native primer KY267 (5′-CAGGAGAGCATAGGGGACCTGG-3′), which was located upstream of the $S_{\gamma 3}$ region, and converted primer (which matched the bisulfite-converted sequence inside the S region) KY262 (5′-CACACCCTTACCTCTCCTAAAC-3′), which was located at the nineteenth repeat of the C57BL/6 $S_{\gamma 3}$ region (equivalent to the twenty-third $S_{\gamma 3}$ repeat of the BALB/c strain). The native primer used to amplify the mouse $S_{\gamma 2b}$ region was KY312 (5′-GGGAAGGTTCATCGGGAAAGGCTG-3′), and the converted primer used for $S_{\gamma 2b}$ amplification was KY315 (5′-CCCCAaCTCCTCCATAaAaaaTAaa-3′).

***In vitro* transcription and gel retardation assay.** Plasmids containing murine switch sequences were transcribed with either T7 or T3 RNA polymerases (Promega) at 37 °C for 1 h in accordance with the manufacturer's instructions. The transcription reaction was terminated by incubating the reaction at 70 °C for 15 min. We added 100 μg of RNase A per 1 μg of transcribed plasmid and incubated the mixture at 37 °C for 30 min. The transcribed plasmid was loaded on a 1% agarose and resolved by electrophoresis in Tris-borate-EDTA (TBE) buffer (89 mM Tris-borate, 2 mM EDTA, pH 8.4) at 70 V for 2 h. We visualized gels by an ultraviolet illuminator after staining them with ethidium bromide.To measure the RNA length, we carried out *in vitro* transcription of the $S_{\gamma 3}$ S regions on the plasmids with T3 or T7 polymerase and radiolabeled ribonucleotide triphosphates. Unincorporated ribonucleotides were removed by an S200 spin column. Free RNA was removed with RNase A. (Specified controls included RNase H at this point.) The RNA-DNA hybrid complex (the plasmid DNA and any RNA that was protected from RNase A digestion by virtue of being base-paired to the DNA) was re-isolated by electrophoresis through a 1% agarose gel. The complex was excised from the agarose gel and purified from the agarose with a silica matrix (GeneClean III; Bio 101, La Jolla, CA). The RNA that had been associated with the plasmid DNA up to this point was then loaded in denaturing buffer onto a 6% denaturing polyacrylamide gel.

Primer extension and PCR reactions. Before transcription with T7 RNA polymerase, pTW-SS91 was linearized with *Xho*I. We verified formation of the RNA-DNA hybrid structure through the use of an aliquot in which radiolabeled [α-^{32}P]rUTP had been added. After transcription, the reaction products were treated with RNase A and passed through a Sephacryl S200 column. The DNA was then resuspended in NEB buffer 1 (10 mM bis-Tris propane-HCl (pH 7), 10 mM $MgCl_2$ and 1 mM dithiothreitol) and a threefold molar excess of extension primer FC34 complementary to the nontemplate G-rich strand was added (typically 1.5 pmol of primer to 0.5 of pmol plasmid DNA). The mixture was then incubated for 15 min at 22 °C. Deoxyribonucleotides (dNTPs) were added to a final concentration of 250 μM with 2 U of Sequenase DNA polymerase (USB, Cleveland, OH), and the extension reaction was allowed to proceed at 22 °C for 5 min. We passed the reaction through a Sephacryl S200 column to remove the primers used in the extension reaction. The equivalent of 10 ng of input DNA substrate was used in PCR reactions with the tail primer and the T7 universal primer. Aliquots of PCR mixture were withdrawn after 15, 20, 25, 30 and 35 cycles and resolved by electrophoresis through a 1.5% agarose gel in TBE buffer run for 1 h at 7 V/cm in the presence of 1 μg/μl of ethidium bromide.

Detection of RNA-DNA hybrid formation. Human minichromosomes were obtained as described[47]. Low molecular weight DNA was digested with *Asp*718 and *Bam*HI at 37 °C overnight to release the S region fragment. When RNase H digestion was required, 2 U of RNase H was added to the restriction digestion mix before the overnight incubation. Digested DNA was resolved on a 0.8% agarose gel by electrophoresis in TBE buffer at 70 V for 2 h. We transferred the DNA to a nylon membrane and probed it with a radiolabeled full-length $S_{\gamma 3}$ probe. Gels were exposed to phosphorimager screens, scanned on a Molecular Dynamics Imager 445SI (Sunnyvale, CA) and analyzed with ImageQuant software version 5.0.

Bisulfite modification assay. We mixed 1 μg of plasmid (or transcribed plasmid) or 20 μg of digested genomic DNA in 30 μl of distilled water with 12.5 μl of 20 mM hydroquinone and 457.5 μl of 2.5 M sodium bisulfite (pH 5.2). The mixture was sealed with mineral oil in a 500-μl microcentrifuge tube and incubated for 16 h at 37 °C in the dark. We purified the bisulfite-treated DNA with the Wizard DNA clean-up system (Promega) according to the manufacturer's instructions. Purified bisulfite-treated DNA was desulfonated with 0.3 M NaOH at 37 °C for 15 min. Desulfonated DNA was recovered by ethanol precipitation and resuspended in TE (pH 8.0).We did PCR with bisulfite-modified DNA as a template and either a pair of native primers, or a native primer paired with a 'converted' primer, the sequence of which matched the conversions anticipated owing to deamination of C to U. We resolved the PCR products on agarose gels and recovered the fragment at the correct size. Purified PCR products were cloned with the TOPO-TA cloning kit (Invitrogen). Plasmid DNA from each clone was purified with the GenElute plasmid miniprep kit (Sigma). Sequencing reactions were done with a SequiTherm Excel II sequencing kit (Epigenetics, Madison, WI) and a Primus 96 Plus thermal cycler (MWG Biotech, High Point, NC). We sequence products automatically on a Li-Cor DNA analyzer, model 4200 (Li-Cor, Lincoln, NE). Although the primer sites were also subjected to bisulfite conversion and, as a result, some molecules lost the ability to act as PCR templates, this background conversion was rare and random. All cloned PCR products from a bisulfite-treated DNA template showed at least background conversion (5–20% of the C residues), possibly owing to DNA breathing during the prolonged incubation at 37 °C. These background conversions, as well as conversions resulting from DNA structure deviation, allowed us to determine which of the two template strands was used in the first PCR cycle for a given cloned molecule (**Fig. 2**). For example, when we read a DNA sequence using the top strand as the reference, molecules derived from top strand showed C-to-T conversions, whereas molecules derived from the bottom showed G-to-A changes (**Fig. 2**). Occasionally, we observed molecules that showed C-to-T conversions for one part of the molecule and G-to-A conversions for the other part of the molecule; we called these 'mosaic molecules'. These were molecules that had changed template during the PCR. Changing of template can also occur during conventional PCR reactions, but it is not detected[48]. We did not include mosaic molecules in the analysis and they represented less than 10% of the total molecules. Because of the repetitiveness of the Sγ3 region, changing the template also generated some molecules with deletions (< 10% of molecules); these molecules were also excluded from this study.

Note: Supplementary information is available on the Nature Immunology *website*.

Acknowledgments

We thank F. Huang and M. Principale for assistance; and I. Haworth and J. Lee for discussions. These studies were supported by grants from the National Institutes of Health (M.R.L.). Manuscripts on *in vitro* class switch R-loops and *in vivo* R-loops[49,50] were retracted very shortly after publication in 2000 because of data alteration by R. B. Tracy. There has been no reliance on the data or reagents of R. B. Tracy in this study or in any other studies by the Lieber laboratory.

Competing interests statement

The authors declare that they have no competing financial interests.

Received 13 January; accepted 11 March 2003.

1. Lieber, M.R. Site-specific recombination in the immune system. *FASEB J.* **5**, 2934–2944 (1991).
2. Lieber, M.R. Pathologic and physiologic double-strand breaks: roles in cancer, aging, and the immune system. *Am. J. Path.* **153**, 1323–1332 (1998).
3. Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J. & Schatz, D.G. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.* **18**, 495–527 (2000).
4. Gellert, M. Recent advances in understanding V(D)J recombination. *Adv. Immunol.* **64**, 39–64 (1997).
5. Dunnick, W.A., Hertz, G.Z., Scappino, L. & Gritzmacher, C. DNA sequence at immunoglobulin switch region recombination sites. *Nucleic Acid Res.* **21**, 365–372 (1993).
6. Stavnezer, J. Antibody class switching. *Adv. Immunol.* **61**, 79–146 (1996).
7. Kinoshita, K. & Honjo, T. Unique and unprecedented recombination mechanisms in class switching. *Curr. Opin. Immunol.* **12**, 195–198 (2000).
8. Snapper, C.M. & Finkelman, F.D. Immunoglobulin class switching. In *Fundamental Immunology* (ed. Paul, W.E.) 831–861 (Lippincott-Raven, Philadelphia, 1999).
9. Gritzmacher, C.A. Molecular aspects of heavy-chain class switching. *Crit. Rev. Immunol.* **9**, 173–200 (1989).
10. Daniels, G.A. & Lieber, M.R. 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 (1995).
11. Stavnezer, J. & Sirlin, S. Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching. *EMBO J.* **5**, 95–102 (1986).
12. Coffman, R.L., Lebman, D.A. & Rothman, P. Mechanism and regulation of immunoglobulin isotype switching. *Adv. Immunol.* **54**, 229–270 (1993).
13. Xu, L. *et al.* Replacement of germ-line ε promoter by gene targeting alters control of immunoglobu-

lin heavy chain class switching. *Proc. Natl. Acad. Sci. USA* **90**, 3705–3709 (1993).
14. Bottaro, A. *et al.* S region transcription *per se* promotes basal IgE class switch recombination but additional factors regulate the efficiency of the process. *EMBO J.* **13**, 665–674 (1994).
15. Lorenz, M., Jung, S. & Radbruch, A. Switch transcripts in immunoglobulin class switching. *Science* **267**, 1825–1828 (1995).
16. Hein, K. *et al.* Processing of switch transcripts is required for targeting of antibody class switch recombination. *J. Exp. Med.* **188**, 2369–2374 (1998).
17. Tashiro, J., Kinoshita, K. & Honjo, T. Palindromic but not G-rich sequences are targets of class switch recombination. *Int. Immunol.* **13**, 495–505 (2001).
18. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription. *Science* **292**, 1863–1876 (2001).
19. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364–366 (1988).
20. Dempsey, L.A., Sun, H., Hanakahi, L.A. & Maizels, N. G4 DNA binding by LR1 and its subunits, nucleolin, and hnRNP D, a role fro G-G pairing in immunoglobulin switch recombination. *J. Biol. Chem.* **274**, 1066–1071 (1999).
21. Reaban, M.E. & Griffin, J.A. Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature* **348**, 342–344 (1990).
22. Reaban, M.E., Lebowitz, J. & Griffin, J.A. Transcription induces the formation of a stable RNA.DNA hybrid in the immunoglobulin α switch region. *J. Biol. Chem.* **269**, 21850–21857 (1994).
23. Tian, M. & Alt, F.W. Transcription induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases *in vitro*. *J. Biol. Chem.* **275**, 24163–24172 (2000).
24. Szurek, P., Petrini, J. & Dunnick, W. Complete nucleotide sequence of the murine g3 switch region and analysis of switch recombination sites in two γ3-expressing hybridomas. *J. Immunol.* **135**, 620–626 (1985).
25. Phoenix, P., Raymond, M., Masse, E. & Drolet, M. Roles of DNA topoisomerases in the regulation of R-loop formation *in vitro*. *J. Biol. Chem.* **272**, 1473–1479 (1997).
26. Masse, E., Phoenix, P. & Drolet, M. DNA topoisomerases regulate R-loop formation during transcription of the rrnB operon in *E. coli*. *J. Biol. Chem.* **272**, 12816–12823 (1997).
27. Clark, S.J., Harrison, J., Paul, C.L. & Frommer, M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* **22**, 2990–2997 (1994).
28. Gough, G.W., Sullivan, K.M. & Lilley, D.M. The structure of cruciforms in supercoiled DNA: probing the single-stranded character of nucleotide bases with bisulphite. *EMBO J.* **5**, 191–196 (1986).
29. Hsieh, C.-L. Stability of patch methylation and its impact in regions of transcriptional initiation and elongation. *Mol. Cell. Biol.* **17**, 5897–5904 (1997).
30. Rothman, P. *et al.* Structure and expression of germline immunoglobulin γ3 heavy chain gene transcripts: implications for mitogen and lymphokine directed class switching. *Int. Immunol.* **2**, 621–627 (1990).
31. Lutzker, S., Rothman, P., Pollock, R., Coffman, R. & Alt, F. Mitogen- and IL-4-regulated expression of germ-line Ig γ2b transcripts: evidence for directed heavy chain class switching. *Cell* **53**, 177–184 (1988).
32. Masukata, H. & Tomizawa, J. A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. *Cell* **62**, 331–338 (1990).
33. Carles-Kinch, K. & Kreuzer, K.N. RNA/DNA hybrid formation at a bacteriophage T4 replication origin. *J. Mol. Biol.* **266**, 915–926 (1997).
34. Lee, D.Y. & Clayton, D.A. Properties of a primer RNA-DNA hybrid at the mouse mitochondrial DNA leading-strand origin of replication. *J. Biol. Chem.* **271**, 24262–24269 (1996).
35. Prichard, M. *et al.* Identification of persistent RNA-DNA hybrid structures within the origin of replication of human cytomegalovirus. *J. Virol.* **72**, 6997–7004 (1998).
36. Baker, T.A. & Kornberg, A. Transcriptional activation of initiation of replication from the *E. coli* chromosomal origin: an RNA-DNA hybrid near *oriC*. *Cell* **55**, 113–123 (1988).
37. Griffith, J.D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514 (1999).
38. Roberts, R.W. & Crothers, D.M. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* **258**, 1463–1466 (1992).
39. Ratmeyer, L., Vinayak, R., Zhong, Y., Zon, G. & Wilson, W.D. Sequence specific thermodynamic and structural properties of DNA-RNA duplexes. *Biochemistry* **33**, 5298–5304 (1994).
40. Shinkura, R. *et al.* The influence of transcriptional orientation on endogenous switch region function. *Nature Immunology* advance online publication, 7 April 2003 (doi:10.1038/ni918).
41. Daniels, G.A. & Lieber, M.R. Strand-specificity in the transcriptional targeting of recombination at immunoglobulin class switch sequences. *Proc. Natl. Acad. Sci. USA* **92**, 5625–5629 (1995).
42. Lee, C.G. *et al.* Quantitative regulation of class switch recombination by switch region transcription. *J. Exp. Med.* **194**, 365–374 (2001).
43. Muramatsu, M. *et al.* Class switch recombination and somatic hypermutation require activation-induced cytidine deaminase (AID), a member of the RNA editing cytidine deaminase family. *Cell* **102**, 541–544 (2000).
44. Revy, P. *et al.* Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell* **102**, 565–575 (2000).
45. Peersen-Mahrt, S.K., Harris, R.S. & Neuberger, M.S. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99–103 (2002).
46. DiNoia, J. & Neuberger, M.S. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43–48 (2002).
47. Hsieh, C.-L., McCloskey, R.P. & Lieber, M.R. V(D)J recombination on minichromosomes is not affected by transcription. *J. Biol. Chem.* **267**, 5613–5619 (1992).
48. Ford, J.E., McHeyzer-Williams, M.G. & Lieber, M.R. Chimeric molecules created by gene amplification interfere with the analysis of somatic hypermutation of murine immunoglobulin genes. *Gene* **142**, 279–283 (1994).
49. Tracy, R.B. & Lieber, M.R. Transcription-dependent R-loop formation at mammalian class switch sequences. *EMBO J.* **19**, 1055–1067 (2000).
50. Tracy, R.B., Hsieh, C. & Lieber, M.R. Stable RNA/DNA hybrids in the mammalian genome: inducible intermediates in immunoglobulin class switch recombination. *Science* **288**, 1058–1061 (2000).