CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the *Igh* locus and antisense transcription in pro-B cells

Stephanie C. Degner[a,1], Jiyoti Verma-Gaur[a,1], Timothy P. Wong[a], Claudia Bossen[b], G. Michael Iverson[c], Ali Torkamani[d], Christian Vettermann[e], Yin C. Lin[b], Zhongliang Ju[f], Danae Schulz[e], Caroline S. Murre[b], Barbara K. Birshtein[f], Nicholas J. Schork[d], Mark S. Schlissel[e], Roy Riblet[c], Cornelis Murre[b], and Ann J. Feeney[a,2]

[a]Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, CA 92037; [b]Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093; [c]Torrey Pines Institute for Molecular Studies, San Diego, CA 92121; [d]Department of Molecular and Experimental Medicine, The Scripps Research Institute, and The Scripps Translational Science Institute, La Jolla, CA 92037; [e]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; and [f]Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY 10461

Edited by Max D. Cooper, Emory University, Atlanta, GA, and approved April 30, 2011 (received for review January 6, 2011)

Compaction and looping of the ~2.5-Mb *Igh* locus during V(D)J rearrangement is essential to allow all V_H genes to be brought in proximity with D_H-J_H segments to create a diverse antibody repertoire, but the proteins directly responsible for this are unknown. Because CCCTC-binding factor (CTCF) has been demonstrated to be involved in long-range chromosomal interactions, we hypothesized that CTCF may promote the contraction of the *Igh* locus. ChIP sequencing was performed on pro-B cells, revealing colocalization of CTCF and Rad21 binding at ~60 sites throughout the V_H region and 2 other sites within the *Igh* locus. These numerous CTCF/cohesin sites potentially form the bases of the multiloop rosette structures at the *Igh* locus that compact during Ig heavy chain rearrangement. To test whether CTCF was involved in locus compaction, we used 3D-FISH to measure compaction in pro-B cells transduced with CTCF shRNA retroviruses. Reduction of CTCF binding resulted in a decrease in *Igh* locus compaction. Long-range interactions within the *Igh* locus were measured with the chromosomal conformation capture assay, revealing direct interactions between CTCF sites 5′ of *DFL16* and the 3′ regulatory region, and also the intronic enhancer (Eμ), creating a D_H-J_H-Eμ-C_H domain. Knockdown of CTCF also resulted in the increase of antisense transcription throughout the D_H region and parts of the V_H locus, suggesting a widespread regulatory role for CTCF. Together, our findings demonstrate that CTCF plays an important role in the 3D structure of the *Igh* locus and in the regulation of antisense germline transcription and that it contributes to the compaction of the *Igh* locus.

lymphocytes | V(D)J recombination | ChIP-seq | insulator | boundary element

Antigen receptors are created through the highly regulated lineage-specific process of V(D)J recombination, creating a diverse repertoire of Ig and T-cell receptors. The generation of the mouse Ig heavy chain in pro-B cells begins with D_H-to-J_H rearrangement on both alleles, followed by V_H-to-D_HJ_H rearrangement. In order for the >100 functional murine V_H genes spread across ~2.5 Mb to gain access to the single D-J rearrangement on that allele, the *Igh* locus undergoes contraction and looping during the pro–B-cell stage of B-cell differentiation (1–5). By measuring spatial distances between 11 small probes spread throughout the *Igh* locus, Jhunjhunwala et al. (2) demonstrated that distal and proximal V_H genes were approximately equidistant from the D genes specifically at the pro–B-cell stage when the V_H genes are rearranging. Computational as well as geometrical approaches have suggested that the locus is organized into rosette-like clusters of loops that compact during rearrangement. Several proteins have been reported to influence *Igh* locus compaction, including Pax5, YY1, and Ikaros (5–7). These proteins and others, such as Ezh2 (8), are also necessary for the rearrangement of distal V_H genes but not proximal V_H genes. This is most likely a consequence of the lack of locus compaction in the absence of these proteins. How all these proteins function and possibly interact to control distal V_H gene rearrangement and *Igh* locus compaction is not yet elucidated.

In addition to the role of these factors in controlling V_H gene rearrangement and locus compaction, proteins involved in higher order chromatin structure and nuclear architecture may be involved. We have hypothesized that the CCCTC-binding factor (CTCF)/cohesin complex may play an important role in antigen receptor locus compaction (9). CTCF is a zinc finger protein that confers insulator function, and it also has been shown to have structural and functional roles in chromatin organization (10, 11). CTCF creates long-range cell type-specific loops at many loci, including *Igf2*/*H19*, β-globin, and IFN-γ (10–15). Cohesin proteins have an established role in sister chromatid cohesion (16) but also participate with CTCF to perform a variety of functions, including transcriptional insulation and long-range chromosomal interactions and looping, presumably by reinforcing the large-scale loops created by CTCF (16–19). Because the CTCF/cohesin complex organizes the 3D structure of the genome by creating long-range loops, we hypothesized that the CTCF/cohesin complex may contribute to the formation of the proposed multiloop 3D structure of the *Igh* locus and of the contracted structure of the *Igh* locus in pro-B cells. If this hypothesis were true, a prerequisite would be that there would be many CTCF binding sites throughout the V_H locus. Indeed, we previously reported >50 sites of CTCF binding throughout the V_H locus in the pro–B-cell stage using chromatin immunoprecipitation on chip (ChIP-on-chip), in addition to the CTCF sites originally described in the 3′ regulatory region (3′RR) (9, 20). We also showed that the cohesin subunit Rad21 was colocalized with CTCF at the selected sites that we tested. Here, we report that cohesin binding sites were colocalized with CTCF at the majority of sites throughout the entire *Igh* locus as determined by ChIP sequencing (ChIP-seq). We then investigated whether CTCF is involved in *Igh* locus compaction. We found that knockdown of CTCF de-

Author contributions: S.C.D., J.V.-G., C.V., M.S.S., R.R., C.M., and A.J.F. designed research; S.C.D., J.V.-G., T.P.W., C.B., G.M.I., C.V., Z.J., D.S., C.S.M., and R.R. performed research; A.T. contributed new reagents/analytic tools; S.C.D., J.V.-G., T.P.W., C.B., G.M.I., A.T., C.V., Y.C.L., D.S., B.K.B., N.J.S., M.S.S., R.R., C.M., and A.J.F. analyzed data; and S.C.D., J.V.-G., and A.J.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE26257).

[1]S.C.D. and J.V.-G. contributed equally to this study.

[2]To whom correspondence should be addressed. E-mail: feeney@scripps.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1019391108/-/DCSupplemental.

creased *Igh* locus compaction in pro-B cells as determined by 3D-FISH. The decrease in compaction was significant, although not as extensive as that in $YY1^{-/-}$ pro-B cells, suggesting it is possible that other proteins also contribute to full locus compaction. Furthermore, we demonstrated long-range chromosomal interactions between the CTCF sites flanking the D_H-J_H-C_H enhancer region, creating a D_H-J_H domain, and knockdown of CTCF decreased this interaction. In addition, we showed that knockdown of CTCF increased D_H and V_H region antisense transcription, most noticeably at Pax5-activated intergenic repeat (PAIR) elements (21). Together, these results suggest that CTCF contributes to the regulation of V(D)J recombination by influencing antisense transcription and the spatial conformation of the *Igh* locus.

Results

Cohesin Is Colocalized with CTCF Throughout the *Igh* Locus. Previously, we reported the locations of sites of CTCF binding throughout the *Igh* locus using ChIP-Chip, and we confirmed that 10 of 10 sites within the *Igh* locus also bound the cohesin subunit Rad21, as determined by ChIP and quantitative PCR (9, 20). To determine whether or not Rad21 was colocalized with CTCF throughout the entire *Igh* locus, we performed ChIP-seq for Rad21 and CTCF using freshly isolated pro-B cells from $Rag1^{-/-}$ mice. In the *Igh* locus, the overall pattern of Rad21 binding was very similar to that of CTCF (Fig. S1*A*). In the proximal half of the V_H locus, which includes all V_H families except J558 and 3609, the CTCF/Rad21 binding sites are all within 150 bp of the recombination signal sequences (RSSs) of V_H genes (Fig. S1*B*). In contrast, all the CTCF/Rad21 binding sites within the distal of half of the V_H locus containing the J558 and 3609 V_H gene families were either far upstream of the coding regions or intergenic (Fig. S1*B*). The majority of sites have both CTCF and Rad21 bound, although some have only CTCF bound. The V_HQ52 gene family is unique in that it has CTCF bound without cohesin (Fig. S1*C*). The ChIP-seq study with its increased sensitivity demonstrated that there are even more sites of CTCF binding within the V_H locus than was indicated by our previous ChIP-Chip study, and it is likely that further depth of sequencing would reveal an even higher concordance of CTCF and Rad21 binding (9).

CTCF Knockdown Decreases *Igh* Locus Compaction. Given the placement of CTCF and cohesin binding sites throughout the *Igh* locus, we previously hypothesized that the CTCF/cohesin complex contributes to the formation of the proposed contracted rosette-like *Igh* locus structure (2). To test this hypothesis, we grew $Rag1^{-/-}$ pro-B cells in short-term culture with IL-7 and stem cell factor (SCF) and then transduced them with retroviruses containing either shRNA targeting CTCF or control scrambled shRNAs. The retroviral constructs also contained GFP to allow purification of transduced cells. Four days later, we sorted GFP^+ pro-B cells and analyzed them for residual CTCF mRNA (Fig. 1*A*) and protein expression (Fig. 1*B*). We then used these cells for 3D-FISH analysis along with $YY1^{-/-}$ and $E2A^{-/-}$ pre–pro-B cells. The cells were probed with three differentially labeled BACs hybridizing to the 3′RR, proximal V_H region, and V-D intergenic region (V_H7183), and just upstream of the V_H locus (V_HJ558) (Fig. 1*C*). All the measurements are plotted in Fig. S2 *B–D*. In Fig. 1*D*, we grasped spatial distances obtained for each cell type into three ranges: <0.3 μm, 0.3–0.5 μm, and 0.5–1.5 μm. In control pro-B cells, the relative distances of most alleles separating J558-7183 and 7183-3′RR probes fell into the <0.3-μm class (69% and 72%, respectively), whereas in CTCF knockdown pro-B cells, the percentage of alleles in this class was significantly reduced (55% and 60%, respectively), although not quite as reduced as in $YY1^{-/-}$ pro-B cells (47% and 62%, respectively). Similarly, the relative distances separating distal

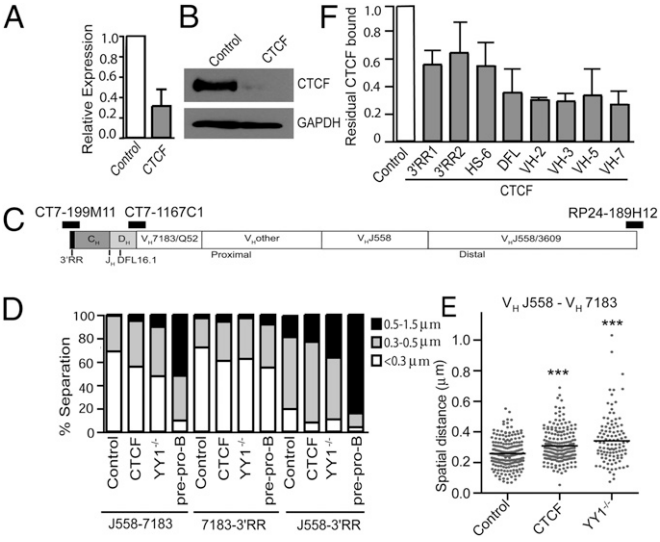

Fig. 1. CTCF knockdown results in decreased *Igh* locus compaction. (*A*) RNA expression in CTCF knockdown and control (scramble shRNA) $Rag1^{-/-}$ pro-B cells. Data are presented as mean ± SEM (n = 3). (*B*) Western blot of CTCF in CTCF knockdown and control pro-B cells. GAPDH served as a loading control. (*C*) Diagram of *Igh* locus indicating the position of the BAC probes. (*D*) *Igh* locus contraction as measured by 3D-FISH in CTCF knockdown and control pro-B cells. $YY1^{-/-}$ pro-B and $E2A^{-/-}$ pre–pro-B cells were also analyzed. The graph represents the percentage of alleles with spatial distances within three ranges: <0.3 μm, 0.3–0.5 μm, and 0.5–1.5 μm. (*E*) Dot plots showing distribution of spatial distances between V_HJ558 and V_H7183 probes. For CTCF knockdown, control, and $YY1^{-/-}$ pro-B cells, 204, 202, and 106 alleles, respectively, were analyzed. ***P < 0.0001 in comparison to control pro-B cells. (*F*) CTCF ChIP in CTCF knockdown and control pro-B cells. Data are presented as mean ± SEM (n = 2).

J558-3′RR were also modestly increased in CTCF knockdown pro-B cells. As expected, the spatial distances in $E2A^{-/-}$ pre–pro-B cells were larger with all the probes. The detailed plot showing overall distribution of spatial distances between J558-7183 probes also demonstrated that the mean spatial distance was increased in CTCF knockdown pro-B cells (0.309 μm) compared with control pro-B cells (0.258 μm), an intermediate value to that in $YY1^{-/-}$ pro-B cells (0.341 μm) (Fig. 1*E* and Table S6). Thus, reduction in CTCF binding results in a modest yet significant decrease in *Igh* locus compaction, although not as extensive as that in $YY1^{-/-}$ pro-B cells. This suggests it is possible that other proteins also contribute to full locus compaction.

Western blotting indicated a large reduction in the total level of CTCF. However, it is likely that some CTCF sites have stronger binding affinity than others, and the reduction in CTCF binding within the *Igh* locus may not be uniform among all the sites. Therefore, we performed ChIP on the sorted GFP^+ pro-B cells that had been transduced with the retroviral constructs (Fig. 1*F*). The results show this to be the case. CTCF is substantially reduced at several sites within the V_H locus and near *DFL16.1*, the most 5′ functional D_H gene but shows more residual binding at the CTCF sites in the 3′RR. This residual level of CTCF could still maintain some looping.

Chromatin Loops Formed Between CTCF Sites at the 3′RR and DFL Region of the *Igh* Locus Create a D_H-J_H Domain. In our ChIP-seq study, we only observed two clusters of CTCF/Rad21 binding outside of the V_H region. The first is a pair of strong CTCF/Rad21 binding sites 3.2 and 5.6 kb 5′ of *DFL16.1* (CTCF/DFL) (Fig. 2*A*). At the 3′RR, there are nine strong sites binding both CTCF and Rad21 spanning ~9 kb downstream of the enhancer portion of the 3′RR (collectively called CTCF/3′RR here). Hy-

IMMUNOLOGY

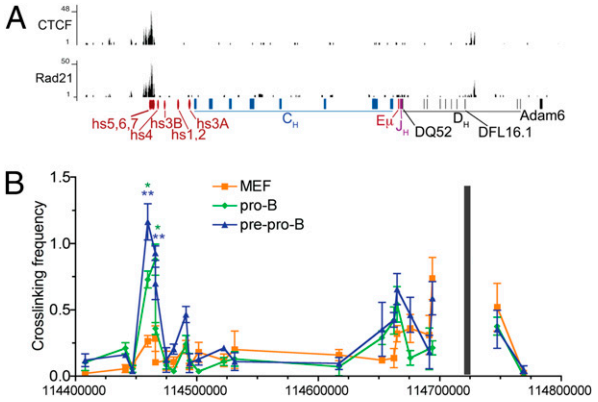

Fig. 2. 3C shows the 3D conformation of the *Igh* locus in *E2A*$^{-/-}$ pre–pro-B cells, *Rag1*$^{-/-}$ pro-B cells, and MEFs. (*A*) Number of reads from the CTCF and Rad21 ChIP-seq experiments in the 3′ portion of the *Igh* locus from the 3′RR to the first V_H gene. A schematic map of the relevant portion of the *Igh* locus is shown, with locations of hs sites (red), constant regions (blue), Eμ (red line), J_H genes (purple lines), and D_H genes (black lines). (*B*) Relative cross-linking frequencies between CTCF/DFL anchor fragment and HindIII fragments within the *Igh* locus in *E2A*$^{-/-}$ pre–pro-B cells, *RAG1*$^{-/-}$ pro-B cells, and MEFs using a CTCF/DFL TaqMan probe (gray bar). Data are presented as mean ± SEM (n = 3). In comparison to MEFs: *P < 0.05; **P < 0.01.

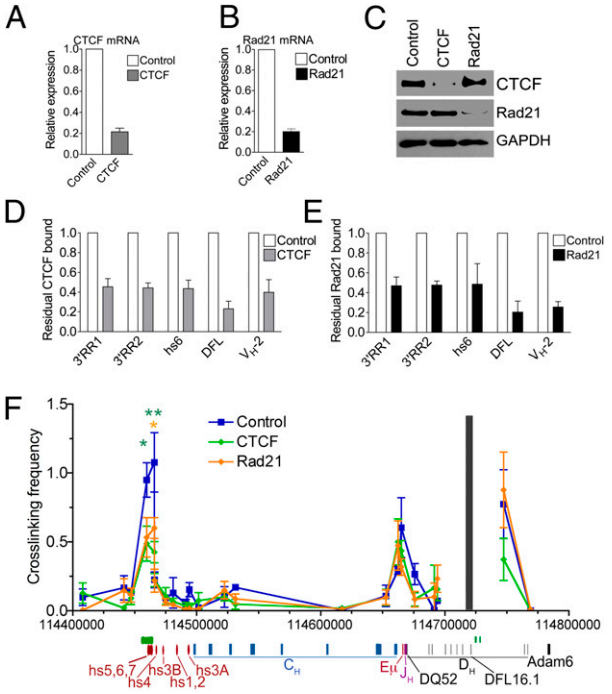

Fig. 3. CTCF and Rad21 knockdown reduces 3D interactions within the *Igh* locus. R2K cells were transduced with control (scramble), CTCF, or Rad21 shRNA retroviruses. Cells were treated with puromycin on days 2–4, and the R2K cells were harvested on day 5 after transduction. Expression levels of (*A*) CTCF and (*B*) Rad21 were measured by quantitative PCR assay, and results were normalized to mouse 18S RNA. Data are presented as mean ± SEM (n = 5). (*C*) Western blot for CTCF and Rad 21. GAPDH was the loading control. CTCF (*D*) and Rad21 (*E*) ChIP assays for CTCF and Rad21 enrichment at selected CTCF sites within the *Igh* locus. Data are presented as mean ± SEM (n = 3 and n = 2, respectively). (*F*) Relative cross-linking frequencies between CTCF/DFL and other HindIII fragments using a CTCF/DFL probe were measured in R2K cells transduced with control, CTCF, and Rad21 shRNA retroviruses. Data are presented as mean ± SEM (n = 4). In comparison to control shRNA: *P < 0.05; **P < 0.01.

persensitive 5 (hs5), hs6, and hs7 each bind CTCF/Rad21, and there are six more CTCF/Rad21 sites extending 6.2 kb downstream of hs7.

We previously hypothesized that the CTCF/DFL sites and CTCF/3′RR sites form a loop creating a domain containing the D_H and J_H genes. We proposed that this loop prevents V_H regions from interacting with the D_H-J_H region before the initiation of V_H-to-DJ_H rearrangement, thus aiding in ordered rearrangement (9). To test whether the CTCF/cohesin sites are involved in long-range interactions within the *Igh* locus, we performed quantitative chromosomal conformation capture (3C) assays using TaqMan probes. We explored interactions between CTCF/DFL and CTCF/3′RR in short-term cultured *Rag1*$^{-/-}$ pro-B cells using a probe at CTCF/DFL, and we found strong interactions between the regions containing CTCF/DFL and CTCF/3′RR (Fig. 2*B*). To determine whether or not these loops were present before the pro–B-cell stage of development, we also tested these interactions in long-term cultured *E2A*$^{-/-}$ pre–pro-B cells and found strong interactions between CTCF/DFL and CTCF/3′RR (Fig. 2*B*). In contrast, we detected minimal interactions between the CTCF/DFL and CTCF/3′RR in murine embryonic fibroblasts (MEFs). These results were confirmed using a TaqMan probe in the 3′RR hs5–7 region (Fig. S3*A*). Interestingly, we also observed looping interactions between CTCF/DFL and the intronic enhancer (Eμ) in both pre–pro-B cells and pro-B cells. In contrast to Eμ, we detected minimal interactions between the CTCF/DFL probe and the other hypersensitive sites of the 3′RR that lack CTCF/Rad21 binding sites, except in *E2A*$^{-/-}$ pre–pro-B cells (Fig. 2*B* and Fig. S3*C*). Thus, CTCF/DFL and CTCF/3′RR sites form loops in pro-B cells and pre–pro-B cells.

Loss of CTCF and Rad21 Has an Impact on Long-Range Chromosomal Interactions of the 3′RR and DFL Region. Next, we examined whether CTCF and Rad21 were required for the observed looping between the 3′RR and DFL sites. To test this, we needed large numbers of cells in which CTCF has been knocked down, precluding the use of sorted GFP$^+$ pro-B cells. We therefore used R2K, an Abelson–murine leukemia virus (A-MuLV) cell line derived from *Rag2*$^{-/-}$ mice on C57BL/6 background. We observed interactions between the CTCF/DFL and CTCF/3′RR sites in these cells (Fig. 3*F*) similar to those we observed in the *Rag1*$^{-/-}$ pro-B cells (Fig. 2*B*); thus, R2K was an appropriate cell line to use for this experiment. We transduced R2K cells with retroviruses expressing control, CTCF, or Rad21 shRNAs. The retroviruses contained a puromycin resistance gene, and R2K cells were selected for 2 d with puromycin after transduction. The cells were harvested 4–5 d after transduction and prepared for 3C analysis. The transduction of the R2K cells led to reduction of CTCF and Rad21 mRNA and protein compared with cells transduced with the retroviruses expressing control shRNA (Fig. 3*A*–*C*). ChIP assays for CTCF and Rad21 showed a two- to threefold reduction in occupancy of CTCF at the CTCF/3′RR and CTCF/DFL sites (Fig. 3 *D* and *E*). The interactions between the CTCF/DFL probe and CTCF/3′RR were reduced more than twofold in R2K cells in which CTCF or Rad21 expression was knocked down (Fig. 3*F* and Fig. S3*D*). To confirm the reduction of interactions between CTCF/DFL and CTCF/3′RR, we also tested these interactions in R2K cells using a TaqMan anchor probe for CTCF/3′RR and obtained similar results (Fig. S3*B*). In contrast, the interactions between CTCF/DFL and Eμ were only minimally reduced in R2K cells in which CTCF or Rad21 levels were knocked down (Fig. 3*F*). These results indicate that loops formed between the 3′RR and DFL regions of the *Igh* are largely mediated by the CTCF/cohesin complex.

CTCF/DFL Has Enhancer-Blocking Activity. CTCF-mediated loop formation is thought to be critical for insulator function (10, 11, 15). Indeed, a region containing the two CTCF/DFL sites was

reported to possess enhancer-blocking activity in an in vitro assay using heterologous promoter/enhancer elements in a nonlymphoid cell line (22). Thus, the CTCF/DFL sites may form an insulator that limits the range of action of Eμ to prevent the transcriptional activation of the V_H region in early B-lineage cells before the DJ recombination step is completed. To test this hypothesis, we stably transfected a pre–B-cell line with a GFP reporter construct consisting of Eμ separated from a V_H promoter (V_HP) by a genomic DNA fragment containing the two CTCF/DFL sites. The construct also contained a phosphoglycerine kinase (PGK)-Neo^R cassette for selection (Fig. S4*A*). Flow cytometry of G418-resistant cells in bulk cultures revealed that the CTCF/DFL region suppressed Eμ-dependent V_HP-driven transcription, resulting in lower median GFP expression, whereas the insertion of an irrelevant stuffer region had no such effect (Fig. S4 *B* and *C*). These results were confirmed by analyzing GFP expression in panels of single-cell clones from each bulk-transfected cell culture (Fig. S4 *D*–*F*). Deletion of the two CTCF/DFL sites abrogated the insulator function of this region, demonstrating the enhancer-blocking or silencing activity of these sites.

CTCF Regulates the Level of Antisense Transcription. Antisense transcription through the D_H locus precedes D_H-to-J_H rearrangement and has been proposed to make the D_H region accessible for subsequent rearrangement (23, 24). Some antisense transcription begins near Eμ and is dependent on the presence of Eμ, whereas other antisense transcription begins near *DST4* (23, 25). Recently, it was shown that antisense transcription decreases just upstream of the CTCF/DFL sites, suggesting that this CTCF region is a boundary that prevents antisense transcription from continuing toward the V_H locus (22). To test whether or not reduction of CTCF would allow the antisense transcription to continue further toward the V_H region, we measured antisense transcription levels in *Rag1*$^{-/-}$ pro-B cells transduced with control or CTCF shRNA retroviruses. Our results revealed that reduction of CTCF leads to an ∼1.5-fold increase in antisense transcription at DFL, which is maintained for another 3.7 kb upstream through the 3′*Adam 6* site. However, by 9 kb upstream of CTCF/DFL, antisense transcription was at the same level as in the control cells, suggesting the presence of a silencer element at or near *Adam 6* (Fig. 4*A*). Thus, decrease of CTCF binding to CTCF/DFL did not result in the extension of antisense transcription into the proximal V_H region. However, the level of transcription within the D_H locus increased (Fig. 4*A*), suggesting a global influence of CTCF on antisense transcription within the D_H-J_H-Eμ domain.

We also tested for the effect of CTCF knockdown on antisense transcription within the V_H locus (Fig. 4*B*). Antisense transcription within the proximal half of the V_HJ558 region was modestly increased. However, a bigger increase in antisense transcription was found in the distal half of the V_HJ558/3609 region near the newly described PAIR elements (21). Sense transcription at V_HJ558 genes was only modestly increased, and the major sense germline transcript, μ°, was unaffected (Fig. 4*C*), demonstrating that CTCF knockdown appeared to have its main effect on modulating antisense transcription.

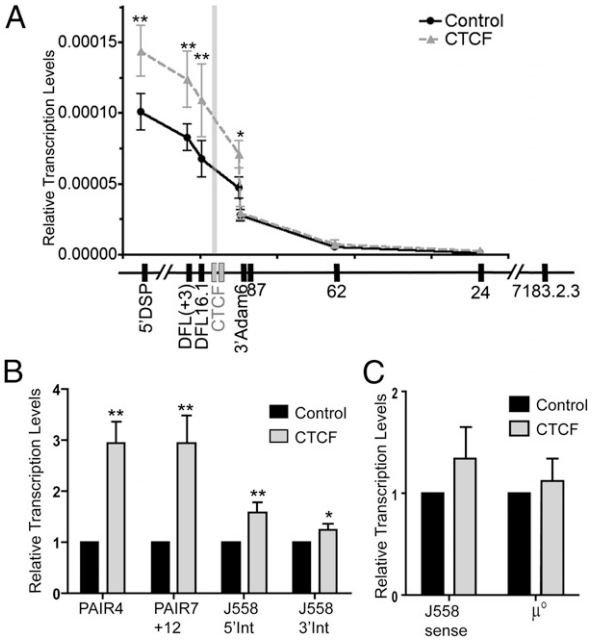

Fig. 4. CTCF knockdown affects the level of antisense transcription in the *Igh* locus. (*A*) Relative antisense transcription levels as measured by quantitative PCR (qPCR) of cultured *Rag1*$^{-/-}$ pro-B cells transduced with control or CTCF shRNAs. The DFL(+3) primer target is located 3 kb downstream of the DFL16.1 gene. The 5′DSP primer targets are located 0.4 kb upstream of the DSP genes. The 24, 62, and 87 primers are located 24, 62, or 87 kb downstream of V_H7183.2.3 (81×). The gray bar indicates the relative location of the CTCF/DFL sites. Results are normalized to mouse 18S RNA and are presented as mean ± SEM (n = 6). (*B*) Relative antisense transcription levels in the V_HJ558 region as measured by qPCR of CTCF knockdown and control *Rag1*$^{-/-}$ pro-B cells. Results are presented as mean ± SEM (n = 5). (*C*) Sense transcription levels from J558 region and μ° were measured by qPCR on CTCF knockdown and control *Rag1*$^{-/-}$ pro-B cells. Results are presented as mean ± SEM (n = 4). In comparison to control pro-B cells: $*P < 0.075$; $**P \leq 0.05$ (details in Table S7).

Discussion

The *Igh* locus is predicted to consist of multiloop rosette-like structures that are present in three domains separated by linker regions in *E2A*$^{-/-}$ pre–pro-B cells and compact to form one domain in pro-B cells (2). Although it is known that the *Igh* locus is compacted in pro-B cells, the proteins that are directly involved in the contraction and looping of the *Igh* locus have not been identified. We previously reported many CTCF binding sites throughout the *Igh* locus, and we proposed that these CTCF sites could form the bases of the multiloop structures of the *Igh* locus and that, together with cohesin, they may aid in locus compaction in the pro–B-cell stage of B-cell development (9). It is possible that proximal V_H regions are recruited by CTCF to surround the cavity of a rosette containing the recombinase (26). To test whether CTCF is involved in locus compaction, we used 3D-FISH to study *Igh* locus compaction in pro-B cells transduced with either control or CTCF shRNA retroviruses. We demonstrated that CTCF does contribute to the formation of the contracted 3D structure of the *Igh* locus in pro-B cells, because knockdown of CTCF results in decreased locus compaction, although the reduction was less than that observed in *YY1*$^{-/-}$ pro-B cells. We attribute this, in part, to the residual CTCF that is left within the *Igh* locus after knockdown (Figs. 1*F* and 3*D*). The ex vivo short-term cultured pro-B cells do not survive well over the long term in the absence of CTCF. At 3 d after knockdown, the percentage of viable GFP^+ cells was the same in the cells transduced with either control or CTCF shRNA containing retroviruses. As the cells progress through days 4 and 5, the percentage of GFP^+ cells drops slightly in cells in which CTCF has been knocked down. Thus, it seems that as the cells completely lose CTCF, they are no longer viable. The cultured *Rag*$^{-/-}$ pro-B cells were even more sensitive to the knockdown of Rad21, precluding analysis of the effect of Rad21 on locus compaction. A second reason why the reduction in locus compaction is modest after CTCF knockdown could be that many proteins are involved in locus compaction, such as Pax5, YY1, Ezh2, and Ikaros (5–8). It is likely that partial reduction of one protein, CTCF, may not be

IMMUNOLOGY

sufficient to undo loops that had been made through the concerted action of more than one protein.

In addition to contributing to the contraction of the *Igh* locus to bring all V_H genes in close proximity to the DJ_H segment to which one V_H gene must rearrange, and in forming the bases of the rosette chromatin structures, CTCF could have a role in creating domains within the *Igh* locus. We previously proposed that CTCF sites could form the base of loops that either exclude or include different regions of the *Igh* locus throughout B-cell development to facilitate specific interactions only at appropriate stages. For instance, we hypothesized that the CTCF/DFL sites may function as a boundary that could separate the D_H and J_H regions from the V_H genes in $E2A^{-/-}$ pre–pro-B cells by looping to the 3′RR and tethering the loop to a site far from the V_H genes (9). This hypothesis was based, in part, on previous studies on *Igh* locus topology (2). One of the probes in that study, h5, was very close to the CTCF/DFL site. In pre–pro-B cells, the h5 probe is located near the 3′RR and is very far away from all the probes in the V_H region. We therefore proposed that this CTCF/DFL-CTCF/3′RR loop would form first and would create a separate domain containing the DJ_H genes. During the pro–B-cell stage, the distal and proximal V_H regions move very close to each other and the D_H probe h5 moves very close to the V_H locus (2). This structural movement of the *Igh* locus would position all V_H genes to be spatially poised for rearrangement to DJ_H.

Our hypothesis was also based, in part, on experiments by Bates et al. (27), which substantiated the idea that the D_H-J_H portion of the *Igh* locus is in a functionally distinct domain from the V_H region. In these experiments, a V_H gene was inserted between CTCF/DFL and DFL16.1 (27). The inserted V_H gene rearranged during the stage of D_H -to-J_H rearrangement. Furthermore, the inserted V_H gene rearranged in thymocytes, which normally only undergo D_H-to-J_H rearrangement but not V_H-to-DJ_H rearrangement (27).

To test the hypothesis that CTCF/DFL and CTCF/3′RR interact, we examined whether loops are formed between these CTCF sites using the 3C assay. In pre–pro-B, pro-B, and R2K pro-B cells, we detected interactions between the CTCF/3′RR and CTCF/DFL sites that were much stronger than interactions seen in MEFs, in which the *Igh* locus is extended. This agrees well with the data demonstrating that the average spatial distances between the BAC downstream of 3′RR and the h5 probe are far greater in MEFs than in $E2A^{-/-}$ pre–pro-B and $Rag^{-/-}$ pro-B cells (2). We also detected interactions between the CTCF/3′RR and CTCF/DFL sites and Eμ in these cells. Eμ has no CTCF sites, but it does contain a YY1 binding site, and we have confirmed by ChIP that YY1 is bound there in pro-B cells (9). Because it has been shown that YY1 and CTCF can interact (28), we propose that it is YY1 binding in Eμ that brings Eμ into this loop. Because we hypothesized that CTCF may allow for long-range chromosomal interactions between the V_H region of the *Igh* locus and the 3′RR and DFL regions, we also tested a selection of primers located near CTCF sites within the V_H locus in 3C, but we were unable to detect interactions with CTCF/3′ RR and CTCF/DFL probes. This was not surprising, because we hypothesize that the interactions among various subsets of the many CTCF sites within the V_H locus are dynamic and are likely to be different in each cell and even to change within a given pro-B cell until a successful V_H-to-DJ_H rearrangement takes place. Thus, they would be very difficult to detect using the 3C assay. In Abelson–MuLV pro-B cells, we demonstrated that the loops formed between CTCF/DFL and CTCF/3′RR were decreased after knockdown of CTCF or Rad21. However, the loop formed with Eμ was only partially reduced. Before the pro–B-cell stage, a CTCF-mediated D_H-J_H-Eμ-C_H loop could be formed. We demonstrated by 3C that this loop remains in the pro–B-cell stage. However, we propose that in pro-B cells, the D_H region loop becomes positioned closer to the V_H locus and interacts in a dynamic fashion with various subsets of the CTCF sites within the 2.5-Mb V_H region, allowing different V_H genes to come within close spatial proximity to the DJ element in each pro-B cell. In this way, a diverse set of V_H genes will rearrange in the population of pro-B cells.

We observed essentially no interactions of CTCF/DFL with hs1,2, the primary enhancer element of 3′RR. The HindIII fragment containing hs3b,4 showed a moderate level of interaction in $E2A^{-/-}$ pre–pro-B cells but low interactions in all the other cell types examined. Because the 3C assay will detect interactions based on close chromosomal proximity, we cannot determine if this low level of interaction at hs4,3b is real or a result of the CTCF-containing hs5–7 being in the adjacent HindIII fragment. These interactions in pro-B cells are very different from those in mature resting and activated B cells, in which hs1,2 as well as hs3b,4 interact extensively with Eμ (29, 30). This difference in looping is not unexpected, because deletion of Eμ greatly reduces V(D)J recombination but does not affect isotype switching (31), whereas deletion of hs3b,4 of 3′RR has the opposite effect (32, 33).

CTCF has been well characterized as the sole protein in vertebrates responsible for insulator activity and for enhancer-blocking activity. Indeed, the CTCF/DFL sites were recently shown to function as insulators in an in vitro assay (22). Experiments presented here confirm this observation in the context of a V_HP and the Eμ enhancer in pro-B cells. One need for a boundary upstream of DFL16 might be to stop the antisense transcription throughout the D_H locus that is observed before D_H-to-J_H rearrangement from continuing on into the V_H locus (23, 24). This D_H antisense transcription has been proposed to create accessibility for D_H-to-J_H rearrangement at a time when the V_H region is inaccessible. The D_H region antisense transcription has recently been shown to drop off just beyond CTCF/DFL (22). We therefore predicted that CTCF knockdown might allow antisense transcription to extend further toward the proximal V_H regions. We observed an increase of antisense transcription for a short distance upstream of CTCF/DFL, but it dropped to control levels at the *Adam6* gene, suggesting that this gene acts as a silencer/boundary for D-region antisense transcription. Surprisingly, we found an increase in antisense transcription from the *Adam6* gene through the *DSP* genes, suggesting a broader regulatory role for CTCF/DFL. Indeed, a very recent study has shown that deletion of a 108-kb intergenic region extending from the proximal V_H genes through *DFL16.1*, including the CTCF/DFL sites, resulted in increased D_H antisense transcription throughout the D_H locus in both B and T cells (25). Because we have shown that CTCF knockdown results in the same increase in D_H antisense transcription, we propose it is the decrease in looping at CTCF/DFL-Eμ-CTCF/3′RR that mediates this enhancing effect on D_H antisense transcription. Because Eμ is required for D_H antisense transcription, it is possible that the decreased looping between the CTCF sites allows Eμ to support D-region antisense transcription better.

Recently, Ebert et al. (21) have described 14 novel regulatory elements in the distal quarter of the V_H locus that consist of juxtaposed sites binding CTCF, E2A, and Pax5, termed PAIR elements. They demonstrate that the PAIR elements direct antisense transcription in a Pax5-dependent manner. Because the CTCF/cohesin site is between the start site of the antisense transcripts and the Pax5/E2A sites, which are presumably in the promoter, CTCF can directly act to regulate this antisense transcription. Indeed, we observed an ~fourfold increase in antisense transcription from PAIR elements after CTCF knockdown. It is possible that Pax5 binding to PAIR may result in a posttranslational modification of CTCF or cohesin, or a change in the composition of complexes binding to CTCF/cohesin, reducing its insulating activity. In contrast, CTCF knockdown did not significantly affect V_HJ558 sense transcription, which is to be expected because there are no CTCF sites near V_HJ558 gene promoters. Also, CTCF knockdown does not affect μ° germline transcription.

In summary, our data indicate that CTCF is involved in the looping of the *Igh* locus during the pro–B-cell stage, during which the *Igh* locus undergoes V(D)J rearrangement. We propose that the CTCF complex forms the bases of the multiloop rosette structure of the *Igh* locus and that it is important in facilitating the compaction of the *Igh* locus during the pro–B-cell stage, most likely in concert with other proteins, such as YY1 and Pax5. We propose that the formation of loops in the V_H region is a dynamic process, in which an ever-changing small subset of CTCF/cohesin sites will be forming loops at any given time, and that this stochastic process will produce different sets of loops in each pro-B cell. In this way, the looping facilitated by the many CTCF/cohesin sites allows for potential rearrangement of V_H genes located throughout the large 2.5-Mb *Igh* locus, thereby permitting the creation of a highly diverse antibody repertoire.

Materials and Methods

Mice and Cell Lines. *Rag1*$^{-/-}$ mice on a C57BL/6 background were obtained from Jackson Laboratories and were maintained at The Scripps Research Institute (TSRI) in accordance with protocols approved by TSRI Institutional Animal Care and Use Committee. *YY1*^{fff} × *mb1-Cre* mice were kindly provided by H. Liu and Y. Shi (Harvard Medical School, Boston, MA) (7). The C57BL/6 MEF cells were obtained from K. Mowen (TSRI). R2K are A–MuLV-derived cell lines from *Rag2*$^{-/-}$ mice on C57BL/6 background and were kindly provided by C. Bassing (University of Pennsylvania, Philadelphia, PA). B6 *Rag1*$^{-/-}$ pro-B cells were isolated and grown as previously described (3). *E2A*$^{-/-}$ pre–pro-B cells were grown as previously described (3, 34).

3C Analysis. The 3C analysis was performed as previously outlined (35). Additional details are provided in *SI Materials and Methods*. Primers are provided in Table S1.

3D-FISH. 3D-FISH was performed as previously described (2). Additional details are provided in *SI Materials and Methods*.

Production and Transduction of CTCF and Rad21 shRNA Retroviruses. Retroviral plasmids containing CTCF shRNA target and control sequences were generously provided by C. Wilson (University of Washington, Seattle, WA) (14). Details of the creation of the Rad21 target construct and the transduction of retroviruses are provided in *SI Materials and Methods* and Table S2. Primers to measure gene expression are listed in Table S3.

Antisense Transcription. Primers for antisense transcription are provided in Table S4.

ChIP and ChIP-Seq. ChIP on pro-B cells was performed largely as previously described (9). Further details are provided in *SI Materials and Methods*, and primers are listed in Table S5. The ChIP-seq dataset is available in the Gene Expression Omnibus (GEO) database under the accession number GSE26257.

Enhancer-Blocking Assay. Linearized insulator reporter constructs were stably transfected by electroporation into an Abelson–MuLV-transformed pre–B-cell line. GFP expression was measured by flow cytometry after 2 wk of selection with G418. Additional details are provided in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Wesley Uykimpang for excellent technical assistance. We thank Dr. Amy Kenter for advice on setting up the TaqMan PCR assay for the 3C experiments. This work was supported by National Institutes of Health Grants R01AI29672 and R01AI82918 (to A.J.F.), R01AI082850 (to C.M.), R01AI40227 (to M.S.S.), R01AI023548 (to R.R.), and R01AI13509 (to B.K.B.), and by UL1 RR025774 (A.T. and N.J.S.). S.C.D. was supported by National Institutes of Health Training Grants F32AI084418 and T32HL07195. C.B. was supported by fellowships from the Swiss National Science Foundation and the European Molecular Biology Organization. C.V. was supported by a Marie Curie International Outgoing Fellowship.

1. Kosak ST, et al. (2002) Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296:158–162.
2. Jhunjhunwala S, et al. (2008) The 3D structure of the immunoglobulin heavy-chain locus: Implications for long-range genomic interactions. *Cell* 133:265–279.
3. Sayegh CE, Jhunjhunwala S, Riblet R, Murre C (2005) Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes Dev*, 19: 322–327 and correction (2008) 22:1717.
4. Roldán E, et al. (2005) Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat Immunol* 6:31–41.
5. Fuxa M, et al. (2004) Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* 18:411–422.
6. Reynaud D, et al. (2008) Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nat Immunol* 9:927–936.
7. Liu H, et al. (2007) Yin Yang 1 is a critical regulator of B-cell development. *Genes Dev* 21:1179–1189.
8. Su IH, et al. (2003) Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat Immunol* 4:124–131.
9. Degner SC, Wong TP, Jankevicius G, Feeney AJ (2009) Cutting edge: Developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. *J Immunol* 182:44–48.
10. Phillips JE, Corces VG (2009) CTCF: Master weaver of the genome. *Cell* 137:1194–1211.
11. Wallace JA, Felsenfeld G (2007) We gather together: Insulators and genome organization. *Curr Opin Genet Dev* 17:400–407.
12. Kurukuti S, et al. (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci USA* 103:10684–10689.
13. Splinter E, et al. (2006) CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* 20:2349–2354.
14. Sekimata M, et al. (2009) CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. *Immunity* 31:551–564.
15. Yusufzai TM, Felsenfeld G (2004) The 5′-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc Natl Acad Sci USA* 101:8620–8624.
16. Nasmyth K, Haering CH (2009) Cohesin: Its roles and mechanisms. *Annu Rev Genet* 43: 525–558.
17. Wendt KS, et al. (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451:796–801.
18. Parelho V, et al. (2008) Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132:422–433.
19. Rubio ED, et al. (2008) CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci USA* 105:8309–8314.
20. Garrett FE, et al. (2005) Chromatin architecture near a potential 3′ end of the igh locus involves modular regulation of histone modifications during B-Cell development and in vivo occupancy at CTCF sites. *Mol Cell Biol* 25:1511–1525.
21. Ebert A, et al. (2011) The distal V(H) gene cluster of the Igh locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells. *Immunity* 34:175–187.
22. Featherstone K, Wood AL, Bowen AJ, Corcoran AE (2010) The mouse immunoglobulin heavy chain V-D intergenic sequence contains insulators that may regulate ordered V(D)J recombination. *J Biol Chem* 285:9327–9338.
23. Bolland DJ, et al. (2007) Antisense intergenic transcription precedes Igh D-to-J recombination and is controlled by the intronic enhancer Emu. *Mol Cell Biol* 27: 5523–5533.
24. Chakraborty T, et al. (2007) Repeat organization and epigenetic regulation of the DH-Cmu domain of the immunoglobulin heavy-chain gene locus. *Mol Cell* 27:842–850.
25. Giallourakis CC, et al. (2010) Elements between the IgH variable (V) and diversity (D) clusters influence antisense transcription and lineage-specific V(D)J recombination. *Proc Natl Acad Sci USA* 107:22207–22212.
26. Lucas JS, Bossen C, Murre C (2010) Transcription and recombination factories: common features? *Curr Opin Cell Biol* 23:1–7.
27. Bates JG, Cado D, Nolla H, Schlissel MS (2007) Chromosomal position of a VH gene segment determines its activation and inactivation as a substrate for V(D)J recombination. *J Exp Med* 204:3247–3256.
28. Donohoe ME, Zhang LF, Xu N, Shi Y, Lee JT (2007) Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell* 25:43–56.
29. Ju Z, et al. (2007) Evidence for physical interaction between the immunoglobulin heavy chain variable region and the 3′ regulatory region. *J Biol Chem* 282:35169–35178.
30. Wuerffel R, et al. (2007) S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. *Immunity* 27:711–722.
31. Perlot T, Alt FW, Bassing CH, Suh H, Pinaud E (2005) Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc Natl Acad Sci USA* 102:14362–14367.
32. Pinaud E, et al. (2001) Localization of the 3′ IgH locus elements that effect long-distance regulation of class switch recombination. *Immunity* 15:187–199.
33. Morvan CL, Pinaud E, Decourt C, Cuvillier A, Cogné M (2003) The immunoglobulin heavy-chain locus hs3b and hs4 3′ enhancers are dispensable for VDJ assembly and somatic hypermutation. *Blood* 102:1421–1427.
34. Ikawa T, Kawamoto H, Wright LY, Murre C (2004) Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. *Immunity* 20:349–360.
35. Hagège H, et al. (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2:1722–1733.