Article

DNA G-Quadruplexes Contribute to CTCF Recruitment

Polina Tikhonova [1] , Iulia Pavlova [1,2], Ekaterina Isaakova [1], Vladimir Tsvetkov [3,4,*] ,
Alexandra Bogomazova [1,5], Tatjana Vedekhina [1], Artem V. Luzhin [6] , Rinat Sultanov [1,5], Vjacheslav Severov [1] ,
Ksenia Klimina [1], Omar L. Kantidze [6] , Galina Pozmogova [1], Maria Lagarkova [1,5,*] and Anna Varizhuk [2,5,*]

[1] Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical
Biological Agency, 119435 Moscow, Russia; tikhonova.polly@mail.ru (P.T.);
pavlova.yui@phystech.edu (I.P.); isaakova.kate@gmail.com (E.I.); abogomazova@gmail.com (A.B.);
neglect1@yandex.ru (T.V.); rhenium112@gmail.com (R.S.); wws83@yandex.ru (V.S.);
ppp843@yandex.ru (K.K.); pozmge@gmail.com (G.P.)
[2] Moscow Institute of Physics and Technology, 141701 Dolgoprudny, Russia
[3] A.V. Topchiev Institute of Petrochemical Synthesis RAS, 119071 Moscow, Russia
[4] WCRC Russia "Digital Biodesign and Personalized Healthcare", Sechenov First Moscow State
Medical University, 119146 Moscow, Russia
[5] Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Federal Research
and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency,
119435 Moscow, Russia
[6] Institute of Gene Biology Russian Academy of Sciences, 119334 Moscow, Russia;
artyom.luzhin@gmail.com (A.V.L.); o_kantidze@mail.ru (O.L.K.)
* Correspondence: v.b.tsvetkov@gmail.com (V.T.); lagar@rcpcm.org (M.L.); annavarizhuk@rcpcm.org (A.V.)

Abstract: G-quadruplex (G4) sites in the human genome frequently colocalize with CCCTC-binding factor (CTCF)-bound sites in CpG islands (CGIs). We aimed to clarify the role of G4s in CTCF positioning. Molecular modeling data suggested direct interactions, so we performed in vitro binding assays with quadruplex-forming sequences from CGIs in the human genome. G4s bound CTCF with Kd values similar to that of the control duplex, while respective i-motifs exhibited no affinity for CTCF. Using ChIP-qPCR assays, we showed that G4-stabilizing ligands enhance CTCF occupancy at a G4-prone site in STAT3 gene. In view of the reportedly increased CTCF affinity for hypomethylated DNA, we next questioned whether G4s also facilitate CTCF recruitment to CGIs via protecting CpG sites from methylation. Bioinformatics analysis of previously published data argued against such a possibility. Finally, we questioned whether G4s facilitate CTCF recruitment by affecting chromatin structure. We showed that three architectural chromatin proteins of the high mobility group colocalize with G4s in the genome and recognize parallel-stranded or mixed-topology G4s in vitro. One of such proteins, HMGN3, contributes to the association between G4s and CTCF according to our bioinformatics analysis. These findings support both direct and indirect roles of G4s in CTCF recruitment.

Keywords: G-quadruplex; chromatin remodeling; CpG methylation; CTCF; HMG proteins

1. Introduction

The guanine quadruplex (G4) structures play a role in DNA replication, transcription, recombination, and DNA repair [1]. G4s may also contribute to epigenetic regulation [2,3] by shaping the DNA methylome [4] and by promoting Yin Yang 1-mediated [5] or CCCTC-binding factor (CTCF)-mediated [6] chromatin looping. G4-chromatin immunoprecipitation sequencing (G4-ChIP-seq) peaks show statistically significant colocalization with CTCF-bound sites [6]. However, the role of G4s in CTCF recruitment has not been explained so far.

Similarly to other zinc-finger (ZF) transcription factors, CTCF binds DNA in a sequencespecific manner, but its cognate sites are rather diverse [7]. A motif that best distinguishes

CTCF binding sites from their flanking regions is a 20-mer with a 14bp core CCGCGNG-GNGGCAG [8]. A different consensus motif, NCANNAG(G/A)NGGC(G/A)(C/G)(T/C), has been revealed by chromatin immunoprecipitation with exonuclease digestion [9]. It is enriched in most CTCF ChIP-seq peaks and partially matches the five-triplet sequence predicted based on the DNA binding specificity of CTCF ZF3-7 (CCAGCAGGGGGCGCT) [10]. None of these sequences match the consensus G4 motif $G_{3+}NG_{3+}NG_{3+}NG_{3+}$, but all of them are G-rich and may overlap G4-prone sites. To clarify whether G4s can be recognized by CTCF, in vitro binding assays are needed.

CTCF sensitivity to DNA methylation is an open question. Hypomethylation was initially shown to promote CTCF binding [11]; but subsequent whole-genome analyses argued against this possibility [12]. Within CTCF binding motifs, methylation effects on the binding efficiency can be negative or positive depending on the position of the methylated CpG [10]. Interestingly, CTCF is frequently recruited to CpG islands (CGIs) [13], which are typically hypomethylated. Within CGIs CTCF binding is mostly invariant and predetermined by sequence [14]. G4s are enriched in CGIs and facilitate their hypomethylation maintenance [4]; thus, they may explain the CGI-CTCF link. This assumption could be verified using bioinformatic approaches.

Methylation is not the only epigenetic factor at play in CTCF positioning. Moderate nucleosome density and regular nucleosome positioning [15,16] are required for CTCF to access to its cognate sites in linker (internucleosomal) DNA. The relation between chromatin compactness and CTCF has not been analyzed thoroughly so far. It is reasonable to assume that internucleosome contacts would hamper CTCF binding to linker DNA similarly to irregular nucleosome positioning and high nucleosome density. Importantly, nucleosome arrangement is controlled by CpG methylation-sensitive histone modifiers and other chromatin remodelers [15,16]. CTCF is also a chromatin remodeler, capable of nucleosome repositioning [17]. Thus, CTCF recruitment to DNA is regulated via several interlinked pathways (Figure 1), and G4s may contribute to each of them. G4s affect methylation via sequestering DNA methyltransferase DNMT1 from its target sites [4] and may promote chromatin remodeling due to their affinity for architectural chromatin proteins and modifiers, including polycomb related complex 2 subunits [18,19].
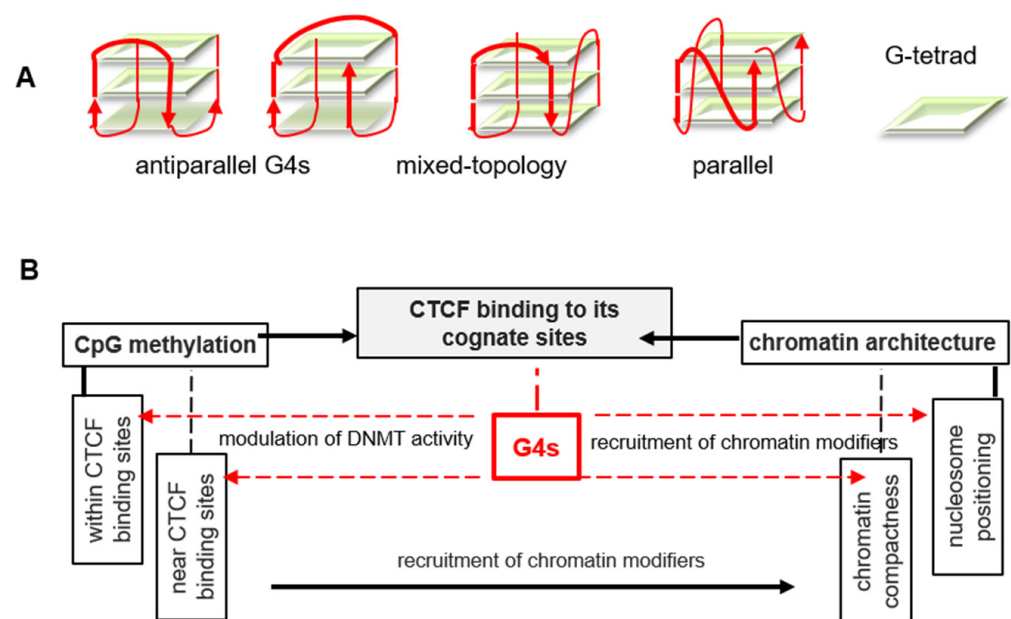

Figure 1. G-quadruplex (G4) structures and their presumed role in CTCF positioning. (**A**) Schematic representation of G4s. (**B**) Summary of the major factors that affects CTCF binding to genomic DNA. Dashed red lines indicate presumed links that are verified in this study.

This study was designed to partially verify the links between G4s, CGIs, chromatin compactness and CTCF recruitment. First, we used ChIP-seq, in silico modeling and in vitro binding assays to test CTCF affinity for G4s. Next, we analyzed previously reported ChIP-seq and bisulfite sequencing data to determine whether G4s account for CTCF recruitment to CGIs and whether DNMT1 plays a substantial role. Finally, to determine whether G4s affect accessibility of CTCF binding sites in internucleosomal DNA, we explored G4 interactions with linker histone analogs and other architectural chromatin proteins.

2. Results

2.1. G4s Colocalize with CTCF-Bound Sites in the Genome and Interact with CTCF In Vitro

CTCF is presumed to be incapable of recognizing G4s [2]. However, evidence against such interactions is limited. In exon 1 of the human telomerase reverse transcriptase (hTERT) gene, G4 formation disrupts CTCF binding, resulting in transcriptional repression [20]. Notably, the CTCF binding site in the hTERT gene is unusual in that it adopts either a hairpin or a G4 structure in vitro depending on the CpG methylation status. The G4/hairpin competition may be an hTERT-specific case rather than a representative example. To get a broader picture, we analyzed genome-wide CTCF occupancy under G4-favoring conditions. For that, we treated K562 cells with a G4 stabilizer pyridostatin (PDS) [21] at a concentration of 10 μM for 24 h and performed ChIP-seq experiments.

The resulting CTCF occupancy profile was similar to that obtained for non-treated (control) cells (Figure 2A,B). Both profiles agreed with the previously reported ChIP-seq data (Figure 2C) [6]. CTCF peaks frequently colocalized with or were flanked by G4 motifs, G4-seq-peaks, and C4-ChIP-seq (BG4) peaks (Figure 2D–F). Importantly, CTCF-bound sites in PDS-treated cells intersected with G4-seq peaks more frequently than in non-treated cells, although colocalization was significant in both cases (Fisher exact test OR = 3.6, p-value $< 1 \times 10^{-300}$ and OR = 3.2, p-value $< 1 \times 10^{-300}$ for PDS-treated and control cells, respectively). In contrast, BG4 peaks were clearly enriched in CTCF-bound sites irrespectively of PDS (Figure 2D).

We next questioned whether prolonged incubation with an excess of G4-stabilizing ligands would enhance CTCF occupancy at G4 sites. For that, we treated K562 cells with 20 μM PDS for 48 h. Phen-DC3, a more potent G4 stabilizer than PDS [22], was tested in parallel at a concentration of 8 μM. Both PDS and Phen-DC3 are presumed to be pan-quadruplex stabilizers [22,23]. However, their effects on conformationally polymorphic G4-prone sites, such as MYC NHE [24], require further investigation. Changes in CTCF occupancy at a G4-prone site (STAT3), G4 clusters colocalized with (BG4) peaks (MYC and VEGFA), and non-G4 sites (USP24 and VPS4A) were evaluated using ChiP-qPCR (the sites are marked with black arrows in Figure 3A; see Table S1 for primer sequences). The ligands increased CTCF occupancy at the G4-prone site, had minor effects on the BG4-positive sites, and did not affect the non-G4 sites (Figure 3B). These data argued against the common assumption that G4 folding attenuates CTCF binding [2], which prompted us to further explore the possibility of G4-CTCF complexes.

We first investigated direct CTCF-G4 interactions in silico. The well-characterized G4 from MYC NHE, previously referred to as Pu27 G4 [25], was selected as a model structure. Docking of MYC-G4 to CTCF yielded a complex that was similar to a previously reported CTCF-duplex complex [10] (Figure 4A). The G4 fit between CTCF ZF4–ZF7, which 'embraced' the G/C-rich B-DNA in reported crystal structures [10], and contacted the protein surface with maximal electrostatic potential. Additional details on the best binding energy conformation of MYC-G4-CTCF and H-bonding in the complex can be found in Figure S1. Briefly, modeling data supported G4 recognition by CTCF.
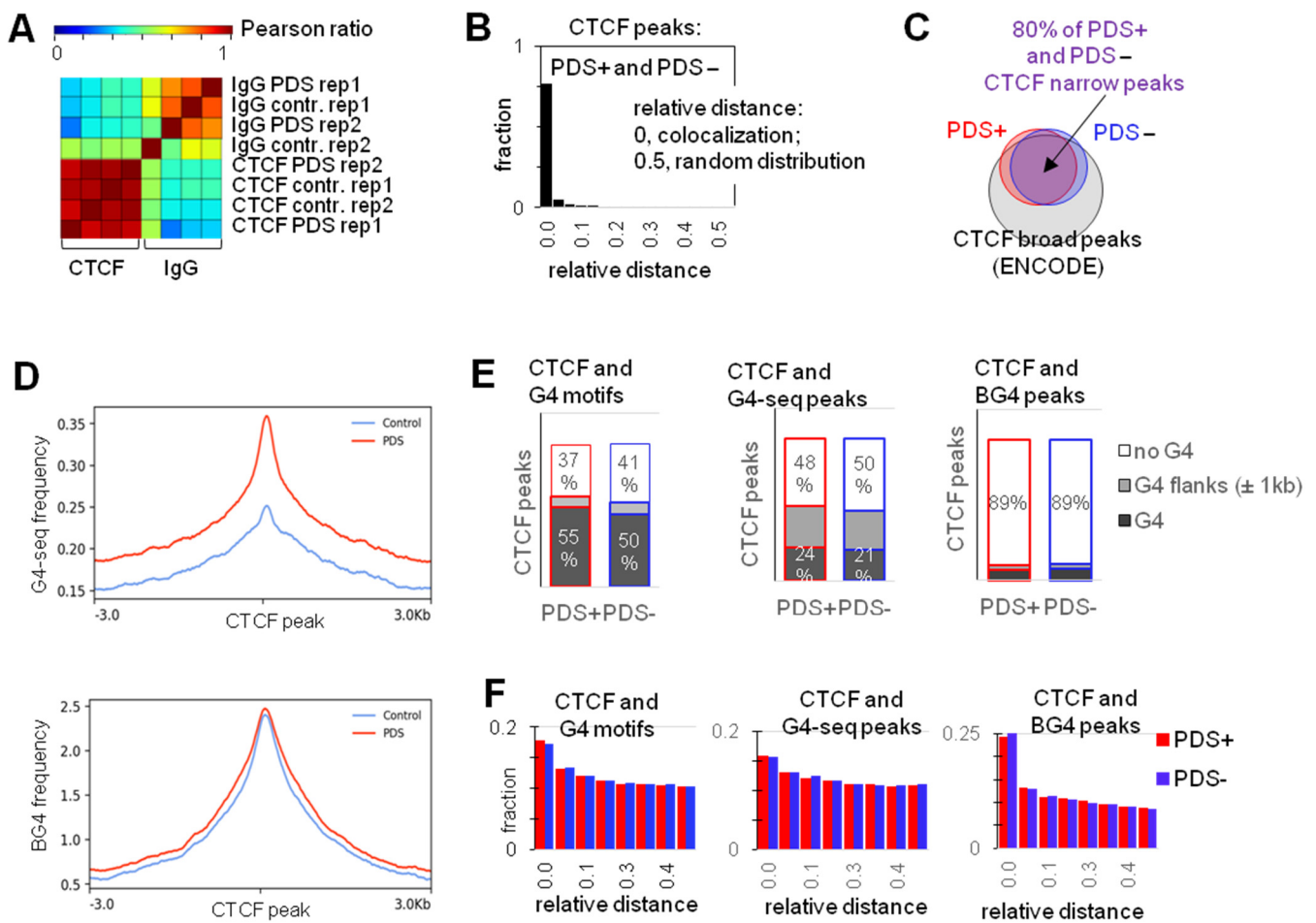
Figure 2. CTCF positioning relative to G4 sites in pyridostatin (PDS)-treated and non-treated cells. (**A**) Summary of the CTCF ChIP-seq data: Pearson correlation between ChIP-seq data sets obtained for PDS-treated (PDS+) and non-treated (PDS−) K562 cells (two biological repeats). (**B**) Relative distance between PDS+ and PDS- CTCF ChIP-seq peaks. (**C**) Venn diagram summarizing the overlap between PDS+/− CTCF peaks and the previously reported CTCF ChIP-seq peaks. (**D**) Metaplots illustrating G4-seq and BG4 peak distribution relative to PDS+ or PDS-(control) CTCF peaks. (**E**) Summary of CTCF peak intersections with G4 sites. (**F**) Relative distances between CTCF peaks and G4 sites.

To verify affinity for CTCF in vitro, we picked three sequences predicted to form stable G4s (red arrows in Figure 3A; Table S2) from MYC, BDNF, and SHANK1 and performed microscale thermophoresis (MST)-based binding assays with fluorophore-labeled CTCF. Respective i-motifs (Table S2) were tested in parallel with the G4s, and the consensus CTCF-binding duplex with a core sequence predicted based on the CTCF ZF3-7 specificity [10] was used as a positive control. Secondary structures of all G4 and i-motif oligonucleotides (ODNs) were confirmed by the characteristic circular dichroism (CD) signatures, i.e., peaks at 265/295 and 288 nm, respectively (Figure 4B). All G4s bound CTCF rather efficiently, with Kd values of 140 ± 60 nM (MYC), 120 ± 30 nM (BDN), and 70 ± 30 nM (SHA). Their affinity was comparable to that of a control duplex (Kd = 80 ± 10 nM), while i-motifs showed no binding (Figure 4C).

This finding agrees with G4 and i-motif distribution in the genome relative to CTCF-bound sites (Figure 4D). Approximately 47% of G4 motifs found within BG4 peaks and only 8% of i-motifs predicted to withstand physiologic conditions [26] colocalized with CTCF-bound sites in K562 cells. We conclude that G4s form complexes with CTCF in vitro and are associated with CTCF in the genome.
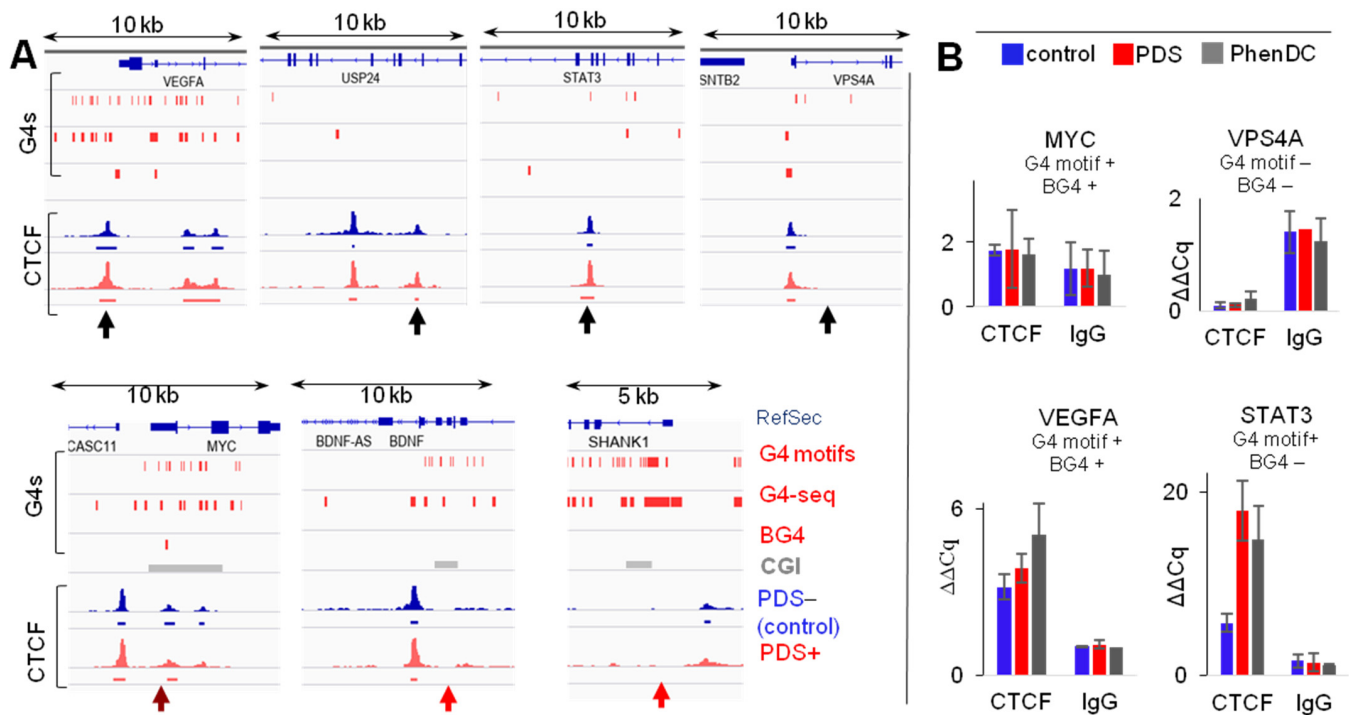
Figure 3. CTCF occupancy at G4 sites in PDS-treated and non-treated cells. (**A**) Genome browser snapshots for VEGFA, USP24, STAT3, VPS4A, MYC, BDNF and SHANK promoter regions/gene bodies harboring G4 motifs, G4-seq peaks, G4 ChIP-seq (BG4) peaks and/or CTCF-bound sites in pyridostatin (PDS)-treated and non-treated K562 cells. Black arrows mark sites analyzed in ChIP-qPCR assays. The CTCF-positive G4-negative site in USP24 was used as a reference, and the CTCF-negative G4-negative site in VPS4A was used as a negative control. Red arrows mark sites predicted to form stable G4s and i-motifs that were used in the binding assays. (**B**) ChIP-qPCR results. The histograms illustrate normalized CTCF occupancy at G4-prone (MYC, VEGFA, and STAT3) and non-G4 (VPS4A) sites relative to the control site (USP24) in non-treated (blue) K562 cells and those treated with PDS (red) or PhenDC3 (grey).

2.2. G4s Account for CTCF Occupancy in CGIs Irrespectively of DNMT1 Inhibition

CTCF binding is CpG methylation-dependent [10] and frequently observed in CGIs [6]. The maintenance of CGI hypomethylation is attributed to DNMT1 inhibition by G4s [4]. Therefore, G4s appear to facilitate CTCF binding to CGI via the DNMT1-dependent mechanism (Figure 5A). We performed statistical analysis of the available data to verify this hypothesis.

Predominant localization of G4s in CGIs, CGI hypomethylation, and partial but significant colocalization between G4s and DNMT1-bound sites have been previously demonstrated [4]. In the K562 cell line, 77% of G4 sites identified by G4-ChIP-seq using the BG4 antibody (representing high-confidence BG4 peaks) overlapped with CGIs, and half intersected with DNMT1-bound sites (Figure 5A, Venn diagram). We investigated distribution of these sites relative to CTCF ChIP-seq peaks and then used X^2 statistics to elucidate whether G4s account for CTCF binding in CGIs.

The frequency of colocalization with CTCF peaks was similar for BG4 peaks inside (63%, 4333/6882) and those outside (62%, 1277/2073) CGIs. On the other hand, BG4-harboring and BG4-lacking CGIs showed markedly different distributions. In most cases (62%, 13249/21399) the latter were CTCF-free, while 9% (1962/21399) flanked and 29% (6188/21399) overlapped with CTCF peaks. In contrast, most BG4-harboring CGIs (72%, 4562/6319) overlapped with CTCF peaks, while 7% (439/6319) flanked CTCF peaks and 21% (1318/6319) were CTCF-free. To summarize, G4s colocalized with CTCF-bound sites irrespectively of CGI presence, while CGIs colocalized with CTCF if they contained G4s (Figure 5B). These results indicate that G4s significantly contribute to the association between CGIs and CTCF ($X^2[1, N = 27,718] = 3848.5$, $p < 0.01$) but not vice versa.
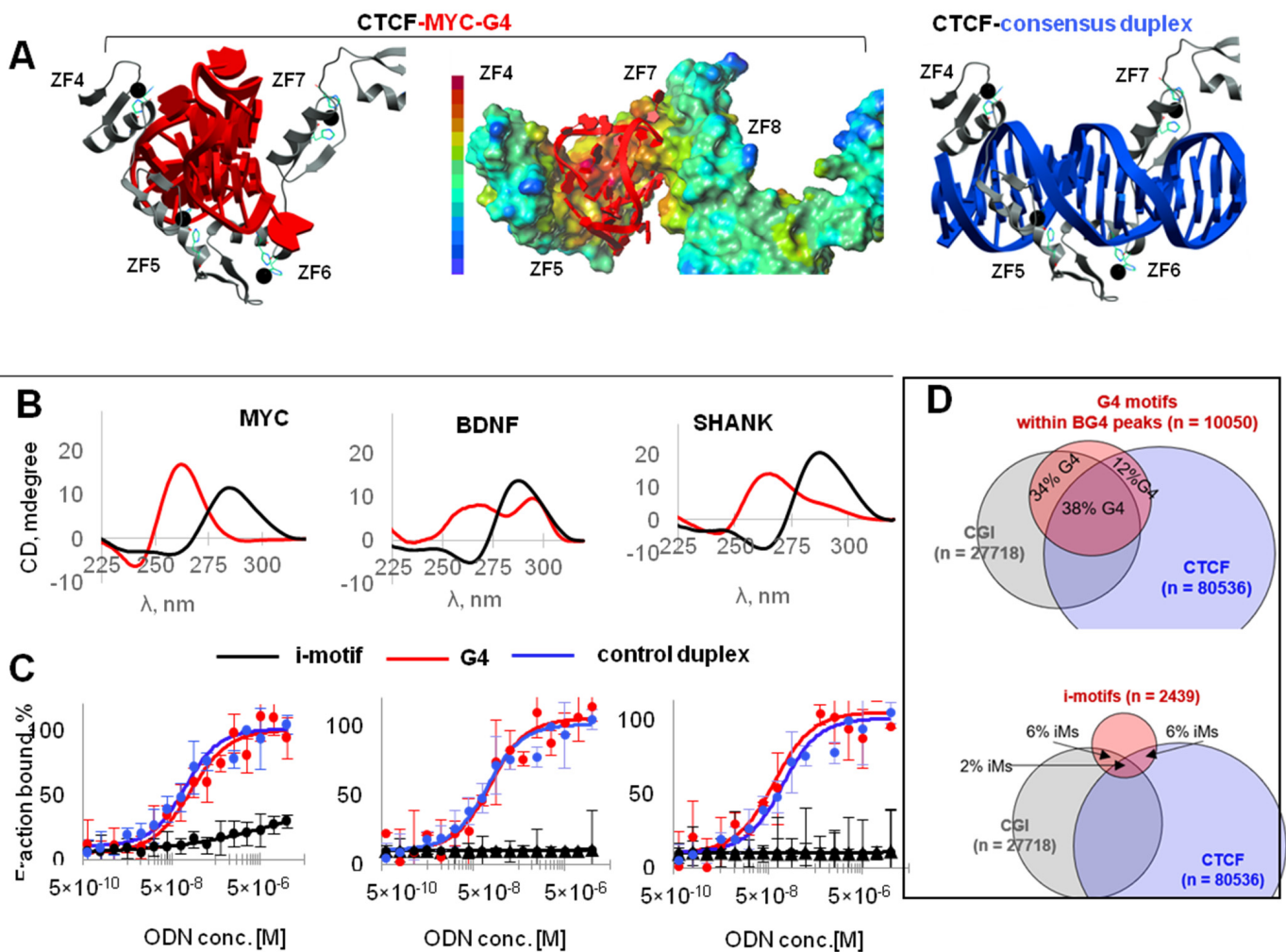
Figure 4. Analysis of CTCF-G4 interactions. (**A**) In silico verification of G4-CTCF binding. The best binding energy conformation of the complex obtained by docking MYC-G4 (red) to CTCF is shown. The complex with the consensus duplex (blue) is shown for comparison. In the middle panel, the surface of MYC-G4-bound CTCF is colored according to the electrostatic potential: from negative (blue) to positive (red). (**B**) Circular dichroism spectra of the 2 μM G4 (red) and iM (black) solutions in 140 mM potassium-phosphate buffer, pH 6.7, supplemented with 10 mM NaCl, obtained at 22 °C. (**C**) Microscale thermophoresis (MST)-based analysis of CTCF interactions the G4s (red), respective i-motifs (black) and the control duplexes (blue). (**D**) Comparison of G4 motif and i-motif intersections with CTCF-bound sites in K562 cells.

We next assessed the contribution of DNMT1. Overlap with CTCF peaks was observed for 62% (858/1385) of DNMT1-lacking BG4-harboring CGIs and 75% (3704/4934) of DNMT1- and BG4-harboring CGIs. Analogously, DNMT1-harboring BG4 peaks within CGIs overlapped CTCF peaks at a slightly higher frequency (68%, 2305/3400) than those lacking DNMT1 (58%, 2028/3482). To summarize, DNMT1-bound sites colocalized with CTCF-bound sites only slightly more frequently than DNMT1-free sites (Figure 5B). This result argues against the decisive role of DNMT1 in CTCF recruitment to CGIs.

Finally, to clarify the role of CpG methylation, we compared methylation levels within CTCF-bound sites, CGIs, and G4 sites in K562 cells using available whole-genome bisulfite sequencing data (Figure 5C). The source data for Figure 5C are provided in Supplementary Archive S1. Methylation level increased in the order CTCF < CGI < BG4 (median: ~10%, 15%, and 30%, respectively).
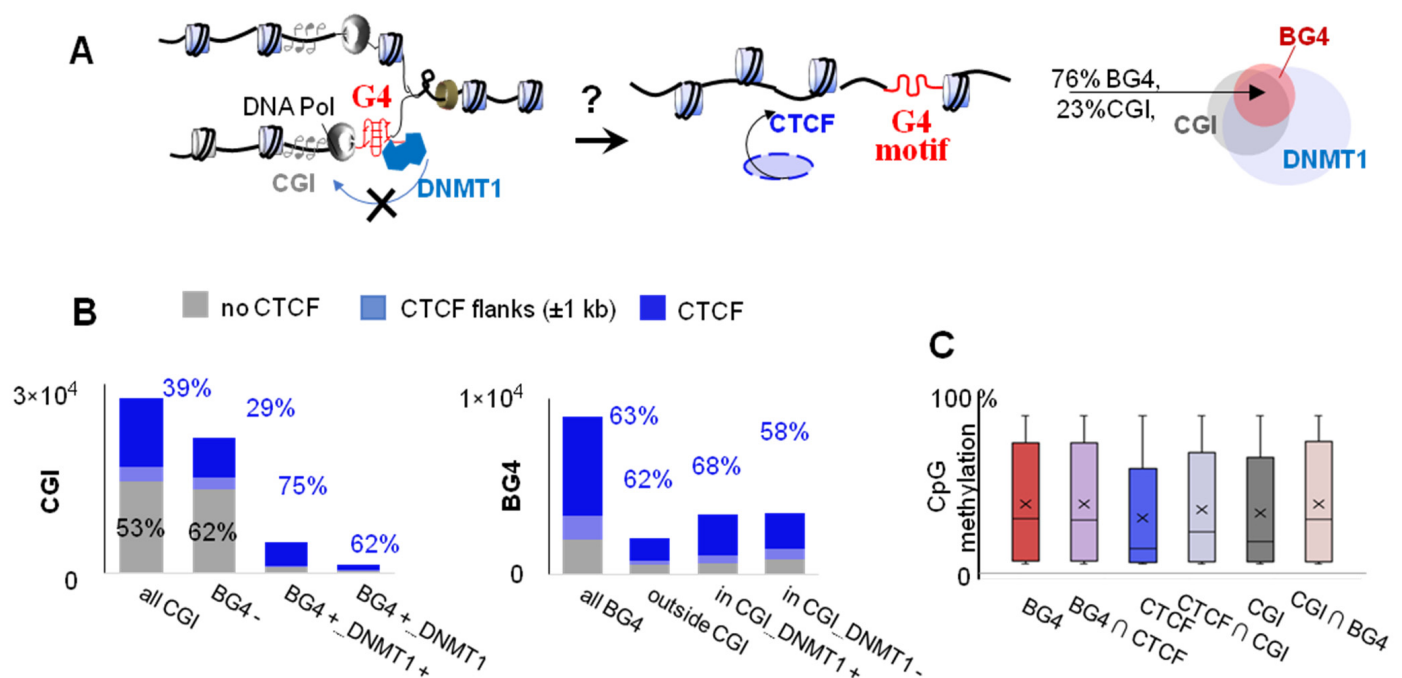
Figure 5. Verification of G4 contribution to the link between CpG islands (CGIs) and CTCF recruitment. (**A**) Hypothetical schemes summarizing the link between G4s, CTCF and CGI methylation: G4s are enriched in CGIs and supposedly protect CGIs from methylation by inhibiting DNMT1, which is favorable for CTCF binding. (**B**) Intersections of CTCF peaks with various subsets of CGIs and BG4 peaks. (**C**) Analysis of average CpG methylation levels in CpG-containing BG4 peaks, CTCF peaks, CGIs and their intersections in K562 cells.

The total average BG4 peak methylation level and those of BG4 peak-overlapping fragments of CTCF and CGI were similar (median: 30%). Although both CGIs and CTCF peaks were hypomethylated, methylation was enhanced at their intersections (mostly BG4-harboring; median: 21%). To summarize, CTCF-bound sites and G4s have slightly higher methylation levels than CGIs in general. These results argue against the decisive role of hypomethylation maintenance in CTCF recruitment to CGIs and supports the role of local methylation.

2.3. G4s May Recruit HMG Proteins That Prevent Chromatin Condensation and CTCF Aggregation

Apart from sequence and methylation, CTCF recruitment depends on nucleosome arrangement [15,16]. We searched for nucleosome regulators among known G4 binders [18,19] and analyzed a group of chromatin architecture-modifying non-histone proteins with intrinsic affinity for bent DNA [27]—high mobility group (HMG) proteins—in a previously reported G4-interactome dataset [19].

We ranked the proteins based on their scores in the reported microarray-based assay with model G4s [19] and picked three hits(HMGN3, HMGN1, and HMGB2) for further analysis (Figure 6A). HMGN1 and HMGN3 are linker histone analogs. They bind to nucleosomal DNA and prevent internucleosome contacts [27,28] to facilitate chromatin decondensation, which renders linker DNA accessible to CTCF. HMGN3 also facilitates chromatin decondensation by recruiting histone acetyltransferases to nearby nucleosomes [29]. HMGB2 typically binds to linker DNA at a nucleosome entry point and acts as a CTCF insulator to prevent its abnormal aggregation [30].
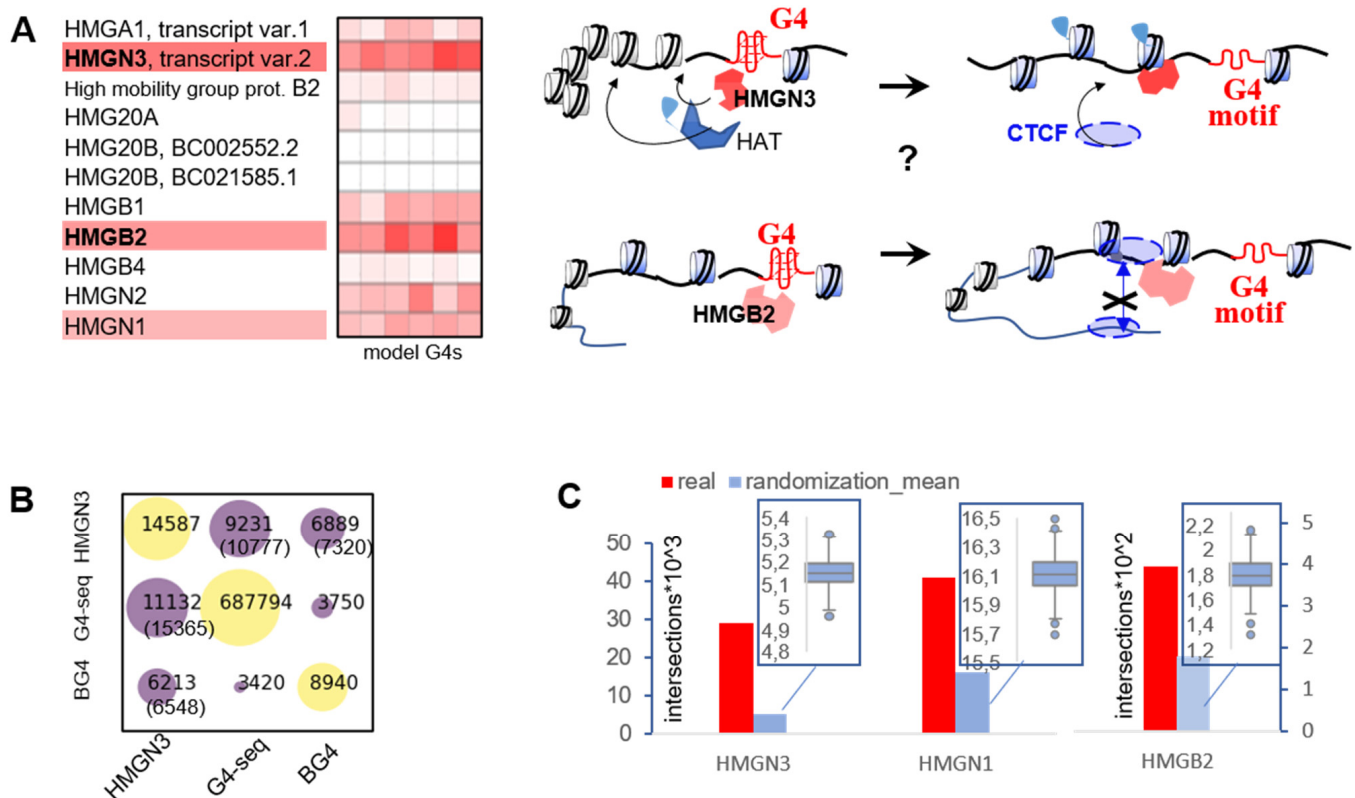
Figure 6. Search for G4-binding chromatin modulators that may affects CTCF positioning. (**A**) A heatmap summarizing G4 binding affinity (while, low; red, high) of HMG proteins in previous assays and hypothetical schemes summarizing the link between G4s, HMG and CTCF: the presumed recruitment of HMGN1 and HMGN3 to G4-prone sites induces chromatin decondensation by the disruption of internucleosome contacts and activation of histone acetyltransferase (HAT). Recruitment of HMGB2 prevents CTCF aggregation. (**B**) Intersection table for HMGN3 ChIP-seq, G4-seq and G4 ChIP-seq (BG4) peaks in K562 cells. The values equal to the number of intersections. Data in parentheses represent the number of G4-seq/BG4 peaks intersecting HMGN3 peaks with flanks (±200 bp). (**C**) Whole-genome analysis of the intersections between G4-seq peaks and HMG protein-bound sites and significance of the G4-HMG correlations (Monte-Carlo simulation results). Red boxes indicate the true number of G4-seq-overlapping HMG protein peaks. The boxplots refer to the randomized peaks.

To verify the association between the HMG proteins and G4s in the human genome, we examined available ChIP-seq data for HMGN3 (K562 cells), HMGN1 (CD4+ T cells), and HMGB2 (IMR90 cells). Average G4-seq coverage of protein-bound sites with flanks (±200 bp to account for the nearest nucleosome) was approximately 22%, 5%, and 4% for HMGN3, HMGN1, and HMGB2, respectively. These values were substantially higher than the average whole-genome G4-seq coverage (1.8%), indicating that G4s are enriched in HMG protein-bound sites. For HMGN3, frequent colocalization with G4s became even more apparent when we switched from G4-seq to BG4 peaks in K562 cells (Figure 6B): 77% of the 8940 BG4s overlapped with HMGN3 peaks and an additional 5% flanked HMGN3 peaks (±200 bp).

HMGB2 showed the least pronounced association with G4s, supposedly because it interacts non-specifically with any DNA kinks [30], and the resulting 'noise signal' partially overshadows specific binding. We assumed non-specific interactions to be minimal at low protein concentrations and analyzed senescent cells (IMR90p28) characterized by low HMGB2 expression levels [31]. G4-seq coverage of HMGB2-bound sites ± 200 bp in these cells (~11%) was almost 3-fold higher than in actively proliferating (IMR90p10) cells (4%), and 7-fold higher than in the whole genome (1.8%). This result suggests that HMGB2 tends to bind G4 sites in the first place.

To assess the significance of the association between G4s and HMG proteins, we first used X^2 statistics. The observed G4-protein peak intersection frequencies exceeded those

predicted by chance ($p < 0.005$), with the X^2 statistic decreasing in the order HMGN3 >> HMGB2 (IMR90p10) > HMGN1 (Figure S2). Next, we performed permutation-based testing with Monte Carlo simulations. A comparison of randomized site intersections and real ones confirmed a non-random distribution in all cases ($p < 0.001$; Figure 6C). Finally, we repeated the permutation-based testing with additional filters applied to the input datasets to ensure that the results were qualitatively independent of the peak-calling procedure (Text box S1).

Given that G4s are associated with active transcription [32] and HMG proteins promote chromatin decondensation [27,28], it is possible that the whole-genome analysis of their relative distributions was biased. We therefore repeated the above experiments using available ATAC-seq data and focusing exclusively on open chromatin regions (Tables S3 and S4, Supplementary Archive S2, and Figure S3). In all cases, the comparison of randomized site intersections with real ones confirmed a non-random distribution ($p < 0.00001$), demonstrating that HMG proteins tend to bind G4 sites both at the whole-genome level and in open chromatin.

2.4. G4s Interact with HMG Proteins and Are Enriched in HMGN3- and CTCF-Bound Sites

The previously reported protoarray assay [19] allowed only semi-quantitative characterization of G4-protein binding; moreover, it was performed with random (model) G4s. We aimed to quantitatively characterize HMG protein binding with representative G4s from the protein occupancy sites and evaluate the binding selectivity. To identify representative G4s, we searched for enriched sequence motifs within HMGN3, HMGB2, and HMGN1-bound sites (Figure S4). In the case of HMGN3, motifs found in both G4-seq peak-intersecting and peak non-intersecting sites included G4-prone sequences or their complements and oligo-T/oligo-A–containing sequences. In the case of HMGB2, we found few motifs, so we also searched for specific sequence patterns in protein peak-intersecting G4 motifs in proliferating and senescent cells using scripts that were developed in-house (Figure S5). For HMGN1, the set of G4-seq peak non-intersecting sites was very large, so we analyzed only G4-seq peak-intersecting sites. The identified motifs included G/C- and A/T-rich and mixed sequences.

Several motif-matching ODNs (Tables S5–S7) were obtained for each protein and analyzed by CD spectroscopy (Figures S6–S8). The ODN sets included duplexes and G4s of various topologies, which enabled preliminary verification of HMG specificity for particular types of secondary structure. In the genomic context, all G4s are presumed to be intramolecular. As such, G4 ODNs that tended to form intermolecular structures (aggregates confirmed by PAGE) were excluded from further analysis (Figures S6–S8). The CD data showed that most G4s had parallel or mixed topologies. We therefore obtained additional ODNs that reportedly adopt antiparallel-stranded quadruplex structures [22CTA [33], HRAS [34], and htel21T18 [35]. Each HMG protein was also tested for binding with non-structured ssDNA (A20 and T20), the model hairpin ds26 [36], and parallel G4 from the microarray assay [19] (positive control). That control G4 has been referred to as G4-2 [19] and CT1 [37] in previous works.

ODN interactions with HMG proteins were analyzed by MST; binding curves and Kd values are shown in Figure 7A and Table 1. Parallel and hybrid G4s bound with HMG proteins in the high nanomolar/low micromolar concentration range. In contrast, antiparallel G4s, motif-matching duplexes, hairpins, and non-structured ODNs showed low or no affinity for HMG proteins. These results explain the partial colocalization of the protein occupancy sites with G4-seq peaks. To summarize, we showed that parallel-stranded and mixed-topology G4s bind with the HMG proteins in vitro and may recruit them in the genome.
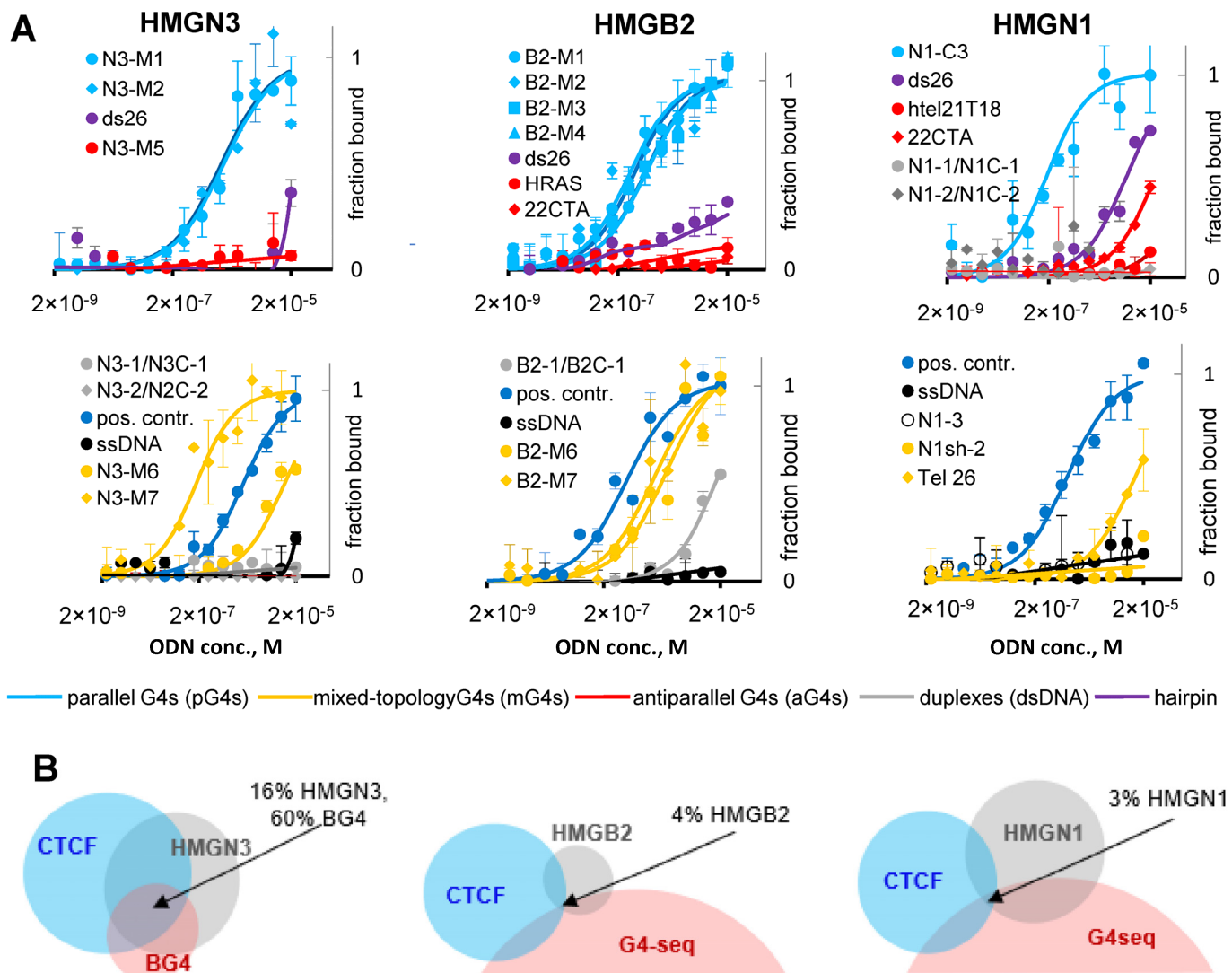
Figure 7. G4-HMG protein interactions and their possible contributions to CTCF positioning. (**A**) Microscale thermophoresis (MST)-based analysis of HMG proteins interactions with representative sequences that comply with enriched motifs (G4s and duplexes), the control hairpin and single-stranded oligonucleotides. Conditions: 50 nM labeled protein, 0–20 μM oligonucleotide, 140 mM potassium-phosphate buffer, pH 7.2 (6.7 for the i-motif N1-3). (**B**) Venn diagrams illustrating the overlaps between G4-ChIP-seq (BG4) peaks, HMGN3-bound sites, and CTCF-bound sites in K562 (left); G4-seq peaks, HMGB2-bound sites, and CTCF-bound sites in IMR-90; and G4-seq peaks, HMGN1-bound sites, and CTCF-bound sites in CD4+ T-cells.

Table 1. Summary of HMG protein interactions with G4s and control oligonucleotides.

ODN	Kd, μM (HMGN3)	Kd, μM (HMGB2)	Kd, μM (HMGN1)
pG4s	1.5 ± 0.6	0.6 ± 0.3	0.2 ± 0.1 (N1–C3); 1.6 ± 0.3 (pos. contr.)
aG4s	>>20	>>20	≥20
mG4s	≥10 (M6); 0.15 ± 0.06 (M7)	4 ± 2 (M6); 1.6 ± 0.6 (M7)	≥10 (Tel 26); >20 (N1sh-2)
hairpin	≥20	≥20	≥10
dsDNA	>>20	>>20	>>20
ssDNA	>>20	>>20	>>20

To verify whether HMG protein recruitment to G4 sites contributes to CTCF binding, we compared relative distributions of HMG and CTCF ChIP-seq peaks. Overall, there was low to moderate colocalization: CTCF peaks overlapped with 53% of HMGN3 peaks, 16% of HMGB2 peaks, and 8% of HMGN1 peaks in the respective cell lines. Substantial portions of the colocalized sites (30% of HMGN3 peaks, 28% of HMGB2 peaks, and 31% of HMGN1 peaks) harbored G4-seq peaks, although overall frequency of triple intersections was low (Figure 7B). Statistical analysis confirmed significance of G4 contribution to the association between HMGN3 and CTCF [X^2(1, N = 37690) = 5574.1, $p < 0.01$]. Switching from G4-seq to G4-ChIP-seq (BG4) peaks increased the frequency of triple intersections (HNGN3-G4-CTCF) in K562 cells. Those triple intersections accounted for 60% of BG4 peaks. We conclude that HMGN3 recruitment to G4 sites may partially contribute to subsequent CTCF positioning.

3. Discussion

We showed that CTCF binds folded G4s and consensus duplexes with comparable affinities (Figure 4). This is hardly surprising, because G4 recognition has been reported previously for other zinc-finger transcription factors with G-rich binding motifs, such as Sp1 [38,39] and MAZ [40]. The latter is functionally rather similar to CTCF, and the two work together as insulators to shape topologically associating domains (TAD) and sub-TAD domains. However, CTCF-G4 and MAZ-G4 complexes are unlikely to exist throughout the G0/G1 phase, except in the cases of persistent (e.g., ligand-stabilized) G4s. The majority of G4s are transient and observed upon replication [41]. Thus, we argue that direct contacts only partially account for CTCF recruitment to G4-prone sites. Histone marks and linker histone analogs that render linker DNA accessible to CTCF may be as important as sequence. We showed that three chromatin modulators (HMGN3, HMGB2, and HMGN1) recognize G4 structures in a topology-specific manner, and the recruitment of HMGN3 to G4 sites may contribute to subsequent CTCF positioning at those sites (Figure 7). As concerns CpG methylation, we argue that global G4-dependent DNMT1 inhibition in CGIs [4] is not crucial for CTCF positioning (Figure 5). However, we do not question the importance of local methylation due to its well-established effects on DNA affinity for zinc fingers [10] or stability of DNA secondary structures [4]. We conclude that a multistep mechanism involving G4-dependent modulation of nucleosome density, positioning, and internucleosome contacts [18,19] and DNA secondary structure-specific CTCF binding may be at play in the genome.

Our findings support G4 involvement in chromatin organization and partially explain G4 enrichment at TAD boundaries [6]. These findings also add to the growing body of links between G4s and the epigenetic machinery, highlighting the prospects of G4s as epigenetic drug targets and raising new concerns about possible side effects of G4-stabilizing small molecule therapeutics [2,42]. Moreover, evidence for G4-CTCF binding adds complexity to the current vision of enhancer-promoter interactions at G4-prone sites [6,43]. One example of such complexity is the MYC case. It is also of particular interest with respect to anticancer strategies based on 3D genome reorganization [44]. In several cancer cell lines, the CTCF binding site upstream of the MYC promoter accounts for long-distance interactions with cancer-specific downstream super-enhancers, which result in elevated MYC expression [45]. In the majority of cell lines, CTCF clustering at the promoter boundary represses MYC transcription, i.e., the CTCF-induced insulation ensures basal rather than elevated expression [46,47]. In K562 cells, we were unable to alter CTCF occupancy in the MYC promoter with G4 ligands, supposedly because G4s (and the cluster of CTCF-bound sites) were already present in the absence of the ligands (Figure 3). It is also possible that PDS failed to stabilize MYC G4, even though it stabilizes the homologous G4 structure Pu22 [22]. The G4-prone site in STAT3 gene that did not coincide with a high confidence BG4 peak and was the most responsive to ligand treatment (Figure 3). Thus, persistent G4s appear to be more challenging than transient ones in terms of fine-tuning with exogenous compounds. We hope that our results along with other emerging evidence

for G4 interference with 3D genome organization [42] will stimulate studies of G4-targeting agents as epigenetic drug candidates.

4. Materials and Methods

4.1. Cell Culture Treatment with G4 Ligands and Chromatin Immunoprecipitation

K562 cells were cultured in RMPI-1640 medium (Paneco, Moscow, Russia) supplemented with 0.4% fetal bovine serum (HyClone GE Healthcare, Greater Milwaukee Area, WI, USA). Cell viability was quantified via trypan blue staining. The G4-stablizing ligands (PDS or PhenDC3) were added to the cell suspensions (1–1.25 mln/mL) to final concentrations of 10 or 20 μM (PDS) and 8 μM (Phen DC3). After 24 h (10 μM PDS) or 48 h (20 μM PDS or 8 μM PhenDC3) of incubation, cells were fixed, and chromatin immunoprecipitation (ChIP) was performed using SimpleChIP® Plus Enzymatic Chromatin IP Kit with magnetic beads (Cell Signalling Technology, Danvers, MA, USA) following the manufacturer's protocole. Briefly, crosslinking was performed via formaldehyde treatment and quenched with glycine. Cells were harvested and nuclei were prepared by incubation with the lysis buffer. Chromatin digestion was performed by treatment with micrococcal nuclease and stopped by EDTA treatment. Next, the lysate was sonicated to obtain DNA fragments of 200–500 bp. Per chromatin immunoprecipitation (ChIP) reaction, ~5–10 μg of digested, cross-linked chromatin was incubated with 2–4 μg CTCF antibody pAb (Active Motif, Carlsbad, CA, USA) overnight at 4 °C. Normal Rabbit IgG (Cell Signaling Technology, Danvers, MA, USA) was used as a negative ChIP control. On the next day, Protein G Magnetic Beads (Cell Signaling Technology, Danvers, MA, USA) were added in each sample and incubated for 6 h at 4 °C. Immobilized complexes were washed two times for 10 min at 4 °C in low salt (1X ChIP buffer) and high salt ($1\times$ ChIP buffer supplemented with 350 mM NaCl) solutions. Samples were incubated with RNase A (Cell Signaling Technology, Danvers, MA, USA) in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) for 30 min at room temperature. The DNA was eluted from the beads and decrosslinked by proteinase K digestion for 2 h at 65 °C. Next, the DNA was purified using DNA purification spin columns and analyzed by Illumina Next Generation sequencing or qRCR.

4.2. ChIP-Seq and Data Analysis

Paired-end libraries were prepared according to the manufacturer's recommendations using NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA). The libraries were indexed with NEBNext Multiplex Oligos kit for Illumina (96 Index Primers, New England Biolabs, Ipswich, MA, USA). Size distribution for the libraries and their quality were assessed by Agilent Bioanalyzer using Agilent DNA High Sensitivity Chips (Agilent Technologies, Santa Clara, CA, USA). The libraries were subsequently quantified by Quant-iT DNA Assay Kit, High Sensitivity (Thermo Scientific, Waltham, MA, USA). DNA sequencing was performed on the HiSeq 2500 platform (Illumina, Madison, WI, USA) according to the manufacturer's recommendations, using the following reagent kits: HiSeq Rapid PE Cluster Kit v2, HiSeq Rapid SBS Kit v2 (200 cycles), HiSeq Rapid PE FlowCell v2 and a 1% PhiX spike-in control. The experiment was repeated twice. Reads for each biological replicate were mapped to the human genome (version hg19) using Bowtie2 (version 2.2.3) with the '——very-sensitive' preset [48]. Non-uniquely mapped reads, PCR duplicates and reads with MAPQ < 30 were filtered out with 'samtools rmdup' and 'samtools view -h -F 256 -q 30'. Peaks were called using PePr (https://github.com/shawnzhangyx/PePr, accessed on 20 December 2020) with a *p*-value cutoff of 0.05 and a sliding window size of 100 bp [49]. The bigWig files were generated using deepTools, version 2.0 [50].

4.3. ChIP-qPCR

For qPCR analysis of the immunoprecipitated samples and inputs, primers intersecting or flanking the CTCF occupancy sites of interest (those intersecting BG4 peaks, G4-seq peaks, G4 motifs or none of the above) were designed using NCBI primer-BLAST and

Eurofins Genomics primer design tools (Table S1). qPCR experiments were performed by QuantStudio5 (Thermofisher Scientific, Waltham, MA, USA) using 96-well white plates. In each experiment, immunoprecipitated samples, primers, and SYBR Green PCR Master Mix—a certified solution of a hot-start DNA polymerase, dNTPs, $MgCl_2$, enhancers, and stabilizers (GenTerra, Russia)—were mixed to give a final volume of 25 μL. The thermocycler program was 95 °C for 5 min (1 cycle) followed by a 2-step reaction of 95 °C for 10 s and 60 °C (for primer pairs STAT3 and VEGFA) or 62 °C (for primer pairs MYC, USP24 and VPS4A) for 30 s (40 cycles). Amplification curves were analyzed using Thermofisher software and ΔΔCq values were calculated in Microsoft Excel to express the relative DNA amplification of the CTCF-occupied region in ligand-treated/non-treated samples to inputs:

ΔΔCq = 2ˆ(−ΔCq_sample)/2ˆ(−ΔCq_input);

ΔCq = Cq_region of interest—Cq_control region

The control region was a G4-lacking site USP24. All primers are specified in Table S1. All experiments were performed in two biological and two technical replicates.

4.4. Molecular Modeling

For G4 docking to CTCF, we used the available model of CTCF zinc fingers 4–10 [10] in complex with a DNA duplex (PDB ID: 5UND) and removed the duplex. The MYC-G4 model, referred to as 'Pu27 truncated' in a previous work (29) was obtained from the Pu24T model (PDB ID: 2N6C) as described previously [51]. A two-step docking procedure was carried out following a published protocol [52]. Briefly, the 1st step included 'rigid' docking using Hex 8.0.0. software [52] and post-processing MM minimization using OPLS force field. At step 2, complexes selected based on the scoring function during post-processing at step 1 were minimized using SYBYL software (Certara, Princeton, NJ, USA) and Powell method. Parameters for interatomic interactions and partial charges on the atoms were taken from Amber7ff02 force field.

4.5. Bioinformatics

G4 motif and i-motif mining was performed using the updated version of imGQFinder [37] and G4/iM-Grinder [53]. G4-seq and G4-ChIP-seq (BG4) data were downloaded from Gene Expression Omnibus. ChIP-seq data for CTCF and other proteins and ATAC-seq data for respective cell lines were downloaded from ENCODE or Gene Expression Omnibus. Data on CpG methylation were obtained from ENCODE. Intersections, evaluation of relative distances, and other manipulations with the sets of genomic intervals were performed using Bedtools [54]. Statistical significance of the intersections between G4 sites and protein-bound sites was verified by chi-squared tests, Fisher tests and permutation-based tests. Enriched motifs in the protein-bound peaks and individual matching sequences were identified using MEME [55] and FIMO [56] algorithms. Details on all bioinformatics analyses are given in Supplementary Text box S1.

4.6. Oligonucleotides, Recombinant Proteins, and Small-Molecule Ligands

Oligonucleotides (ODNs) were obtained from Litekh, Moscow, Russia (purity ≥ 95%, HPLC). Recombinant mouse CTCF with polyhistidine (His) tag was purchased from My-BioSource, San Diego, CA, USA. Recombinant human HMGB2 with His tag was purchased from Abcam, Cambridge, UK. Recombinant human HMGN1 and HMGN3 with His tags were purchased from LifeSpan BioSciences, Seattle, WA, USA. For microscale therophoresis (MST) assays, the proteins were labeled with the His-Tag Labeling Kit RED-tris-NTA (NanoTemper Technologies, Munich, Germany) according to the manufacturer's protocol. G4 ligands pyridostatin (PDS) and bisquinolinium-derivatized phenanthroline-dicarboxamide (Phen DC3) were obtained from Sigma-Aldrich (St. Louis, MO, USA).

4.7. Circular Dichroism Spectroscopy and Electrophoresis

ODN solutions (40 μM) in 140 mM potassium phosphate buffer (pH 7.2 or 6.7) containing 10 mM NaCl were heated to 95 °C for five min and ice-cooled on ice (for G4s and i-motifs) or cooled slowly to room temperature (for duplexes) to facilitate correct folding. These annealed samples were used as stock solutions for microscale thermophoresis assays. For circular dichroism (CD) measurements and polyacrylamide gel electrophoresis (PAGE), the pre-annealed samples were diluted with the respective buffers to a final concentration of 1–5 μM. CD spectra were recorded at room temperature using a Chirascan spectrophotometer (Applied Photophysics, Leatherhead, UK) and a quartz cuvette with an optical path length of 10 mm. Nondenaturing polyacrylamide gel electrophoresis (PAGE) was performed in a standard Tris–borate–EDTA (TBE) buffer (pH 8) at a gel concentration of 20%. Low molecular weight marker 10–100-nt ssDNA (Affymetrix, Santa Clara, CA, USA) was used as a control. The gels were run for 2 h at 200 V at room temperature with a 1× TBE with 10 mM KCl buffer. Oligonucleotide bands were stained with SYBR Gold (Thermo Fisher Scientific, Waltham, MA, USA) and visualized using a Gel Doc scanner (Bio-Rad, Hercules, CA, USA).

4.8. Microscale Thermophoresis

Labeled proteins were mixed with two-fold serial dilutions of unlabeled ODNs to a final protein concentration of 50 nM and varying ODN concentrations (from 20 μM to 0.61 nM). The mixtures were stored at room temperature for 15 min prior to MST measurements. MST curves were registered using Monolith NT.115, equipped with a RED/GREEN detector, and standard capillaries (NanoTemper, Munich, Germany) at 22 °C with MST monitoring by fluorescence of the labeled protein. The dependence of its normalized fluorescence on the concentration of the unlabeled oligonucleotide was analyzed using MO.Affinity Analysis software (NanoTemper, Munich, Germany). To obtain dissociation constant values, experimental data were fitted to the Kd model.

Data Availability Statement: CTCF ChIP-seq data for PDS-treated and non-treated K562 cells have been deposited with Gene Expression Omnibus under accession number GSE173074. The updated version of ImGQFinder is an open source collaborative initiative available in the GitHub repository (https://github.com/RCPCM-GCB/ImGQFinder, accessed on 15 November 2020).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2020**, *2*, 123–136. [CrossRef]
2. Mukherjee, A.K.; Sharma, S.; Chowdhury, S. Non-duplex G-Quadruplex Structures Emerge as Mediators of Epigenetic Modifications. *Trends Genet.* **2019**, *35*, 129–144. [CrossRef]
3. Varizhuk, A.; Isaakova, E.; Pozmogova, G. DNA G-Quadruplexes (G4s) Modulate Epigenetic (Re)Programming and Chromatin Remodeling Transient Genomic G4s Assist in the Establishment and Maintenance of Epigenetic Marks, While Persistent G4s May Erase Epigenetic Marks. *Bioessays* **2019**, *41*, e1900091. [CrossRef]
4. Mao, S.Q.; Ghanbarian, A.T.; Spiegel, J.; Cuesta, S.M.; Beraldi, D.; Di Antonio, M.; Marsico, G.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol.* **2018**, *25*, 951–957. [CrossRef]
5. Li, L.; Williams, P.; Ren, W.; Wang, M.Y.; Gao, Z.; Miao, W.; Huang, M.; Song, J.; Wang, Y. YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat. Chem. Biol.* **2021**, *17*, 161–168. [CrossRef]
6. Hou, Y.; Li, F.Y.; Zhang, R.X.; Li, S.; Liu, H.D.; Qin, Z.H.S.; Sun, X. Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. *Epigenetics* **2019**, *14*, 894–911. [CrossRef] [PubMed]
7. Ohlsson, R.; Renkawitz, R.; Lobanenkov, V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **2001**, *17*, 520–527. [CrossRef]
8. Kim, T.H.; Abdullaev, Z.K.; Smith, A.D.; Ching, K.A.; Loukinov, D.I.; Green, R.D.; Zhang, M.Q.; Lobanenkov, V.V.; Ren, B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **2007**, *128*, 1231–1245. [CrossRef] [PubMed]
9. Rhee, H.S.; Pugh, B.F. Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution. *Cell* **2011**, *147*, 1408–1419. [CrossRef] [PubMed]
10. Hashimoto, H.; Wang, D.; Horton, J.R.; Zhang, X.; Corces, V.G.; Cheng, X. Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. *Mol. Cell* **2017**, *66*, 711–720.e3. [CrossRef]
11. Wiehle, L.; Thorn, G.J.; Raddatz, G.; Clarkson, C.T.; Rippe, K.; Lyko, F.; Breiling, A.; Teif, V.B. DNA (de)methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries. *Genome Res.* **2019**, *29*, 750–761. [CrossRef]
12. Heberle, E.; Bardet, A.F. Sensitivity of transcription factors to DNA methylation. *Essays Biochem.* **2019**, *63*, 727–741. [PubMed]
13. Vavouri, T.; Lehner, B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biol.* **2012**, *13*, R110. [CrossRef]
14. Teif, V.B.; Beshnova, D.A.; Vainshtein, Y.; Marth, C.; Mallm, J.P.; Hofer, T.; Rippe, K. Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res.* **2014**, *24*, 1285–1295. [CrossRef]
15. Barutcu, A.R.; Lian, J.B.; Stein, J.L.; Stein, G.S.; Imbalzano, A.N. The connection between BRG1, CTCF and topoisomerases at TAD boundaries. *Nucleus-Phila* **2017**, *8*, 150–155. [CrossRef]
16. Barisic, D.; Stadler, M.B.; Iurlaro, M.; Schubeler, D. Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature* **2019**, *569*, 136–140. [CrossRef]
17. Owens, N.; Papadopoulou, T.; Festuccia, N.; Tachtsidi, A.; Gonzalez, I.; Dubois, A.; Vandormael-Pournin, S.; Nora, E.P.; Bruneau, B.G.; Cohen-Tannoudji, M.; et al. CTCF confers local nucleosome resiliency after DNA replication and during mitosis. *Elife* **2019**, *8*, 8. [CrossRef]
18. Makowski, M.M.; Grawe, C.; Foster, B.M.; Nguyen, N.V.; Bartke, T.; Vermeulen, M. Global profiling of protein-DNA and protein-nucleosome binding affinities using quantitative mass spectrometry. *Nat. Commun.* **2018**, *9*, 1–10. [CrossRef]
19. Vlasenok, M.; Levchenko, O.; Basmanov, D.; Klinov, D.; Varizhuk, A.; Pozmogova, G. Data set on G4 DNA interactions with human proteins. *Data Brief* **2018**, *18*, 348–359. [CrossRef]
20. Li, P.T.; Wang, Z.F.; Chu, I.T.; Kuan, Y.M.; Li, M.H.; Huang, M.C.; Chiang, P.C.; Chang, T.C.; Chen, C.T. Expression of the human telomerase reverse transcriptase gene is modulated by quadruplex formation in its first exon due to DNA methylation. *J. Biol. Chem.* **2017**, *292*, 20859–20870. [CrossRef]
21. Rodriguez, R.; Muller, S.; Yeoman, J.A.; Trentesaux, C.; Riou, J.F.; Balasubramanian, S. A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.* **2008**, *130*, 15758–15759. [CrossRef]
22. De Cian, A.; DeLemos, E.; Mergny, J.L.; Teulade-Fichou, M.P.; Monchaud, D. Highly efficient G-quadruplex recognition by bisquinolinium compounds. *Am. Chem. Soc.* **2007**, *129*, 1856–1857. [CrossRef]
23. Muller, S.; Sanders, D.A.; Di Antonio, M.; Matsis, S.; Riou, J.F.; Rodriguez, R.; Balasubramanian, S. Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. *Org. Biomol. Chem.* **2012**, *10*, 6537–6546. [CrossRef] [PubMed]

24. Le, H.T.; Miller, M.C.; Buscaglia, R.; Dean, W.L.; Holt, P.A.; Chaires, J.B.; Trent, J.O. Not all G-quadruplexes are created equally: An investigation of the structural polymorphism of the c-Myc G-quadruplex-forming sequence and its interaction with the porphyrin TMPyP4. *Org. Biomol. Chem.* **2012**, *10*, 9393–9404. [CrossRef] [PubMed]

25. Phan, A.T.; Modi, Y.S.; Patel, D.J. Propeller-type parallel-stranded g-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.* **2004**, *126*, 8710–8716. [CrossRef]

26. Wright, E.P.; Huppert, J.L.; Waller, Z.A.E. Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res.* **2017**, *45*, 2951. [CrossRef] [PubMed]

27. Postnikov, Y.V.; Bustin, M. Functional interplay between histone H1 and HMG proteins in chromatin. *Biochim. Biophys. Acta (BBA)-Bioenerg.* **2016**, *1859*, 462–467. [CrossRef]

28. Murphy, K.J.; Cutter, A.R.; Fang, H.; Postnikov, Y.V.; Bustin, M.; Hayes, J.J. HMGN1 and 2 remodel core and linker histone tail domains within chromatin. *Nucleic Acids Res.* **2017**, *45*, 9917–9930. [CrossRef]

29. Barkess, G.; Postnikov, Y.; Campos, C.D.; Mishra, S.; Mohan, G.; Verma, S.; Bustin, M.; West, K.L. The chromatin-binding protein HMGN3 stimulates histone acetylation and transcription across the Glyt1 gene. *Biochem. J.* **2012**, *442*, 495–505. [CrossRef]

30. Zirkel, A.; Nikolic, M.; Sofiadis, K.; Mallm, J.P.; Brackley, C.A.; Gothe, H.; Drechsel, O.; Becker, C.; Altmuller, J.; Josipovic, N.; et al. HMGB2 Loss upon Senescence Entry Disrupts Genomic Organization and Induces CTCF Clustering across Cell Types. *Mol. Cell* **2018**, *70*, 730–744.e6. [CrossRef] [PubMed]

31. Aird, K.M.; Iwasaki, O.; Kossenkov, A.V.; Tanizawa, H.; Fatkhutdinov, N.; Bitler, B.G.; Le, L.; Alicea, G.; Yang, T.L.; Johnson, B.; et al. HMGB2 orchestrates the chromatin landscape of senescence-associated secretory phenotype gene loci. *J. Cell Biol.* **2016**, *215*, 325–334. [CrossRef]

32. Hansel-Hertsch, R.; Beraldi, D.; Lensing, S.V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **2016**, *48*, 1267–1272. [CrossRef] [PubMed]

33. Lim, K.W.; Alberti, P.; Guedin, A.; Lacroix, L.; Riou, J.F.; Royle, N.J.; Mergny, J.L.; Phan, A.T. Sequence variant (CTAGGG)(n) in the human telomere favors a G-quadruplex structure containing a G center dot C center dot G center dot C tetrad. *Nucleic Acids Res.* **2009**, *37*, 6239–6248. [CrossRef]

34. Membrino, A.; Cogoi, S.; Pedersen, E.B.; Xodo, L.E. G4-DNA Formation in the HRAS Promoter and Rational Design of Decoy Oligonucleotides for Cancer Therapy. *PLoS ONE* **2011**, *6*, e24421. [CrossRef]

35. Liu, C.D.; Zhou, B.; Geng, Y.Y.; Tam, D.Y.; Feng, R.; Miao, H.T.; Xu, N.N.; Shi, X.; You, Y.Y.; Hong, Y.N.; et al. A chair-type G-quadruplex structure formed by a human telomeric variant DNA in K+ solution. *Chem. Sci.* **2019**, *10*, 218–226. [CrossRef] [PubMed]

36. Anselmet, A.; Mayat, E.; Wietek, S.; Layer, P.G.; Payrastre, B.; Massoulie, J. Non-antisense cellular responses to oligonucleotides. *Febs Lett.* **2002**, *510*, 175–180. [CrossRef]

37. Varizhuk, A.; Ischenko, D.; Tsvetkov, V.; Novikov, R.; Kulemin, N.; Kaluzhny, D.; Vlasenok, M.; Naumov, V.; Smirnov, I.; Pozmogova, G. The expanding repertoire of G4 DNA structures. *Biochimie* **2017**, *135*, 54–62. [CrossRef]

38. Raiber, E.A.; Kranaster, R.; Lam, E.; Nikan, M.; Balasubramanian, S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res.* **2011**, *40*, 1499–1508. [CrossRef]

39. Da Ros, S.; Nicoletto, G.; Rigo, R.; Ceschi, S.; Zorzan, E.; Dacasto, M.; Giantin, M.; Sissi, C. G-Quadruplex Modulation of SP1 Functional Binding Sites at the KIT Proximal Promoter. *Int. J. Mol. Sci.* **2020**, *22*, 329. [CrossRef]

40. Xiao, T.J.; Li, X.; Felsenfeld, G. The Myc-associated zinc finger protein (MAZ) works together with CTCF to control cohesin positioning and genome organization. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, 7. [CrossRef]

41. Valton, A.L.; Prioleau, M.N. G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends Genet.* **2016**, *32*, 697–706. [CrossRef]

42. Reina, C.; Cavalieri, V. Epigenetic Modulation of Chromatin States and Gene Expression by G-Quadruplex Structures. *Int. J. Mol. Sci.* **2020**, *21*, 4172. [CrossRef] [PubMed]

43. Hegyi, H. Enhancer-promoter interaction facilitated by transiently forming G-quadruplexes. *Sci. Rep.* **2015**, *5*, 9165. [CrossRef]

44. Kantidze, O.L.; Gurova, K.V.; Studitsky, V.M.; Razin, S.V. The 3D Genome as a Target for Anticancer Therapy. *Trends Mol. Med.* **2020**, *26*, 141–149. [CrossRef]

45. Schuijers, J.; Manteiga, J.C.; Weintraub, A.S.; Day, D.S.; Zamudio, A.V.; Hnisz, D.; Lee, T.I.; Young, R.A. Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism. *Cell Rep.* **2018**, *23*, 349–360. [CrossRef]

46. Lutz, M.; Burke, L.J.; Barreto, G.; Goeman, F.; Greb, H.; Arnold, R.; Schultheiss, H.; Brehm, A.; Kouzarides, T.; Lobanenkov, V.; et al. Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res.* **2000**, *28*, 1707–1713. [CrossRef]

47. Filippova, G.N.; Fagerlie, S.; Klenova, E.M.; Myers, C.; Dehner, Y.; Goodwin, G.; Neiman, P.E.; Collins, S.J.; Lobanenkov, V.V. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell. Biol.* **1996**, *16*, 2802–2813. [CrossRef]

48. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]

49. Zhang, Y.X.; Lin, Y.H.; Johnson, T.D.; Rozek, L.S.; Sartor, M.A. PePr: A peak-calling prioritization pipeline to identify consistent or differential peaks from replicated ChIP-Seq data. *Bioinformatics* **2014**, *30*, 2568–2575. [CrossRef]

50. Ramirez, F.; Ryan, D.P.; Gruning, B.; Bhardwaj, V.; Kilpert, F.; Richter, A.S.; Heyne, S.; Dundar, F.; Manke, T. deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **2016**, *44*, W160–W165. [CrossRef] [PubMed]

51. Pavlova, I.I.; Tsvetkov, V.B.; Isaakova, E.A.; Severov, V.V.; Khomyakova, E.A.; Lacis, I.A.; Lazarev, V.N.; Lagarkova, M.A.; Pozmogova, G.E.; Varizhuk, A.M. Transcription-facilitating histone chaperons interact with genomic and synthetic G4 structures. *Int. J. Biol. Macromol.* **2020**, *160*, 1144–1157. [CrossRef] [PubMed]

52. Macindoe, G.; Mavridis, L.; Venkatraman, V.; Devignes, M.D.; Ritchie, D.W. HexServer: An FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.* **2010**, *38*, W445–W449. [CrossRef]

53. Belmonte-Reche, E.; Morales, J.C. G4-iM Grinder: When size and frequency matter. G-Quadruplex, i-Motif and higher order structure search and analysis tool. *NAR Genom. Bioinform.* **2020**, *2*, lqz005. [CrossRef] [PubMed]

54. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [CrossRef] [PubMed]

55. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.Y.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef]

56. Grant, C.E.; Bailey, T.L.; Noble, W.S. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **2011**, *27*, 1017–1018. [CrossRef]