The Myc-associated zinc finger protein (MAZ) works together with CTCF to control cohesin positioning and genome organization

Tiaojiang Xiaoa, Xin Lia, and Gary Felsenfelda,1

*aLaboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, MD 20892-0540

Contributed by Gary Felsenfeld, December 30, 2020 (sent for review November 6, 2020; reviewed by Peter Fraser and Tom Maniatis)

The Myc-associated zinc finger protein (MAZ) is often found at genomic binding sites adjacent to CTCF, a protein which affects large-scale genome organization through its interaction with cohesin. We show here that, like CTCF, MAZ physically interacts with a cohesin subunit and can arrest cohesin sliding independently of CTCF. It also shares with CTCF the ability to independently pause the elongating form of RNA polymerase II, and consequently affects RNA alternative splicing. CTCF/MAZ double sites are more effective at sequestering cohesin than sites occupied only by CTCF. Furthermore, depletion of CTCF results in preferential loss of CTCF from sites not occupied by MAZ. In an assay for insulation activity like that used for CTCF, binding of MAZ to sites between an enhancer and promoter results in down-regulation of reporter gene expression, supporting a role for MAZ as an insulator protein. Hi-C analysis of the effect of MAZ depletion on genome organization shows that local interactions within topologically associated domains (TADs) are disrupted, as well as contacts that establish the boundaries of individual TADs. We conclude that MAZ augments the action of CTCF in organizing the genome, but also shares properties with CTCF that allow it to act independently.

MAZ | CTCF | cohesin arrest | insulation | RNA Pol II pausing

T

he Myc-associated zinc finger protein (MAZ) was identified and characterized as a regulatory protein associated with MYC gene expression (1), and independently as a regulatory factor called Pur-1 that can activate an insulin promoter in HeLa cells (2). It was also identified as a serum amyloid A-activating factor and named SAF-1 (3). Previous studies have shown that MAZ plays important roles in regulation of a wide variety of genes (4–6). MAZ is a six zinc finger protein with a G-rich binding motif. It shares a part of its motif with other factors, notably the three zinc finger protein Sp1, but binding affinity of MAZ is greater to MAZ sites that have been tested (6).

Although MAZ acts as a transcription factor, our interest in MAZ initially related to the observation that, in K562 human erythroleukemia cells, there is an unusually high correlation of sites occupied by CTCF, the chromatin architectural factor, with adjacent bound MAZ (7). We confirm this correlation in K562 and other cell types. This raises the question of whether the properties of CTCF/MAZ double binding sites differ from those of CTCF-only sites. Most CTCF sites are associated with cohesin (8–12), CTCF known to help form chromatin loop domains generated by the action of cohesin, which brings together pairs of distant sites on chromatin. Defined loop boundaries are created when a traveling cohesin encounters a bound and correctly oriented CTCF molecule (13, 14). We asked whether CTCF-associated MAZ could contribute to this process.

We show here that, in K562 cells, the cohesin component Rad21 is significantly more likely to be associated with a CTCF/MAZ double site than with a site where CTCF alone is bound. Consistent with the abundance of CTCF/MAZ/Rad21 sites, we find that MAZ and Rad21 interact with each other in nuclear extracts as well as in an in vitro pull-down assay, suggesting that the free energy of this interaction contributes to the effect of MAZ binding on CTCF affinity for DNA. There are a relatively small number of genomic sites where MAZ is bound independently of CTCF but is associated with Rad21. Depletion of MAZ results in loss of Rad21 from these sites.

These results suggest that MAZ plays a role in genome organization that supports and complements that of CTCF. In many genomic locations, it binds to sites adjacent to CTCF, stabilizes CTCF binding, and provides an additional contribution to the arrest of cohesin. Consistent with an ability to interfere with such A

Significance

The protein CTCF plays a major role in large-scale organization of the genome. Binding sites for the protein MAZ are found adjacent to many CTCF sites. We show that, at such double sites, MAZ stabilizes CTCF binding. MAZ, like CTCF, acts independently as an “insulator” element to block the effects of a distal enhancer on a promoter, and, like CTCF, it can block the advance of a transcribing RNA polymerase II, leading to alternative RNA splicing patterns. Depletion of MAZ causes loss of short-range interactions within the nucleus and disruption of some longer-range interactions. Thus, MAZ plays a complementary role to CTCF in the nucleus, enhancing the organizational properties of CTCF and displaying many functions related to genome organization.

Author contributions: T.X., X.L., and G.F. designed research; T.X. and X.L. performed research; T.X., X.L., and G.F. analyzed data; and T.X. and G.F. wrote the paper.

Reviewers: P.F., Florida State University; and T.M., Columbia University Medical Center. The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To whom correspondence may be addressed. Email: gary.felsenfeld@nih.gov.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2023127118/-/DCSupplemental.

Published February 8, 2021.
(HeLa) to 58% (K562) of CTCF sites are located next to sites occupied by MAZ. Since K562 cells show the greatest overlap between CTCF and MAZ, we focused our subsequent analysis on these cells. To characterize this overlap further, we analyzed the separation distances between CTCF motifs within known CTCF binding sites and MAZ motifs within MAZ binding sites. MAZ shows some preference for binding upstream of CTCF motifs (Fig. 1B). The peak separation distance is 40 to 50 bp. CTCF plays an important role in arresting the movement of cohesin and helping to establish chromatin loop domains (17–21). The presence of MAZ at a large fraction of CTCF sites raises the question whether MAZ contributes in any way to cohesin positioning. In K562 cells, the majority of MAZ sites that colocalize with the cohesin component Rad21 are at CTCF sites. About 70% of CTCF binding sites in K562 cells are also occupied by cohesin (Fig. 1C), so it is not surprising that cohesin is present at CTCF/MAZ double sites. However, we find that CTCF sites in K562 cells that are adjacent to MAZ are more likely to sequester Rad21 (88% of sites) than are CTCF sites that are not near MAZ (56%; Fig. 1D). Consistent with this observation, a double knockdown of MAZ and CTCF results in significantly greater loss of Rad21 than knockdown of CTCF alone (Fig. 1E).

These results suggest that MAZ contributes to localization of cohesin in the genome. This could be because MAZ acts synergistically with adjacent CTCF, and arrests cohesin only in cooperation with it. Alternatively, MAZ might be able to arrest cohesin sliding independently of CTCF. To investigate these possibilities, we calculated the binding changes compared to controls in the ChiP-seq data generated from the stMAZ and siCTCF knockdown K562 cells. Our analysis confirmed a significant decrease in CTCF and Rad21 binding at the CTCF and MAZ double sites upon MAZ depletion (Fig. 2 A and B). In addition, both CTCF and Rad21 binding were further reduced at the CTCF and MAZ double sites in the CTCF and MAZ double-knockdown cells (SI Appendix, Fig. S1 A–D). In contrast, CTCF and Rad21 binding were not significantly altered at CTCF-only sites upon MAZ depletion (Fig. 2 A–C). We note that the CTCF and Rad21 binding sites that exhibited reduced binding upon MAZ depletion showed 75% overlap in position and had similar distribution patterns in the genome (Fig. 2 D–F).

Not surprisingly, genome-wide annotation reveals that MAZ sites have a genomic distribution similar to those of CTCF (Fig. 1F). Among these sites, a number of MAZ/Rad21 sites (~3,000) in K562 cells that were not associated with CTCF lost Rad21 upon MAZ knockdown, indicating that MAZ does not require CTCF to arrest cohesin movement (Fig. 3 A and B). Thus, MAZ and cohesin can be associated in the genome in the absence of CTCF, and, at such sites, loss of MAZ results in loss of cohesin (Fig. 3B).

Interaction of MAZ and Rad21. The frequent genome-wide association of MAZ and Rad21 raised the question whether the two proteins interact physically. To test this possibility, we purified bacterially expressed MBP-tagged MAZ and used it in immunoprecipitation experiments to incubate with nuclear extracts from K562 cells (SI Appendix, Methods). The MBP pull-down assay showed that Rad21 was present in the anti-MBP immunoprecipitated fraction (Fig. 3 C and D). In an alternative experiment, we expressed Flag-tagged MAZ in HEK293T cells. Anti-Flag immunoprecipitation showed that Rad21 coprecipitates with Flag-tagged MAZ (Fig. 3E). These results were confirmed by mass spectrometry analysis of similar anti-Flag immunoprecipitates. As shown in SI Appendix, Table S1, among the proteins that coprecipitate with MAZ are many of the components of the cohesin complex, including Rad21, SMC1A, and SMC3.

Although this establishes a MAZ-cohesin interaction, it does not distinguish between direct interaction and interactions mediated by a cofactor in the extract. We therefore repeated pull-down experiments using purified MBP-tagged MAZ expressed in...
MAZ is required for stabilizing the binding of both CTCF and cohesin in chromatin. (A) Box plot of CTCF binding changes after MAZ KD (iMAZ) compared to control at CTCF and MAZ double binding sites (1), at CTCF-only binding sites (2), or at MAZ-only binding sites (3). P values for all comparisons were calculated by the two-sided Kruskal–Wallis H-test (\(***P < 0.001\)). (B) Box plot of Rad21 binding changes after MAZ KD compared to control at CTCF and MAZ double-binding sites (1), at CTCF-only binding sites (2), or at MAZ-only binding sites (3). P values for all comparisons were calculated by the two-sided Kruskal–Wallis H-test (\(***P < 0.001\)). (C) Representative plots showing that both CTCF and Rad21 binding were reduced at CTCF and MAZ double-binding sites after MAZ knockdown. Identity of each ChIP-seq is listed on the left, and two replicates are shown in blue for CTCF IP and in red for Rad21 IP. (D) Overlap between the reduced numbers of CTCF binding sites and Rad21 binding sites that were analyzed in A and B. After MAZ knockdown, 18,945 CTCF binding sites showed reduced binding, and 18,570 Rad21 binding sites showed reduced binding compared to that in the control. About 74% of these sites overlapped. (E) Distribution of the CTCF reads at sites that showed reduced binding after MAZ knockdown in A. (F) Genome-wide distribution of the Rad21 binding sites that showed reduced binding after MAZ knockdown in B.

CTCF sites at these regions are adjacent to bound MAZ. In contrast, very few CTCF sites not associated with MAZ are marked by H3K4 trimethylation (Fig. 4B). To explore further the regulatory role of MAZ, we analyzed RNA-seq data from K562 cells (SI Appendix, Table S2). These data show that 290 genes are up-regulated and 277 are down-regulated twofold or more when MAZ is depleted. Of these, 162 of the up-regulated and 204 of the down-regulated genes are associated with MAZ/H3K4Me3 sites (Fig. 4C). Surprisingly, 75 to 80% of these are adjacent to CTCF sites, whereas only 27% of all MAZ/H3K4Me3 sites are shared with CTCF in the whole genome. Thus, genes with expression most sensitive to MAZ depletion appear to be enriched in CTCF binding sites associated with promoters (Fig. 4C). A summary of the effects of MAZ knockdown on expression of the entire gene population is given in a volcano plot in Fig. 4D, and a list of the affected genes is shown in SI Appendix, Table S2. The results of a Gene Ontology (GO) analysis for the most affected genes are presented in Fig. 4E. We note that MAZ appears to be associated with regulation of the mitotic cell cycle, which may be connected to its interaction with the cohesin complex (23).

MAZ as an Insulator Protein. CTCF is well known to function as an “enhancer blocking” factor when placed between an enhancer and promoter (24, 25). To determine whether MAZ has similar properties, we made use of a construct similar to those used in bacteria and Rad21 expressed in the TnT Coupled Wheat Germ Extract Systems (Methods). As shown in Fig. 3F, the two proteins coprecipitate, confirming that Rad21 and MAZ interact directly. However, cohesin components SMC1A and SMC3 do not interact directly with MAZ (SI Appendix, Fig. S2B).

Given the proximity of MAZ binding sites to many CTCF binding sites, we asked whether there was any physical interaction between MAZ and CTCF. Coimmunoprecipitation and mass spectrometry studies described above to identify potential MAZ cofactors might interact in vivo when bound to adjacent DNA sites. To investigate this, we knocked down CTCF in K562 cells (Fig. 3, A and B). In K562 cells, 78% of MAZ binding sites are marked by H3K4 trimethylation (H3K4Me3), a modification typically associated with active promoters (22) (Fig. 4A), suggesting that MAZ may play a regulatory role in expression of a wide variety of genes (Discussion). We asked whether CTCF was associated with MAZ at these H3K4Me3 regions. We find that 41% (12,501 of 30,371) of the MAZ binding sites at Regulatory Sites: Partnership with CTCF. In K562 cells, 78% of MAZ binding sites are marked by H3K4 trimethylation (H3K4Me3), a modification typically associated with active promoters (22) (Fig. 4A), suggesting that MAZ may play a regulatory role in expression of a wide variety of genes (Discussion). We asked whether CTCF was associated with MAZ at these H3K4Me3 regions. We find that 41% (12,501 of 30,371) of the MAZ binding sites are next to a CTCF binding site, whereas only 27% of all MAZ-H3K4Me3 sites are shared with CTCF (Fig. 4A). To further investigate this, we knocked down CTCF in K562 cells (Fig. 3, A and B). In K562 cells, 78% of MAZ binding sites are marked by H3K4 trimethylation (H3K4Me3), a modification typically associated with active promoters (22) (Fig. 4A), suggesting that MAZ may play a regulatory role in expression of a wide variety of genes (Discussion). We asked whether CTCF was associated with MAZ at these H3K4Me3 regions. We find that 41% (12,501 of 30,371) of the MAZ binding sites at Regulatory Sites: Partnership with CTCF. In K562 cells, 78% of MAZ binding sites are marked by H3K4 trimethylation (H3K4Me3), a modification typically associated with active promoters (22) (Fig. 4A), suggesting that MAZ may play a regulatory role in expression of a wide variety of genes (Discussion). We asked whether CTCF was associated with MAZ at these H3K4Me3 regions. We find that 41% (12,501 of 30,371) of the MAZ binding sites at Regulatory Sites: Partnership with CTCF. In K562 cells, 78% of MAZ binding sites are marked by H3K4 trimethylation (H3K4Me3), a modification typically associated with active promoters (22) (Fig. 4A), suggesting that MAZ may play a regulatory role in expression of a wide variety of genes (Discussion). We asked whether CTCF was associated with MAZ at these H3K4Me3 regions. We find that 41% (12,501 of 30,371) of the MAZ binding sites at Regulatory Sites: Partnership with CTCF. In K562 cells, 78% of MAZ binding sites are marked by H3K4 trimethylation (H3K4Me3), a modification typically associated with active promoters (22) (Fig. 4A), suggesting that MAZ may play a regulatory role in expression of a wide variety of genes (Discussion). We asked whether CTCF was associated with MAZ at these H3K4Me3 regions. We find that 41% (12,501 of 30,371) of the
many studies of CTCF, in which a sequence that binds the insulator protein to be tested is placed between an enhancer and a promoter that drives expression of luciferase (25). As shown in Fig. 5 A and B, the construct with a MAZ binding motif, when transfected into HEK293T or HeLaS3 cells, expressed luciferase at lower levels than those observed with a construct in which the MAZ binding motif was mutated, suggesting that MAZ acts as an insulator to block the interaction between enhancer and promoter.

To confirm this property of MAZ, we repeated the experiments with the constructs containing the MAZ binding sites, but used RNA interference (RNAi) to knock down MAZ expression (Fig. 5 C–E). This resulted in loss of the enhancer-blocking effect of the MAZ element: luciferase expression from the wild-type MAZ construct (pIHLMAZwtE) was almost the same as that from the mutant construct (pIHLMAZmuE) lacking MAZ binding (Fig. 5F). In complementary experiments, we introduced a plasmid containing MAZ binding motifs can form G-quadruplex structures (26) (Discussion). To address this problem, we analyzed ChIP-seq data in K562 cells for the elongating form of RNA Pol II (Pol II S2P) and identified the Pol II S2P sites that were adjacent to bound MAZ. (G) CTCF is more stable when MAZ is nearby (+). Shown are results of two independent experiments in which CTCF protein was partially depleted with siRNA. ChIP-seq data were used to measure CTCF binding in K562 cells. After CTCF depletion, more CTCF binding sites are retained when MAZ is nearby than in the absence of MAZ (–).

Fig. 3. Interactions of MAZ with cohesin. (A) Rad21 lost from binding sites after MAZ knockdown, showing that most such sites are not associated with CTCF (CTCF,–2973). Sites shown are common to three separate ChIP-seq experiments for CTCF and for Rad21 before and after MAZ knockdown. MAZ binding sites are restricted to those sites detected from the ENCODE MAZ ChIP-seq data. (B) MAZ-dependent binding of Rad21 at a site that has no CTCF. Examples of ChIP-seq data showing that Rad21 binding is reduced after MAZ depletion at MAZ sites lacking CTCF binding. (C) Purification of MBP-tagged MAZ protein. MBP-tagged MAZ was expressed in bacteria and purified with MBP affinity beads. Western blot shows that Rad21 coprecipitates with MBP-MAZ. (D) Pull-down assay. Purified MBP-MAZ was added to nuclear extracts from K562 cells and immunoprecipitated with anti-MBP beads. Western blot shows that Rad21 coprecipitates with MBP-MAZ. (E) Co-IP. Flag-MAZ was expressed in HEK293T cells, then purified with anti-Flag beads. Western blot shows that Rad21 coprecipitates with Flag-MAZ. (F) Pull-down assay. MBP-tagged MAZ, expressed in bacteria and purified with MBP affinity beads, was used to immunoprecipitate Rad21 expressed in vitro in the TnT Coupled Wheat Germ Extract Systems (SI Appendix, Table S5). Western blot shows that Rad21 interacts with MBP-MAZ. (G) CTCF is more stable when MAZ is nearby (+). Shown are results of two independent experiments in which CTCF protein was partially depleted with siRNA. ChIP-seq data were used to measure CTCF binding in K562 cells. After CTCF depletion, more CTCF binding sites are retained when MAZ is nearby than in the absence of MAZ (–).
absence of a MAZ site. Consistent with the observation that MAZ can pause the elongating form of Pol II, we found that Pol II S2P was enriched at downstream MAZ binding sites relative to upstream exons (Fig. 6E). An example of MAZ effects on splicing is shown in Fig. 6F. Thus, MAZ strongly promotes Pol II pausing-associated splicing in the absence of CTCF.

**Location of MAZ Between Closely Spaced Genes.** The early studies of MAZ binding-site effects on RNA Pol II noted that a genomic MAZ site located between a pair of closely spaced genes can contribute to transcription termination and prevent interference with the downstream gene (29, 30). We examined the genomic population of closely spaced genes and divided it into pairs oriented in the same direction, those facing (transcribing) toward each other and those facing away from each other, in all cases with a separation from 2 kb to 5 kb. We then asked in each case what fraction had an occupied MAZ binding site in K562 cells. Over 40% of closely spaced genes that are transcribed in the same direction have a MAZ site between them (SI Appendix, Fig. S7C). This probably reflects the high incidence of MAZ sites near promoters (Fig. 4A), which also explains the observation that over 60% of divergent genes (transcribed away from each other) are separated by MAZ, since the presence of two promoters in the intergenic space increases the probability of finding at least one MAZ site. Similarly, the absence of a promoter explains why only 20% of convergent genes (transcribed toward each other) have a MAZ site between them. It therefore does not appear that MAZ binding sites are present in unexpectedly high abundance between closely spaced genes. However, our RNA Pol II pausing data support the idea that a MAZ site between closely spaced genes transcribed in the same direction will impede the incursion into the downstream gene of a polymerase transcribing the upstream gene, as previously suggested (29, 30).

**Effects of MAZ Depletion on Large-Scale Genome Organization.**

Given the association of MAZ with cohesin, we asked what effect depletion of MAZ might have on large-scale genome organization. We carried out Hi-C analysis comparing genomic interactions in control K562 cells with interactions in cells in which MAZ had been knocked down. After evaluating the quality of the Hi-C libraries, we pooled the data and detected about 150 million normal paired reads for K562 control and K562 siMAZ knockdown cells (SI Appendix, Table S3). The whole-genome chromosomal interaction heat map is similar between the control and MAZ-depleted cells. The A and B compartments also are not significantly changed after depletion of MAZ (SI Appendix, Fig. S4).

Differential TAD analysis (Fig. 7A and B) shows that cells depleted of MAZ have reduced interactions within a TAD compared to those in control cells, whereas there is little or no effect on inter-TAD contacts. Consistent with this conclusion, an aggregate TAD analysis (ATA) shows a marked decrease in intra-TAD contact frequency, corresponding to the strong blue diagonal. The blue points at the 5′ and 3′ borders of the ATA analysis indicate that loss of MAZ also decreases contact frequency in the 5′ and 3′ boundaries of individual TADs (Fig. 7B). In agreement with this, the genome-wide insulation scores at 10-kb resolution aligned to TAD borders are lower in cells in which MAZ is depleted (Fig. 7C). Differential TAD analysis (Fig. 7A and B) shows that cells depleted of MAZ have reduced interactions within a TAD compared to those in control cells, whereas there is little or no effect on inter-TAD contacts. Consistent with this conclusion, an aggregate TAD analysis (ATA) shows a marked decrease in intra-TAD contact frequency, corresponding to the strong blue diagonal. The blue points at the 5′ and 3′ borders of the ATA analysis indicate that loss of MAZ also decreases contact frequency in the 5′ and 3′ boundaries of individual TADs (Fig. 7B). In agreement with this, the genome-wide insulation scores at 10-kb resolution aligned to TAD borders are lower in cells in which MAZ is depleted (Fig. 7C). Differential TAD analysis (Fig. 7A and B) shows that cells depleted of MAZ have reduced interactions within a TAD compared to those in control cells, whereas there is little or no effect on inter-TAD contacts. Consistent with this conclusion, an aggregate TAD analysis (ATA) shows a marked decrease in intra-TAD contact frequency, corresponding to the strong blue diagonal. The blue points at the 5′ and 3′ borders of the ATA analysis indicate that loss of MAZ also decreases contact frequency in the 5′ and 3′ boundaries of individual TADs (Fig. 7B). In agreement with this, the genome-wide insulation scores at 10-kb resolution aligned to TAD borders are lower in cells in which MAZ is depleted (Fig. 7C). Differential TAD analysis (Fig. 7A and B) shows that cells depleted of MAZ have reduced interactions within a TAD compared to those in control cells, whereas there is little or no effect on inter-TAD contacts. Consistent with this conclusion, an aggregate TAD analysis (ATA) shows a marked decrease in intra-TAD contact frequency, corresponding to the strong blue diagonal. The blue points at the 5′ and 3′ borders of the ATA analysis indicate that loss of MAZ also decreases contact frequency in the 5′ and 3′ boundaries of individual TADs (Fig. 7B). In agreement with this, the genome-wide insulation scores at 10-kb resolution aligned to TAD borders are lower in cells in which MAZ is depleted (Fig. 7C). Differential TAD analysis (Fig. 7A and B) shows that cells depleted of MAZ have reduced interactions within a TAD compared to those in control cells, whereas there is little or no effect on inter-TAD contacts. Consistent with this conclusion, an aggregate TAD analysis (ATA) shows a marked decrease in intra-TAD contact frequency, corresponding to the strong blue diagonal. The blue points at the 5′ and 3′ borders of the ATA analysis indicate that loss of MAZ also decreases contact frequency in the 5′ and 3′ boundaries of individual TADs (Fig. 7B). In agreement with this, the genome-wide insulation scores at 10-kb resolution aligned to TAD borders are lower in cells in which MAZ is depleted (Fig. 7C).
These changes show that, although MAZ does not have the large-scale effects on genome organization associated with CTCF, it nonetheless appears to play an important role in the pattern of genomic contacts, principally within TADs, and it contributes to the integrity of TADs.

Discussion

We have shown that MAZ, often bound to DNA in association with CTCF, plays an important role in stabilizing CTCF binding and directly or indirectly contributes to the arrest of the cohesin complex at CTCF sites. To a lesser extent, it also contributes independently of CTCF to the organization of the genome through its interaction with cohesin. MAZ interacts directly with the Rad21 component of the cohesin complex and indirectly with other members of the complex. Although the majority of MAZ/cohesin (Rad21) sites in the genome are adjacent to CTCF sites, we identify in K562 cells about 3,000 MAZ/Rad21 sites that are not associated with CTCF alone (Fig. 1E). That is significantly greater than that resulting from depletion of MAZ and CTCF results in loss of cohesin from some of those sites (Fig. 3G). It seems possible that cohesin, bound to both MAZ and CTCF at these double sites, may be responsible for the observed contribution of MAZ to CTCF binding affinity. The bound MAZ–CTCF–cohesin cluster would then constitute a mutually stabilizing complex (Fig. 8D).

MAZ shares other properties with CTCF. In an experiment similar to one used frequently (25, 31) to demonstrate the enhancer-blocking properties of CTCF, we incorporated MAZ sites between an enhancer and the promoter of a reporter gene (Fig. 5). When the wild type construct is transfected into HEK293T or HeLa S3 cells, expression of the reporter gene is inhibited relative to that of Renilla luciferase activity (Fig. 5A). Similarly, it has been shown that a series of four MAZ sites inserted in a minigene construct can interfere with transcription elongation (29). MAZ also makes an indirect contribution to cohesin binding when it occupies a site adjacent to bound CTCF: almost 90% of CTCF/MAZ sites in K562 cells are associated with cohesin, compared to 56% of CTCF-only sites (Fig. 1D). Depletion of both MAZ and CTCF results in loss of cohesin from some of those sites that is significantly greater than that resulting from depletion of CTCF alone (Fig. 1E). The presence of MAZ also helps to stabilize CTCF binding at such double sites. Knocking down CTCF results in preferential loss of CTCF from sites that are not adjacent to MAZ (Fig. 3G). It seems possible that cohesin, bound to both MAZ and CTCF at these double sites, may be responsible for the observed contribution of MAZ to CTCF binding affinity. The bound MAZ–CTCF–cohesin cluster would then constitute a mutually stabilizing complex (Fig. 8D).
and affect splicing choices (5). As noted above, MAZ sites not associated with CTCF are frequently located at promoters, and a genome-wide survey shows that, at many promoters, there are multiple MAZ binding motifs. As has been pointed out in many papers (26, 35), when such G-rich sequences are located close to one another and present in a transcribed region, the G-rich strand as well as a G-rich RNA transcript may be able to adopt the G quadruplex structure. It seems probable that G quadruplex formation is at least partially responsible for the behavior reported by Yokahara and Proudfoot (30). More recently, sequences with the potential to form G quadruplex have been identified upstream of a variety of genes, and G quadruplex-specific antibodies have revealed the presence of such sequences at TAD domain boundaries (36, 37). It should be remembered, however, that antibodies specific for a non-B form DNA will stabilize that form (38, 39). Thus, antibodies reveal only the presence of potential G quadruplex-forming sequences.

This raises the question of the relationship of MAZ binding to quadruplex structure and the role of each in regulation of gene expression. In the case of single binding motifs, or where multiple motifs are not close together, G quadruplex will not form. Most importantly, although there is evidence that MAZ can bind to the quadruplex, published data clearly show that it binds more strongly to the corresponding duplex (40), so that bound MAZ will preferentially stabilize its binding site in the duplex conformation.

In resolution of this question, two kinds of evidence support the conclusion that important properties of MAZ sites depend on bound MAZ, but not the MAZ motif alone. First, the ability of a MAZ binding site to act as an enhancer-blocking insulator depends on binding of the MAZ protein (Fig. 5). Overexpression of MAZ increases blocking activity, and depletion of MAZ reduces it. Second, we find that, at many sites in the genome, bound MAZ, but not the MAZ motif alone, is responsible for pausing a transcribing RNA Pol (Fig. 6 A–C). We identified many sites occupied by both MAZ and the elongating form of RNA Pol II, but not occupied by CTCF (SI Appendix, Fig. S7A). Such a configuration is consistent with a role for MAZ in RNA Pol II pausing; the further observation that knocking down MAZ results in loss of both RNA Pol II and MAZ confirms that MAZ protein is both necessary and sufficient for pausing of Pol II at a MAZ binding site.

It has been shown that pausing of RNA Pol II can result in alternative splicing choices during transcription, and this has been demonstrated explicitly in the case of CTCF (28). It seems likely that MAZ would have similar effects on splicing outcomes. An example of the effect of MAZ knockdown on splicing choices is shown in Fig. 6F. We note that among the proteins that coprecipitate with MAZ (SI Appendix, Table S1) are the U1, U2, U4, U5, and U6 members of the spliceosome complex (41, 42), a property shared with CTCF.

Xiao et al.
The Myc-associated zinc finger protein (MAZ) works together with CTCF to control cohesin positioning and genome organization
It is interesting to compare the properties of MAZ with those of YY1, another zinc finger protein that contributes to genome organization (43–45). YY1 interacts with the SMC1 subunit of cohesin (44) and helps stabilize promoter–enhancer interactions in the genome (43, 45). In K562 cells, most YY1 binding sites (77%) are close to MAZ sites (SI Appendix, Fig. S6A). However, most (73%) of these YY1/MAZ sites are not close to CTCF (SI Appendix, Fig. S6B). Thus, one of the properties of MAZ, in stabilizing CTCF binding, is distinct from the properties of YY1. The stabilization of enhancer–promoter interactions by YY1 is proposed to involve contact between YY1 proteins bound at those sites and further strengthened by interaction with cohesin (45). Unlike YY1, many MAZ sites are located near promoters, but few are near enhancers (SI Appendix, Fig. S6 C and D). Although, like YY1, MAZ helps stabilize short-range interactions, it also affects TAD organization (see below), probably because of its ability to arrest cohesin sliding.

Fig. 7. MAZ and genome organization in K562 cells. (A) Hi-C data analysis was conducted using the GENOVA R package (18). Dot plot of differential TAD analysis results. K562 cells depleted of MAZ have fewer interactions within a TAD (dark blue; TAD distance N+0) compared to those in K562 control cells. Log2 ratio of the contact frequency in siMAZ compared to control is plotted. (B) Aggregate TAD analysis results. The differential signal between K562 control and K562 siMAZ was calculated at 10-kb resolution. Blue indicates reduced signal, and red indicates increased signal. There is a decrease in contacts within the TADs and at TAD borders in the K562 siMAZ. (C) Alignment of border strength of TADs. The border strength of TADs is reduced in K562 siMAZ compared to that in K562 Ctr. Genome-wide insulation scores were calculated in 10-kb resolution using the findTADsAndLoops.pl script in Homer. Heat map was generated with deepTools. Circles mark borders of TADs. (D) Aggregate peak analysis at 2.5-kb resolution. (E) Example of Hi-C contact matrices for a zoomed-in region on chromosome 11 (chr11:1,500,000–2,500,000). This region covers the imprinting control region of the h19 locus. Heat map was generated at 5-kb resolution using the Juicer program. The upper right half contact matrices are generated from K562 Ctr, while the lower left half contact matrices are generated from K562 siMAZ. The locations showing changes of contact frequency are marked in red circles. Loops shown below the heat map were generated with the hichipper program and visualized with WashU Epigenome Browser. Virtual 4C maps were generated using the coordinate chr11:2,015,001–2,020,000, which corresponds to the location of h19 gene, as viewpoint. ChIP-seq and gene tracks were generated with the pyGenomeTracks program (Methods).

Fig. 8. Model of MAZ action. (A) Proposed model for the effect of bound MAZ on adjacent CTCF, in which the shared interaction with cohesin (red ring) provides increased stability of CTCF and cohesin. Loss of bound MAZ from these sites reduces CTCF binding affinity. (B) Pol II S2P sliding arrested by MAZ. (C) Loss of bound CTCF may allow cohesin to slide to a nearby MAZ site.
MAZ Contributes to Genome Organization. CTCF plays the predominant role in organizing the genome, probably because it is so effective at arresting sliding, both of cohesin and RNA Pol II. MAZ, though it shares these properties, is probably less efficient. Nonetheless, the Hi-C data in Fig. 7 show that MAZ also plays an important role in genome organization on a different scale from CTCF. Disruption of MAZ primarily disrupts interactions within TADs, as well as reducing the contacts that establish the boundaries of individual TADs. It seems likely that the effect on intra-TAD contacts reflects the ability of MAZ to stop sliding cohesin complexes. As we have shown above, MAZ is capable not only of arresting cohesin on its own, but also contributes significantly to the ability of CTCF to do so; some effects of loss of MAZ on long-range organization may reflect loss of bound CTCF stabilized by adjacent MAZ binding. Perhaps surprisingly, the rearrangements in genomic organization seen on MAZ deletion with CTCF stabilized by adjacent MAZ binding. Perhaps surprisingly, the rearrangements in genomic organization seen on MAZ deletion within TADs, as well as reducing the contacts that establish the architecture is resilient.

There is reason to think that some of the properties of MAZ are shared with other zinc finger proteins. For example, our lab has shown that Vez1 is capable of pausing RNA Pol II and affecting RNA splicing outcomes (46). However, these proteins do not appear to have formed the collaborative arrangement with CTCF that is reflected in the high abundance in the genome of paired CTCF/MAZ sites. That pairing results in stabilization of the CTCF/DNA interaction and must also contribute to the arrest of cohesin. If CTCF is the “master organizer of the genome,” MAZ may be the master’s apprentice.

1. K. B. Marcu, A. J. Patel, Y. Yang, Differential regulation of the c-MYC P1 and P2 promoters in the absence of functional tumor suppressors: Implications for mechanisms of deregulated MYC transcription. Curr. Top. Microbiol. Immunol. 224, 47–56 (1997).
2. G. C. Kennedy, W. J. Rutter, Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates an insulin promoter in heterologous cells. Proc. Natl. Acad. Sci. U.S.A. 89, 11498–11502 (1992).
3. A. Ray, B. K. Ray, Isolation and functional characterization of cDNA of serum amyloid A-activating factor that binds to the serum amyloid A promoter. Mol. Cell. Biol. 18, 7327–7335 (1998).
4. M. Haller, J. Au, M. O’Neill, D. J. Lamb, 16p11.2 transcription factor MAZ is a dosage sensitive regulator of genitourinary development. Proc. Natl. Acad. Sci. U.S.A. 115, E1849–E1858 (2018).
5. N. D. Robson-Dixon, M. A. Garcia-Blanco, MAZ elements alter transcription elongation and silencing of the fibrinogen growth factor receptor 2 exon lib. J. Biol. Chem. 279, 29075–29084 (2004).
6. S. Her, R. CLaycomb, T. C. Tai, D. L. Wong, Regulation of the rat phenylethanolamine N-methyltransferase gene by transcription factors Sp1 and MAZ. Mol. Pharmacol. 64, 1180–1188 (2003).
7. K. Zhang, N. Li, R. R. Ainsworth, W. Wang, Systematic identification of protein interactions mediating chromatin looping. Nat. Commun. 7, 12249 (2016).
8. Y. H. Jung et al., Maintenance of CTCF- and transcription factor-mediated interactions from the gametes to the early mouse embryo. Mol. Cell 75, 154–171.e5 (2019).
9. K. L. Lyu, M. J. Rowley, V. G. Corces, Architectural proteins and pluripotency factors cooperate to orchestrate the transcriptional response of HESCs to temperature stress. Mol. Cell 71, 940–955.e7 (2018).
10. J. E. Phillips-Cremins et al., Architectural protein subclassies shape 3D organization of genomes during lineage commitment. Cell 153, 1281–1295 (2013).
11. M. J. Rowley, V. G. Corces, Organizational principles of 3D genome architecture. Nat. Rev. Genet. 19, 789–800 (2018).
12. K. Van Bortle et al., Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol. 15, R82 (2014).
13. S. S. Rao et al., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680 (2014).
14. S. S. Rao et al., Cohesin loss eliminates all loop domains. Cell 171, 305–320.e24 (2017).
15. Y. Li et al., Characterization of constitutive CTCF/cohesin loci: A possible role in establishing topological domains in mammalian genomes. BMC Genomics 14, 553 (2013).
16. R. Mourad, L. Li, D. Giguver, Uncovering direct and indirect molecular determinants of chromatin loops using a computational integrative approach. PLoS Comput. Biol. 13, e1005538 (2017).
17. G. A. Busslinger et al., Cohesin is positioned in mammalian genomes by transcription, CTCF and WapI. Nature Genet. 44, 503–507 (2017).
18. J. S. Haahr et al., The cohesin release factor WAPL restricts chromatin loop extension. Cell 165, 693–707.e14 (2017).
19. S. Hadjur et al., Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature 460, 410–413 (2009).
20. T. Mishiro et al., Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. EMBO J. 28, 1234–1245 (2009).
21. V. Parelo et al., Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132, 422–433 (2008).
22. L. M. Soares et al., Determinants of histone H3K4 methylation patterns. Mol. Cell 68, 773–785.e5 (2017).
23. J. M. Peters, A. Tedeschi, J. Schmitz, The cohesin complex and its roles in chromosome biology. Genes Dev. 22, 3089–3114 (2008).
24. A. C. Bell, A. G. West, G. Felsenfeld, The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98, 387–396 (1999).
25. K. Ishihara, M. Oshimura, M. Nakao, CTCF-dependent chromatin insulator is linked to epigenetic remodeling. Mol. Cell 23, 733–742 (2006).
26. N. Kim, The interplay between G-quadruplexes and transcription. Curr. Med. Chem. 26, 2898–2917 (2019).
27. A. R. Kornblihtt, CTCF: From insulators to alternative splicing regulation. Cell Res. 22, 450–452 (2012).
28. S. Shukla et al., CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479, 74–79 (2011).
29. R. Ashfield et al., MAZ-dependent termination between closely spaced human complement genes. EMBO J. 13, 5655–5667 (1994).
30. M. Yokahia, N. J. Proudfoot, Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. Mol. Cell 3, 593–600 (1999).
31. K. S. Wendt, J. M. Peters, How cohesin and CTCF cooperate in regulating gene expression. Chromosome Res. 17, 201–214 (2009).
32. M. Yokahia, N. J. Proudfoot, Transcriptional termination and coupled polyadenylation and splicing in vivo. EMBO J. 19, 3770–3777 (2000).
33. M. Guillerou, N. J. Proudfoot, Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. Cell 132, 983–995 (2008).
34. J. Song et al., Two consecutive zinc fingers in Sp1 and in MAZ are essential for interactions with cis-elements. J. Biol. Chem. 276, 30439–30441 (2001).
35. J. A. Capra, K. Paeschke, M. Singh, V. A. Zakian, G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. PLoS Comput. Biol. 6, e1000861 (2010).
36. Y. Hou et al., Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. Epigenetics 14, 894–911 (2019).
37. V. B. Kaiser, C. A. Semple, Chromatin loop anchors are associated with genome instability in cancer and recombination hotspots in the germ line. Genome Biol. 19, 101 (2018).
38. M. Bartas et al., The presence and localization of G-quadruplex forming sequences in the domain of bacteria. Molecules 24, 1711 (2019).
39. E. M. Lafer, R. Sousa, R. Ali, A. Rich, B. D. Stollar, The effect of anti-Z-DNA antibodies on the B-DNA-Z-DNA equilibrium. J. Biol. Chem. 261, 6438–6443 (1986).
40. V. Brázda, L. Hároniková, J. C. Liao, M. Fojta, DNA and RNA quadruplex-binding proteins. Int. J. Mol. Sci. 15, 17493–17517 (2014).
41. J. Abelson, A close-up look at the spliceosome, at last. Proc. Natl. Acad. Sci. U.S.A. 114, 4288–4293 (2017).
42. J. M. Dowen et al., Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell 159, 374–387 (2014).
43. J. A. Beagan et al., YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res. 27, 1139–1152 (2017).
44. X. Pan et al., YY1 controls Igκ repertoire and B-cell development, and localizes with condensin on the Igκ locus. EMBO J. 32, 1168–1182 (2013).
45. A. S. Weintraub et al., YY1 is a structural regulator of enhancer-promoter loops. Cell 171, 1573–1588.e28 (2017).
46. H. Gowher, K. Brick, R. D. Camerini-Otero, G. Felsenfeld, Vezf1 protein binding sites genome-wide are associated with pausing of elongating RNA polymerase II. Proc. Natl. Acad. Sci. U.S.A. 109, 2370–2375 (2012).
47. H. Belaghzal, J. Dekker, J. H. Gibcus, Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. Methods 123, 56–65 (2017).
48. B. Gel et al., regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics 32, 289–291 (2016).