Nucleic Acids Research, 2024, **52**, 9328–9339
https://doi.org/10.1093/nar/gkae591
Advance access publication date: 16 July 2024
Synthetic Biology and Bioengineering

Re-engineered guide RNA enables DNA loops and contacts modulating repression in *E. coli*

Yunshi Yang [1,†], Iris Rocamonde-Lago [1,†], Boxuan Shen [1,2], Ieva Berzina [1], Johanna Zipf [1] and Björn Högberg [1,*]

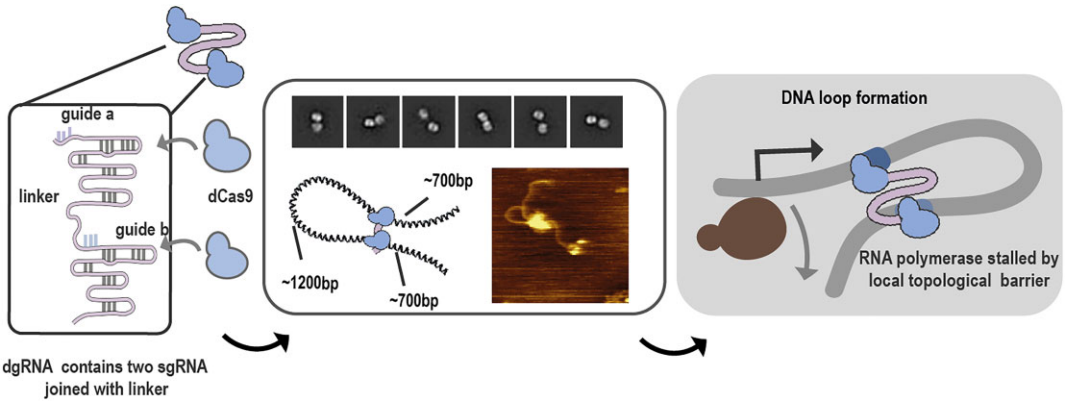
[1]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Solna, Stockholm 17177, Sweden
[2]Biohybrid Materials, Department of Bioproducts and Biosystems, Aalto University, 00076 Aalto, Finland

[*]To whom correspondence should be addressed. Tel: +46 8524807036; Email: bjorn.hogberg@ki.se
[†]The first two authors should be regarded as Joint First Authors.

Abstract

RNA serves as information media as well as molecular scaffold in nature and synthetic systems. The single guide RNA (sgRNA) widely applied in CRISPR techniques exemplifies both functions, with a guide region bearing DNA base-pairing information, and a structural motif for Cas9 protein scaffolding. The scaffold region has been modified by fusing RNA aptamers to the tetra-stem loop. The guide region is typically not regarded as a pluggable module as it encodes the essential function of DNA sequence recognition. Here, we investigate a chimera of two sgRNAs, with distinct guide sequences joined by an RNA linker (dgRNA), regarding its DNA binding function and loop induction capability. First, we studied the sequence bi-specificity of the dgRNA and discovered that the RNA linker allows distal parts of double-stranded DNA to be brought into proximity. To test the activity of the dgRNA in organisms, we used the LacZ gene as a reporter and recapitulated the loop-mediated gene inhibition by LacI in *E. coli*. We found that the dgRNA can be applied to target distal genomic regions with comparable levels of inhibition. The capability of dgRNA to induce DNA contacts solely requires dCas9 and RNA, making it a minimal system to remodel chromosomal conformation in various organisms.

Graphical abstract



Introduction

Since the initial identification of the CRISPR locus (1), great research efforts have been dedicated to the understanding of the CRISPR-Cas system as an adaptive immune system in microbes. Focusing on the molecular mechanism underlying the defensive function of CRISPR, Cas9 was discovered as an unusually large Cas protein (2) and later demonstrated its DNA endonuclease activity was demonstrated *in vitro*. Later on, the discovery of CRISPR RNA (crRNA) (3) and trans-activating CRISPR (tracrRNA) (4) showed that the flexibility of the CRISPR-Cas system was enabled by encoding target information in the nucleic acid. Soon after the elucidation of the essential components of CRISPR-Cas9 for DNA cleavage in the endogenous environment, it was harnessed for carrying out precise gene editing in various organisms (5–7). The CRISPR-Cas9 system was further simplified by fusing crRNA and tracrRNA into a single transcript coined as single guide RNA (sgRNA) (8). By encoding the 20 nucleotides guide sequence in a sgRNA, the CRIPSR-Cas9 system can be easily reprogrammed to target almost any arbitrary gene locus. The

Received: January 17, 2023. **Revised:** May 23, 2024. **Editorial Decision:** June 21, 2024. **Accepted:** June 25, 2024

highly programmable gene targeting function of the CRISPR-Cas9 system was later repurposed for DNA transcription interference by creating an endonuclease activity dead Cas9 (dCas9) (9). The CRISPR-dCas9 system has become a platform for epigenetic modification by fusing a transcriptional activator or repressor to dCas9 (10–14) in conjunction with the modification of the sgRNA (15). In recent years, more CRISPR systems were uncovered, with capability extended to RNA targeting (16) and protease function (17). The expanding CRISPR toolkit has been continuously driving and transforming the research paradigm in molecular biology across many fields.

A prominent question in cell biology is to interpret the functional outcome of genome organization. The higher-order structure of the genome provides physical premises for the molecular machinery to exert a regulatory effect by interacting with functional genomic elements, serving as a key regulatory layer for gene expression. With recent technical advances, the long-range (18) DNA contacts at a global level could be identified from an ensemble average, regarded as the fundamental mechanism for genome organization. However, the causal-effect relationship between genome structure and transcription outcome remains disputed due to conflicting study results. Research about specific domains revealed an altered expression pattern after structural reorganization (19,20), while global disruption of topologically associated domains only led to dysregulation of a minor portion of genes (21). To systematically evaluate the effect of regulatory elements and potential targets, the CRISPR-Cas9 system has been used to screen the function of regulatory elements through either genetic disruption or repression (22,23). Despite the advantage of the CRISPR system to flexibly target functional genomic elements, these available approaches cannot determine the effect derived from physical contacts sustained by genomic architecture, focusing instead on perturbing regulatory elements such as enhancers. In recent years, the CRISPR-dCas9 system has been engineered to enable loop formation through heterodimerization of a modified dCas9 protein directed to separate genomic targets. Thus, demonstrating the feasibility and functional implication of introducing DNA loops by dCas9 dimerization. These methods can be controlled by additional chemicals (24) or physical cues (25) to induce loop formation, as well as *via* incorporation of orthogonal dCas9 subtypes, to allow heterodimer formation (24,26). Another method used a single type of engineered dCas9 fusion protein while supplementing an additional bridge protein to allow DNA loop formation (25).

Here, we introduce a concise strategy focused on redesigning the guide RNA to allow the formation of DNA contacts. We first added a non-structured RNA linker to connect two sgRNA units with different guide sequences in tandem, referred to as double guide RNA (dgRNA) (Figure 1A, Supplementary Figure S9). We next validated that the dgRNA, as a single transcript *in vitro*, can recruit two dCas9 proteins to form a ribonucleoprotein (RNP), and that the formed complex was capable of binding to the corresponding DNA targets, bringing them into spatial proximity. To further investigate the function of the dgRNA-dCas9 RNP in organisms, we targeted the LacZYA locus in *Escherichia coli,* and measured the gene expression through a β-galactosidase assay. We observed that strains expressing dgRNA-dCas9 display more potent and prolonged inhibition of the LacZ gene, at the same level of the Lac repressors, in comparison to sgRNA-mediated CRISPR interference, implying a loop-mediated gene regulatory effect. Following this finding, we next investigated the physical searching space between the two guide regions of the dgRNA measured by its DNA length. Our results showed that the efficiency of the dgRNA-mediated regulatory effect was reduced with increasing distance between DNA target regions, indicating a limited searching space of the dgRNA-dCas9 RNP.

In general, we designed a bivalent dgRNA for dCas9 and studied its functional mechanism by mapping the molecular interactions between dgRNP and target DNA. We have demonstrated that by joining two separate sgRNAs together with an RNA linker, the dCas9 RNP could be devised for inducing DNA loop *in vitro* with sequence specificity, and for regulating gene expression *in vivo*.

Materials and methods

Strain construction

All *E. coli* strains for the β-galactosidase assay were derived from NEBExpress (NEB). The *E. coli* strains were transformed with a pdCas9-bacteria plasmid (a gift from Stanley Qi, Addgene plasmid #44 249, https://www.addgene.org/44249/) and a distinct guide RNA plasmid modified from a pgRNA-bacteria plasmid (a gift from Stanley Qi, Addgene #44 251, https://www.addgene.org/44251/). The transformation was carried out by heat shock at 42°C for 45 seconds. Transformed *E. coli* strains were subsequently selected on LB-agar plates with 100 μg/ml of Ampicillin (Sigma-Aldrich) and 34 μg/ml of Chloramphenicol (Thermo Fisher Scientific). To express each gRNA plasmid, DNA templates were ordered from Thermo Fisher Scientific and subcloned into the pgRNA-bacteria plasmid by replacing the original gRNA expressing region between SpeI and HindIII restriction sites. The gRNA plasmids expressing double dual guide RNAs were subcloned by BioCat GmbH. Sequences can be found in Supplementary Table S1.

Western blot and Coomassie staining

To check the expression levels of dCas9 under the different culture conditions, the cultures were centrifuged at 4000 rcf for 10 minutes and the supernatant was discarded. The cell pellets were frozen at –80°C and thawed to room temperature prior to the experiment. The cells were resuspended in a 1:1 mixture of 1× PBS and 2× Laemmli buffer (Bio-Rad, #1 610 737) containing 2-mercaptoethanol (Sigma-Aldrich, #M6250-100ML). To lyse the cells, the samples were sonicated for 5 min in intervals of 10 s on and 10 s off. Finally, the cell lysates were heated at 95°C for 2 min and centrifuged at 6000 rcf for 2 min. The total amount of protein on the samples was measured by absorbance at 280 nm with NanoDrop and 200 μg was loaded to 10% SDS-PAGE gels. The gels were run in 25 mM Tris base (Sigma-Aldrich, #T4661), 192 mM Glycine (VWR, #24403.367), 0.2% SDS (Sigma-Aldrich, #11 667 289 001) buffer for 30 min at 120 V followed by another 60 min at 200 V. The gels for whole cell lysate staining were incubated with 0.3% w/v Coomassie Brilliant Blue (Sigma-Aldrich, #27816-100G), 0.2 M citric acid (Sigma-Aldrich, #251275-100G); and de-stained overnight under rotation. The Coomassie stained gels were imaged using a GE LAS 4000 imager.

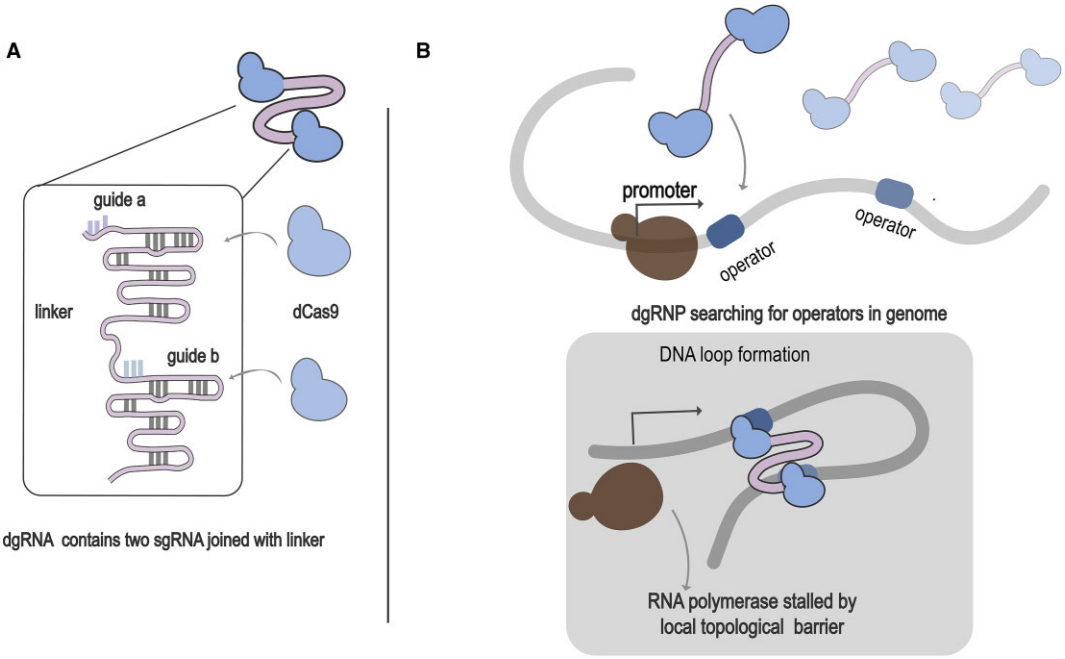


Figure 1. Re-designed bivalent guide RNA directing dCas9 to induce loop-mediated gene regulation in *E. coli*. (**A**) Each dgRNA comprises two sgRNAs with distinct guide sequences connected by an adjustable linker. The dgRNA recruits the dCas9 at stoichiometry of 1:2 to form a ribonucleoprotein. (**B**) The dgRNP imposes a regulatory effect on gene expression by modifying the genomic DNA topology. The dgRNP recognizes separate operators in the genome and forms hybrid products by bringing distal operators into spatial proximity to form DNA contacts or loops. The modified genomic topology affects gene expression by introducing topological complexity to genomic regions adjacent to the target gene.

The samples for western blot were transferred to a nitrocellulose blotting membrane (Cytiva) by applying a constant voltage of 20 V for 50 min. The membranes were blocked with 5% Blotting-Grade Blocker (Bio-Rad) diluted in TBST (0.1% Tween) for 1 hour at room temperature. Then, the membranes were incubated overnight at 4°C with mouse anti-Cas9 monoclonal IgG1 antibody (Thermo Fisher) diluted 1:500 in TBS 0.1% Tween (Sigma-Aldrich). The day after, the unbound antibodies were washed away with TBST three times for 15 min each. Then, the membranes were incubated for 2 hours at room temperature with Alexa 647-conjugated goat anti-mouse IgG antibody (Thermo Fisher) diluted 1:1000 in TBST. After washing with TBST, the membranes were imaged using a GE LAS 4000 imager.

β-Galactosidase assay

Before carrying out the β-galactosidase assay, 10 μl of overnight culture were inoculated into 90 μl of Lysogeny broth, LB (VWR), supplemented with 1 mM IPTG (VWR), 100 μg/ml of Ampicillin and 34 μg/ml of Chloramphenicol. When reaching mid log phase, the bacteria were induced with or without 2 μM of uTC and grown for another 2 h. A volume of 100 μl from each culture was measured by using the yeast β-galactosidase assay kit (Pierce) following the manufacturer's instructions. The absorbance at 420 nm and 600 nm was measured to calculate the β-galactosidase activity using the following equation:

$$\beta - galactosidase\ activity = \frac{1000 \times A_{420}}{t \times V \times OD_{600}}$$

t = time of incubation in minutes and V = volume of culture used for measurement in ml.

In vitro transcription of RNA

All gRNAs used for electrophoretic mobility shift assay (EMSA) and transmission electron microscopy (TEM) were produced by *in vitro* transcription using the T7 HighYieldAid kit (Thermo Fisher Scientific) for 4 h at 37°C, followed by digestion with DNAseI (Thermo Fisher Scientific). The transcripts were further purified with a spin column using the E.Z.N.A. total RNA kit (Omega Bio-Tek) following the instructions of the product.

Gel electrophoretic mobility shift assay

DNA substrates were prepared by amplifying DNA ultramers with Cy5- or Cy3-labelled primers at the 5′ end, ordered from Integrated DNA Technologies (IDT). PCR products were then subsequently purified using a PCR purification kit (Qiagen, #28 104). The RNPs were constituted at room temperature in dCas9 buffer consisting of 20 mM Tris–HCl (pH 7.5), 100 mM KCl, 5 mM $MgCl_2$, 5% glycerol and 1 mM DTT. For sgRNAs, the ratio of dCas9:sgRNA was 1:1.2 and for dgRNA the ratio of dCas9:dgRNA was 2.4:1 with a final RNA concentration of 250 nM. The RNPs were added to the fluorophore labelled DNA substrates at 100 nM with equal volume and incubated for 1 hour at 37°C. After incubation, the samples were run on a 4% native PAGE gel at 300 V for 20 min at 4°C in 0.5× TBE supplemented with 5 mM $MgCl_2$. After gel electrophoresis, the gels were imaged with a GE LAS 4000 imager. Sequences used for EMSA can be found in Supplementary Table S2.

Negative stain transmission electron microscopy

DNA substrates were prepared by amplifying DNA ultramers with desthiobiotin-labelled primers at the 5′ (IDT). The RNPs were constituted at room temperature in dCas9 buffer consist-

ing of 20 mM Tris–HCl (pH 7.5), 100 mM KCl, 5 mM $MgCl_2$, 5% glycerol, 1 mM DTT with a final concentration of 250 nM of 2.4:1 dCas9:dgRNA ratio. After preparation, the mixture of RNA and dCas9 was incubated at room temperature for 10 min, the labelled DNA substrates at 100 nM were added to RNPs with equal volume and incubated for 1 h at 37°C. The DNA in complex with the RNPs was bound to streptavidin MyOne C1 Dynabeads (Life Technologies) and washed twice with dCas9 binding buffer. The products bound to the Dynabeads were eluted with 5 mM D-Biotin (Thermo Fisher Scientific). The eluted sample was diluted to 5 nM of DNA concentration before being applied onto TEM grids. All samples were prepared with formvar coated copper grids (Electron Microscopy Sciences, #FCF400-CU), and were glow discharged before use. For the preparation of the TEM grids the sample was applied to the carbon-coated surface and incubated for 20 s followed by a wash with water for 4 s. The grids were subsequently stained with uranyl acetate for 20 s and air dried for 45 min before imaging. Micrographs were collected using SerialEM on a Talos 120C G2 fitted with a Ceta-D detector at a 73000x magnification. Subsequent data processing was carried out using CryoSPARC v4.0.0 (27). Particles were picked using template picking with templates generated using manually picked particles from the s20 dataset. The particles were then cleaned using two rounds of 2D classification, with the final 2D classes used in the analysis selected from the second round. Measurement of distances was carried out by comparison to a published atomic model of dCas9 (PDB: 6K3Z) using UCSF Chimera (28) and ImageJ (29).

Atomic force microscopy

A freshly cleaved mica surface was pre-treated with 5 mM Nickel sulfate for 1 min at room temperature and washed three times with water before being dried with a nitrogen airflow. The sample was immediately deposited on the mica surface for 1 min and washed with water three times before being air dried with a nitrogen airflow. Atomic force microscopy (AFM) images were taken with a JPK instruments Nanowizard 3 ultra with an Olympus Biolever mini cantilever in AC mode. Bruker SCANASYST-AIR probes (Bruker) with a spring constant of 0.4 N/m were used for imaging with driving frequency at ∼80kHz.

Chromosome conformation capture assay (3C)

For the 3C assay, the bacteria was grown as described for the β-galactosidase assay. Before fixation, the OD_{600} was adjusted to 0.6 for all the samples, and 5 ml of each bacterial culture was cross-linked with 1% final v/v of formaldehyde for 10 min at room temperature. The fixation reaction was quenched with glycine (final concentration of 130 mM) for 5 min at room temperature followed by an additional 10 min on ice. The cells were pelleted by centrifugation at 10 000 rcf and 4°C for 15 min, and washed with TE buffer (10 mM Tris, 2 mM EDTA, pH 8.0). After another centrifugation step, the cells were resuspended in 400 µl of 25 mg/ml lysozyme (Thermo Fisher Scientific, #10 076 813) in TE buffer and lysed for 10 min at room temperature. To denature the proteins on the sample, 10% SDS was added to a final concentration of 0.5% and incubated for 5 min at 65°C followed by 10 min at 37°C. The SDS was quenched by adding 10% Triton X-100 to a final concentration of 1%. In the first 3C experiment, the samples were digested overnight in a 200 µl reaction containing 120 µl of the cell lysate, 480 U of HhaI (NEB, #R0139L), and 1x rCutSmart™ Buffer (NEB, #B6004S). The digestion reaction was carried out for 18 h at 37°C, before adding an additional 480 U of HhaI and incubating for 4 more hours at 37°C. For the second round of 3C experiment, the samples were incubated with a combination of 480 U of Alu I (Thermo Scientific, #ER0012) and 480 U of HaeIII (NEB, #R0108M). After overnight incubation, the volume of the reaction was adjusted to 500 µl and incubated for an additional 5 h at 37°C. To terminate the digestion, 10% SDS was added to a final concentration of 0.5% for 20 min at 80°C while shaking at 1000 rpm. Next, the samples were ligated for 16 h at 16°C and 300 rpm in a 800 µl reaction containing final concentrations of 0.9% Triton X-100, 0.1 mg/ml BSA, 10 000 U of T4 DNA Ligase (NEB, #M0202M) and 1× T4 DNA Ligase Reaction Buffer with 10 mM ATP (NEB, #B0202S). The following day, the reaction was incubated for another hour at 25°C while shaking at 450 rpm and terminated by adding a final concentration of 10 mM EDTA and incubating for 20 min at 85°C. To remove the RNA, the samples were treated with a final concentration of 0.1 mg/ml of RNase I (Thermo Fisher Scientific, #EN0601) for 45 min at 37°C and 450 rpm. The cross-linking was reversed from the digested and ligated samples with a final concentration of 1 mg/ml of Proteinase K (Thermo Fisher Scientific, #EO0491) and incubation for 16 h at 53°C and 450 rpm. The DNA was extracted by adding an equal volume of phenol-chloroform-isoamyl alcohol (Sigma Aldrich), mixing by inversions and centrifuging at maximum speed for 15 min. The upper aqueous phase was transferred to a new tube and the DNA was precipitated with a volume excess of ice-cold absolute ethanol, incubation at –20°C overnight and pelleting by centrifugation at maximum speed for 15 min at 4°C. The DNA pellet was washed with 70% ethanol once, centrifuged again and air dried. The final DNA pellet was resuspended in TE buffer and the concentration was measured by Qubit using the Invitrogen Qubit 1× dsDNA High Sensitivity (HS) and Broad Range (BR) Assay Kit (Thermo Fisher Scientific, #Q33230) according to the instructions of the manufacturer. The digestion and ligation efficiency was examined by 1.5% agarose gel electrophoresis. The relative contact frequencies between the O3 proximal site P2 and several sites along the region of interest were determined by analyzing 50 ng of each DNA sample with qPCR using the Luna Universal qPCR Master Mix (NEB, #M3003) according to the kit's instructions.

Design of primers used for 3C-qPCR was illustrated in Figure 5A, sequences of which can be found in Supplementary Table S3. A pair of 3C primers was used to normalize the baseline contact frequency as technical control by targeting HhaI or HaeIII + AluI-digested fragments outside of the dgRNA-targeted region. The quantity of contacts was thus calculated as: normalized quantity $= 2^{(Ct(ROI)-Ct(TC))}$, from C_t values derived from primers for the genomic region of interest (ROI) and outside the region as technical control (TC).

Quantification and statistical analysis

All data from β-galactosidase assay in this work were collected from three independent biological replicates. All statistical analyses were presented as means ± SD. Significant differences between the values under different experimental conditions were subjected to two-tailed Student's *t*-test. For all tests, P value < 0.05 was considered as statistically

significant. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$ and ns, no significance versus the control groups.

Results

A gRNA consisting of two sgRNAs connected by an RNA linker demonstrates sequence bispecificity

The sgRNA in the CRISPR-Cas9 system is an engineered RNA derived from the fusion of the 3′ end of crRNA to the 5′ end of tracrRNA by a hairpin, to mimic the native tertiary structure essential for Cas9 activation and DNA cleavage (8) (Figure 2A). To avoid perturbing the functional conformation of each sgRNA, we used an RNA linker of a poly-adenine sequence connecting the 5′ end of 'guide a' to the 3′ end of 'guide b' (Figure 2A, Supplementary Figures S1 and S9).

First, we investigated the function of the dgRNA by validating that it can recruit dCas9 to form an active RNP to recognize and bind to two DNA targets. We designed the dgRNA with two distinct guides, 'a' and 'b' (depicted in Figure 2A), and synthesised the corresponding double-stranded DNA substrates 'a' and 'b', each labeled with Cy5 and Cy3 respectively. Our experimental procedure involved titrating dCas9 over the dgRNA, utilizing a Cy5-labeled DNA substrate 'a' as a fluorescent tag to track the formation of the RNP complex. Gel electrophoresis analysis of the resulting products revealed a shift in mobility from faster to slower speeds with increasing concentrations of dCas9. This shift indicated the progressive formation of the dgRNP, comprised of two dCas9 moieties (as illustrated in Figure 2C). Additionally, at higher concentrations of dCas9, we observed a faint band above the primary products, suggesting the possibility of alternative conformations adopted by the dgRNP:DNA complex. To comprehensively assess the binding dynamics of the dgRNA to the two distinct DNA targets, we employed an electrophoretic mobility shift assay (EMSA) to analyze the products resulting from the interaction of the dgRNP with DNA substrates (as depicted in Figure 2B and D). As a control, we included a sgRNA with guide sequence 'a' to distinguish the stoichiometry of dCas9 in the products from the different sample combinations. As shown in Figure 2D, the dgRNP effectively bound to both Cy5- labelled 'a' and Cy3-labelled 'b' DNAs confirming the binding activity of both guide RNA units of the dgRNA construct. Furthermore, in the presence of both DNA substrates, we observed additional bands formed with slower mobility and showing overlapping Cy5 and Cy3 fluorescence. This observation suggests that the dgRNP is capable of simultaneously engaging with both DNA targets. Additionally, minor products were detected above the major bands in both fluorescent channels, consistent with the minor conformations observed in Figure 2C, further supporting the notion of alternate dCas9 conformations in complex with target DNA. We extended our investigation by examining the binding of the dgRNP to a DNA substrate with both target sites separated by a 119 bp spacing (as depicted in Supplementary Figure S2C). Analysis of the EMSA results revealed products formed by the dgRNP bound to the DNA substrate showing faster mobility compared to the DNA substrate incubated with two separate sgRNPs, indicative of DNA loop formation. The EMSA assay confirmed the formation of products with the expected stoichiometry between dgRNP and the target DNA, highlighting the specificity of the interaction. However, the detection of products with multiple conformations raises intriguing questions regarding the detailed structural arrangement of the formed complex, warranting further exploration into its structural details.

We conducted further experiments to evaluate the endonuclease activity of a Cas9:dgRNA RNP, aiming to fully elucidate the functionality of the dgRNA. We systematically titrated Cas9 to the dgRNA and incubated the resulting RNP with Cy5-labelled substrates 'a' and 'b' separately. Gel electrophoresis analysis of the reaction products revealed a notable reduction in cleavage efficiency for substrate 'b' compared to substrate 'a' (as shown in Supplementary Figure S2). Extending the incubation time of the dgRNP with substrate 'b' led to improvement in cleavage, yet a substantial portion remained uncleaved. This disparity between the binding and cleavage efficiencies of dgRNP with substrate 'b' corroborate the findings from previous studies regarding the impact of 5′ end modifications of the sgRNA (30). Specifically, the addition of a linker to the 5′ end of the sgRNA, akin to the structure of guide 'b' in dgRNA, minimally affected the formation of the R-loop of the Cas9 RNP with the target DNA, but significantly impaired the cleavage rate.

Combining the insights gained from this experiment with our previous findings we elucidated the influence of the RNA linker on the dgRNA functionality. While the presence of the linker facilitated simultaneous binding to a pair of target sequences, it also resulted in compromised functionality when coupled with Cas9. This observation underscores the intricate interplay between guide RNA design and enzyme activity of Cas9.

dgRNA mediates DNA loop formation *in vitro*

We delved deeper into the binding capabilities of the dgRNP to a DNA substrate harboring two distal target sites. Utilizing linearized pUC19 as our substrate, dgRNA was engineered to target two sites that are approximately 2400 bp apart. Given the potential necessity for flexibility between the two guide RNA units of the dgRNA when binding to two sites on the same DNA molecule, we systematically varied the RNA linker length from 20 to 40 nucleotides. Employing EMSA, we visualized the complexes formed by pUC19 bound with dgRNP, observing mobility patterns akin to those of pUC19 bound with two separate sgRNAs targeting the same pair of sites (Figure 3A). This indicated a stoichiometry of dCas9 to target DNA of 2:1.

To gain structural insights into these complexes, we purified the DNA bound with dgRNP and subjected it to negative-stain transmission electron microscopy (NS-TEM). Our analysis revealed the connection of two dCas9 molecules bridged by RNA, indicative of looped DNA. Notably, dgRNP samples with varying linker lengths exhibited an average distance between the mass centers of the dCas9 units of around 100 Å (Figure 3B), confirming their connection by the RNA linker. However, increasing the RNA linker length did not lead to a noticeable increase in the mean distance between dCas9 molecules (Figure 3B, Supplementary Figure S3E), possibly due to resolution limitations or biases inherent of NS-TEM.

To determine the capability of the dgRNP to induce DNA loop formation, we imaged a linearized pUC19 bound with dgRNP using atomic force microscopy (AFM). We tested the dgRNA with a 40 nt linker targeting two pairs of sites located 1200 and 2300 bp apart on the linearized pUC19. As controls for comparison we also imaged the dgRNP and lin-

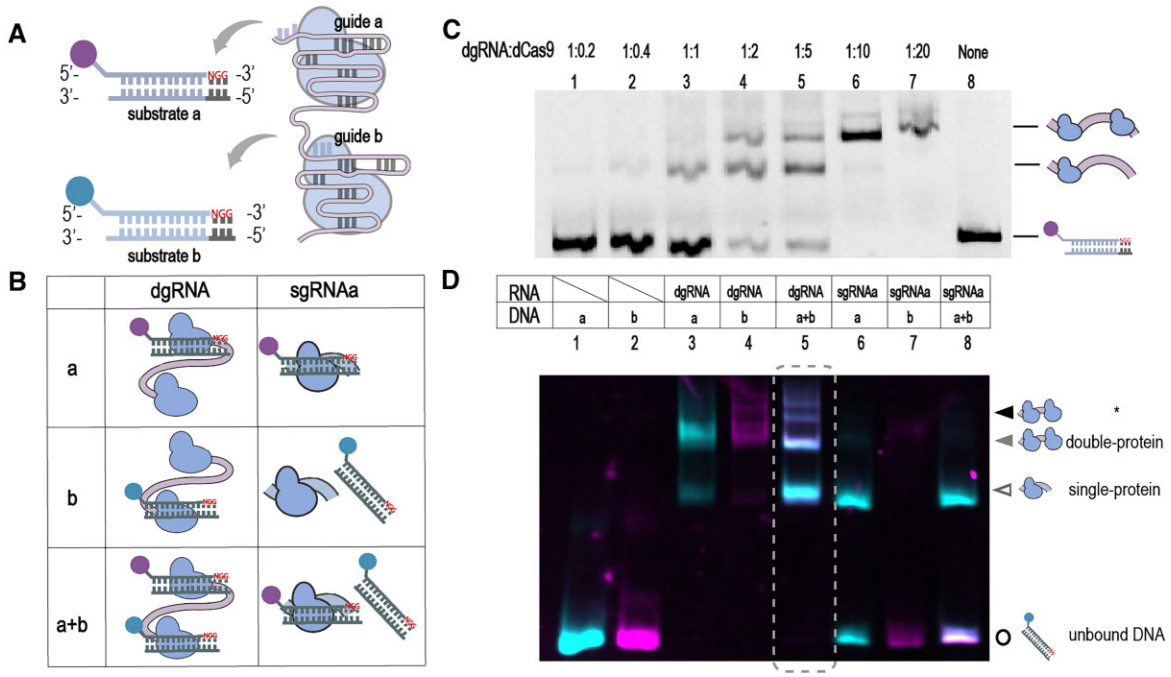


Figure 2. The dgRNA demonstrates sequence bi-specificity for DNA targets. (**A**) Illustration of the dgRNA guiding dCas9 to recognize separate DNA targets. The dgRNA consists of two sgRNA, with a 5′ end-proximal guide sequence referred to as 'guide a' and a 3′ end-proximal guide referred to as 'guide b'. (**B**) Schematics for the assay in (D). Two separate DNA substrates, substrate 'a' and substrate 'b', were used to keep track of the dgRNP-DNA complex formation. Each substrate contains the target sequence corresponding to two guides of dgRNA and were labelled with Cy5 or Cy3 fluorophores. The dgRNP was tested under three conditions, including with Cy5-labelled substrate 'a', with Cy3-labelled substrate 'b', or with both substrates 'a + b'. The sgRNP with guide 'a' was tested under the same conditions as a control. (**C**) EMSA measuring the titration of dCas9 over the dgRNP while binding to a Cy5-labelled single DNA substrate. (**D**) EMSA visualizing the dgRNP binding to two separate DNA targets simultaneously. Lanes 1–2 show labelled DNA substrates, lanes 3–8 show the corresponding products formed by the RNP and DNA substrate 'a' or 'b' as indicated above. Open circle, fluorophore-labelled DNA substrates; open triangle, products associated with a single protein; solid grey triangle, products associated with double proteins; solid black triangle, products associated with double proteins in minor conformation. Dashed box highlights the sample of dgRNP incubated with both substrates. Products were visualized using native PAGE gel; images were taken in Cy3 and Cy5 channels and merged, with uncropped images in separate channels shown in Supplementary Figure S2.

earised plasmid (Figure 3D). The dgRNPs were incubated with pUC19 before being imaged by AFM. Compared to pure pUC19, we observed dgRNP binding to pUC19 with a few binding modes. As expected, using AFM we could observe the dgRNP bound to both targets, and the two targeted sites with dCas9 were contracted by the RNA linker to form a loop (Figure 3C, Supplementary Figure S4A, B). For the dgRNPs targeting different sites on pUC19, the loop lengths were consistent with the designed distance between binding sites. For the dgRNP targeting two sites at approximately 150 bp from ends of linearized pUC19, separated by approximately 2400 bp, we observed recircularization of linearized pUC19. We observed smaller DNA loop formation with dgRNP targeting sites separated by 1200 bp and located ∼800 bp away from both ends. The formation of a DNA loop at targeted sites with the anticipated size demonstrated the programmability of the dgRNP to induce DNA loops.

From the AFM images we observed several products formed between dgRNP and linearized DNA, including DNA loop, DNA bound with dgRNP at one target site, and DNA bound with two dgRNPs at both sites in tandem (Figure 3E and F, Supplementary Figures S5 and S6). To better analyse the frequency of the DNA loop detected in the sample, we quantified the number of products belonging to each category across the field-of-view. DNA bound with dgRNP at one site was further identified as two subcategories: with one protein and with tethered protein (Figure 3E-F). The difference was in the morphology of the RNP imaged at the binding site and was included to avoid bias due to ambiguity. Notably, the observed one protein could be a result of the limited resolution of AFM tips to resolve two dCas9s connected by an RNA linker at given sites. DNA bound with dgRNPs at both sites in tandem was identified as DNA bound with more than two proteins, again to avoid the ambiguity in the number of proteins per site. From the quantification, we observed various frequencies of DNA loop with different dgRNPs. With the dgRNP recognising targets separated by 160 bp, the yield was around 50% (Figure 3E). When the distance was increased to 2100 bp, the yield dropped to 12% (Figure 3F). Such discovery revealed that the efficiency of loop formation could be largely affected by the distance between target sites.

While AFM data enriched our understanding by offering the topographical details of dgRNPs bound with DNA, it is essential to acknowledge the limitations in interpreting this data. Differentiating genuine DNA loops from other structural features poses a challenge, thus our quantitative analysis of various structural categories offers statistics under the best-case scenario for looping efficiency. Overall, the quantitative analysis from AFM, coupled with TEM data, underscores the capability of the dgRNA to form dgRNP complexes with two dCas9 units, facilitating DNA loop formation at specific sites *in vitro*.

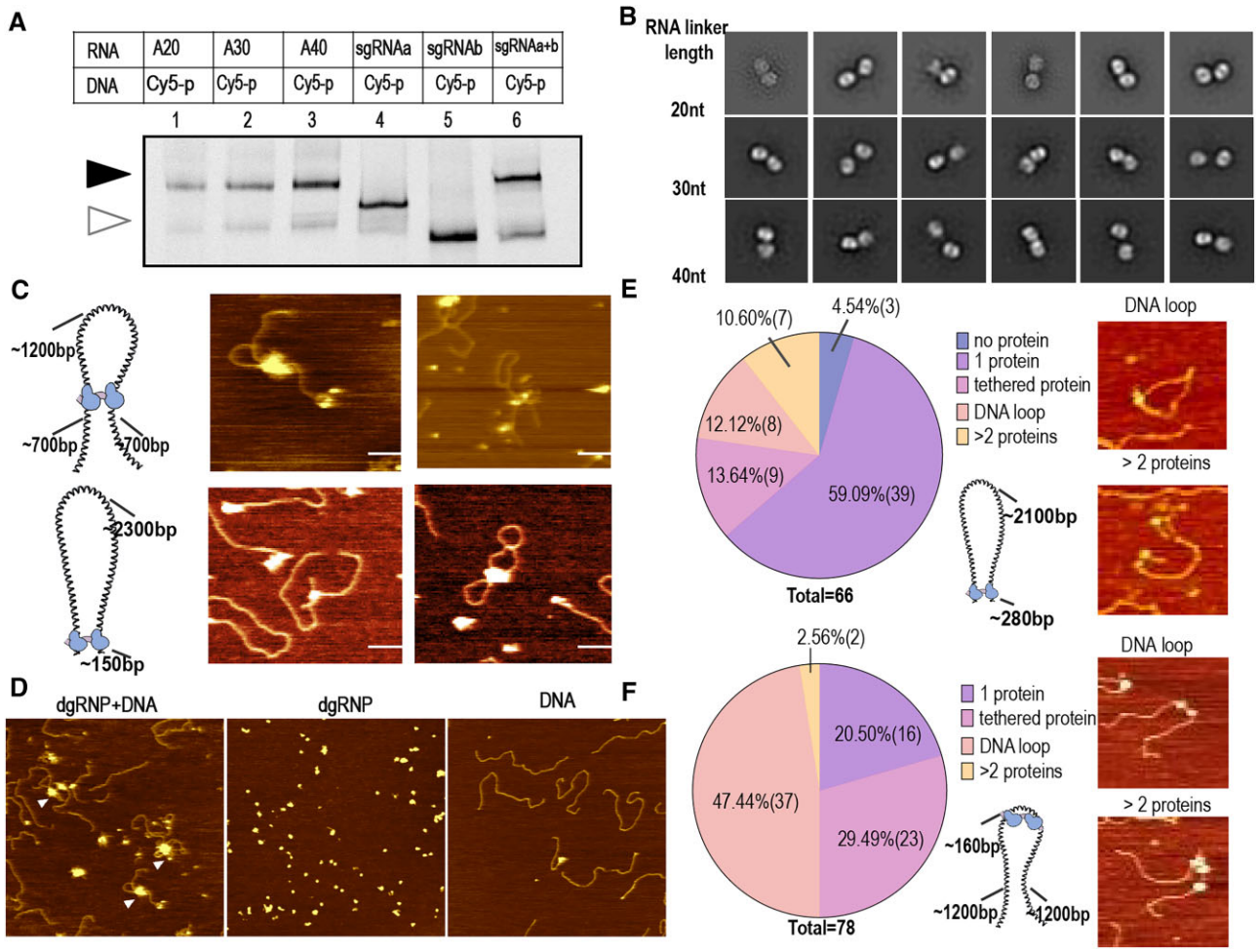


Figure 3. dgRNP induces DNA loop formation *in vitro*. (**A**) EMSA measuring dgRNP of varied linker lengths binding to DNA substrates. A DNA substrate with Cy5-label was used for monitoring the mobility of the RNP-bound complexes. dgRNP with linker length of 20 nt, 30 nt, 40 nt formed a complex with DNA (lanes 1–3, solid black triangle). Lanes 4–5 show Cy5-DNA bound with sgRNA-a and sgRNA-b (open triangle). Lane 6 shows Cy5-DNA bound with sgRNP-a and sgRNP-b (solid black triangle). (**B**) 2D class averages from NS-TEM of purified dgRNP-bound to DNA. Each row shows classes of dgRNPs with corresponding linker length. Each class shows the density from the two dCas9 proteins of the complex, with undetectable signals from RNA or DNA. (**C**) Cropped AFM images of dgRNP targeting at different pairs of sites on the same linearized DNA substrate. Left panel shows cartoon illustrations of the target sites on the substrate, with cropped AFM images of the DNA loop formed between dgRNP DNA substrates on the right; uncropped images shown in Supplementary Figure S4. Scale bars, 100 nm. (**D**) AFM images of the DNA substrate incubated with dgRNP, dgRNP alone and linearized DNA substrates alone. White triangles indicate dgRNP. (**E, F**) Characterization of the products formed between dgRNP and the DNA substrate from AFM images. Each sector corresponds to a category of products identified from AFM images, with cropped images of representative complexes from four categories. The percentages of each product type are shown on a graph pie, with the actual counts of each category from each sample are shown in brackets. Cartoon illustrations depict the target sites of corresponding dgRNPs on DNA substrates. Uncropped AFM images are shown in Supplementary Figures S5 and S6.

dgRNA directing dCas9 to recapitulate DNA loop-mediated gene repression of LacZ in *E. coli*

As the dgRNA demonstrated the capability of loop induction, we next tested whether the dgRNA can induce loop formation and regulate gene expression in a cellular environment. To study the function of the dgRNP *in vivo*, we used a canonical DNA loop model, the Lac operon in *E. coli* (31). The Lac operon consists of a set of genes essential for lactose metabolism, of which the LacZ gene expression is regulated by a Lac repressor (LacI) via a loop-mediated mechanism. In the absence of allolactose, or its analogue IPTG, Lac repressors recognize the primary operator O1 and one of the auxiliary operators O2 or O3 flanking the promoter of LacZ (Figure 4A). The Lac repressor binding to separate operators could dimerize to form a tetramer inducing DNA loop formation between operators. Allolactose or IPTG binds and sequesters the Lac repressors, forcing their release from the operators, and resulting in an increased LacZ expression. Therefore, by using IPTG, the operators are free to be bound by the dgRNP.

By co-expressing the dCas9 and the dgRNA targeting neighbour sites of the two auxiliary operators O2 and O3, we studied the regulatory effect of dgRNP on the LacZ expression. It has been reported that the target strand affects the binding stability of dCas9, potentially due to the competition between RNA polymerase and dCas9 (9). Therefore, we tested targeting either the template strand or non-template strand of the LacZ gene. To further compare the effect between sgRNP and dgRNP, we constructed *E. coli* strains expressing sgRNAs targeting the same neighbouring sites of O2 or O3 operators. Constructed *E. coli* strains expressing sgRNP targeting O2 and O3 on the non-template strand of LacZ were termed as FO2 and FO3, and targeting O2 and O3 on the

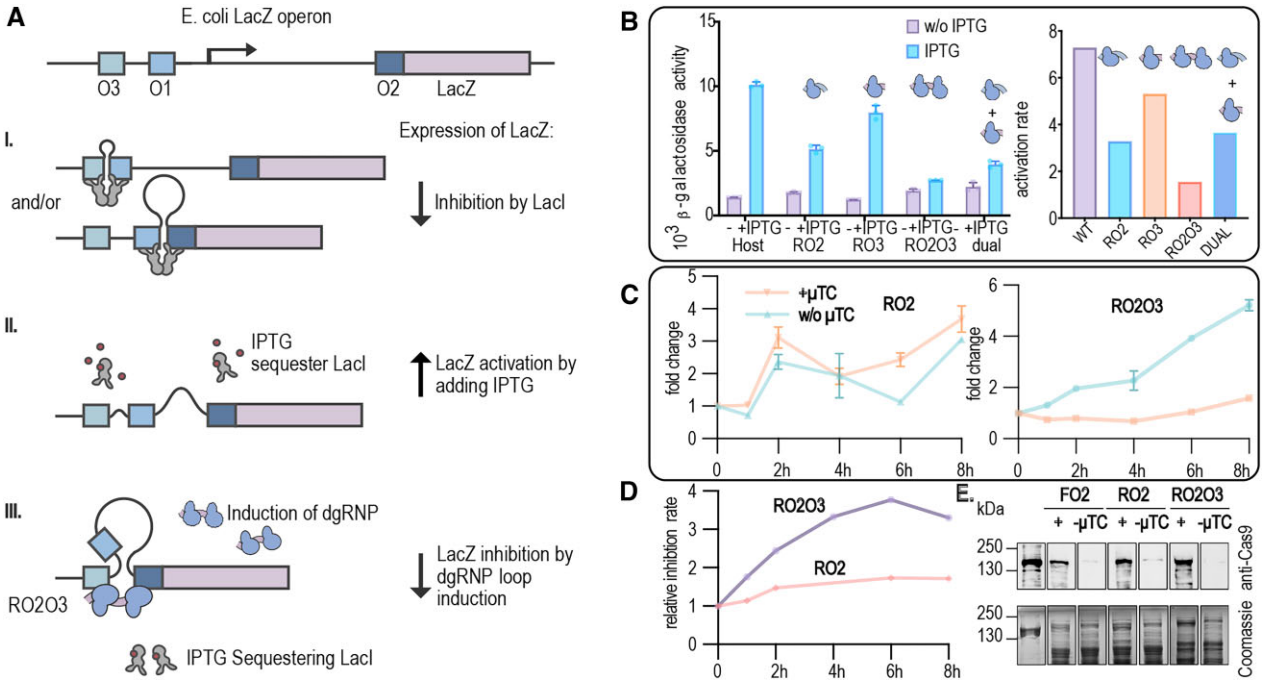


Figure 4. dgRNP represses LacZ expression in *E. coli* with a mechanism distinct from normal CRISPRi. (**A**) Schematic illustration for LacZ regulation of *E. coli* in a native system and by dgRNP. In *E. coli,* operator O1, the primary operator overlapping with the LacZ promoter; O2 and O3, the auxiliary operators; IPTG, the analog of allolactose. I, II and III-panels show the hypothesised influence of LacI and dgRNP on the DNA and the resulting LacZ expression. (**B**) β-Galactosidase activity measurement of *E. coli* expressing the gRNA targeting operator O2 and O3. RO2, strain expressing sgRNA targeting the sequence 15 bp upstream of O2 on the template strand of LacZ; RO2O3, strain expressing the dgRNA targeting both sequences 15 bp upstream of O2 and 21 bp upstream of O3 on the template strand of LacZ; dual, strain expressing two separate sgRNAs targeting the same sequences as RO2O3. Constructed strains were induced to express dCas9 with µTC 2 h before adding IPTG. The β-galactosidase activity was measured 2 h after adding IPTG. Mean values and SD from biological triplicates are shown. Calculated activation rate by IPTG is provided on the right panel. Two-sided *t*-test was performed to compare the expression level between the different conditions. (**C**) Time-course measurement of the β-galactosidase activity fold change in constructed *E. coli* strains after adding IPTG. Calculated fold change of the β-galactosidase activity was plotted against time. Mean values and SD of biological triplicates are shown. (**D**) Calculated inhibition rate by the RNP during an 8 h-time course in the constructed *E. coli* strains after adding IPTG. Inhibition rate was calculated by the ratio of β-galactosidase activity of each strain without dCas9 induction versus dCas9 induction. (**E**) Western blot of dCas9 in the constructed strains. Top panel shows the Alexa 647 signal from a secondary antibody against the anti-dCas9 primary antibody. Bottom panel shows Coomassie staining of the same amount of total protein from corresponding strains. Uncropped images can be found in Supplementary Figure S8.

template strand termed as RO2 and RO3. Correspondingly, strains expressing dgRNPs targeting both O2 and O3 on the template strand were referred to as RO2O3, and FO2O3 referring to strains expressing dgRNPs targeting both sites on the non-template strand. After the induction of dCas9 expression, we performed a β-galactosidase assay to measure the expression level of LacZ in strains expressing different guide RNAs. Our results showed that when targeting the non-template strand of LacZ, adding IPTG failed to increase β-galactosidase activity in FO2 and FO3 (Supplementary Figure S7), while in the wild type strain the IPTG led to a significant increase. The same effect was seen with FO2O3, with almost unchanged β-galactosidase activity with IPTG (Figure 4B, Supplementary Figure S7). By contrast, RO2O3 showed significantly lower activation rate after adding IPTG compared to RO2 and RO3, indicating a more potent repression mechanism in addition to a road blocking effect by the sgRNPs (9). We measured the expression level of dCas9 in sgRNP and dgRNP expressing strains with Western blot before and after induction with µTC (Figure 4E), showing comparable dCas9 expression level across strains. To distinguish the increased repression effect from the dosage effect due to dgRNPs binding to both targets in tandem as seen *in vitro*, we also included a control strain co-expressing both sgRNA-O2 and sgRNA-O3 referred to as dual-sgRNA. We observed reduced activation rate of LacZ in the dual-sgRNA strain compared to the strain expressing only sgRNA-O3, comparable to the strain expressing sgRNA-O2 (Figure 4B). But the activation rate was still higher than the strain expressing dgRNA. The more potent repression of dgRNA thus further implied an additional mechanism other than simply a road blocking effect by the sgRNP. In addition to co-expressing the two sgRNAs, we expanded our approach by constructing strains co-expressing two separate dgRNAs. Each dgRNA was designed to target either RO2 or RO3, along with a sequence absent in the host genome (Supplementary Figure S12A, B). This dual-dgRNA system enabled recognition of both RO2 and RO3 target sites while simultaneously tethering dCas9 to a missing target sequence. Our findings revealed that the dual-dgRNA strategy resulted in enhanced repression of LacZ compared to dual-sgRNA, albeit still exhibiting weaker inhibition compared to dgRNA-RO2O3. Our results demonstrated that the dgRNP could recapitulate the LacZ repression mechanism of the Lac repressor, implying the formation of a DNA loop between targeted sites.

We next monitored the repression rate of LacZ along an 8 h time course. We monitored the fold change of LacZ expression at different timepoints after adding IPTG by mea-

suring the expression level of LacZ in RO2O3 and RO2. Our results showed that dgRNP can sustain potent repression of LacZ up to 8 h after dCas9 induction with μTC, compared to a cumulative increase of β-galactosidase activity without dCas9 induction. While sgRNP showed a synchronized fold change of LacZ at conditions with dCas9 induction or without (Figure 4C). We also calculated the inhibition rate of dgRNP and sgRNP, showing a higher repression rate induced by dgRNP compared to sgRNP along the 8 h period. Furthermore, the dgRNP sustained the repression of LacZ up to 6 h with a mild drop of repression rate at 8 h (Figure 4D). We hypothesize that the mild repression by sgRNP targeting the template strand could be explained by helicase activity of RNA polymerase destabilizing the RNA:DNA duplex formed between the sgRNA and the template strand (9). The potent and sustained repression induced by the dgRNP suggests that the dgRNP forms a more stable complex with the genomic DNA via a loop-mediated mechanism, enabling higher resistance to destabilization factors *in situ* such as RNA polymerase by modifying local genomic topology. In summary, these results demonstrate that the dgRNP could recapitulate the Lac repressor-mediated LacZ repression, implying the dgRNP-mediated DNA loop formation *in situ*. These results further demonstrated the loop formation induced by our dCas9-dgRNA system.

In addition, we conducted Chromosome Conformation Capture assays (3C) to investigate the local genome organization within the targeted genomic region affected by the dgRNP (Supplementary Figure S10A). Contact frequencies between the O3 proximal site (P2) and various sites across the adjacent genomic region, including the O2 proximal site (P4), were measured across multiple samples. While we could detect increased likelihood for loop contacts with our dgRNA compared to sgRNA, the control with the wild-type LacI gave no indication of looping with 3C in our hands. It appears that the 3C assay failed to resolve the native DNA loop mediated by LacI in the host strain (Supplementary Figure S10B, C), even when increasing the digestion frequency of the region (Supplementary Figure S11B), as evidenced by consistent contact frequencies between P2 and P4 regardless of IPTG treatment (Supplementary Figures S10B, C and S11B). This could be due to the main repression contribution of the LacI system coming from the induction of a small 91 bp loop between O3 and O1 (like in Figure 4A II, top), which would probably be too short to be resolved with 3C. Nevertheless, this underscores the persistent challenge of resolving DNA loops at the scale of hundreds of base pairs, or lower, in *E. coli*, emphasizing the need for further advancements in research tools within the field.

Distance between the targeted genomic regions affects the efficiency of LacZ repression

To test the effect of distance between anchor sites on loop induction, we programmed the dgRNA to target O2 and a series of distal target sites across the LacZYA operon. All target sites were distributed across the LacZYA locus downstream of O2 and were named by ascending number from O4 to O8 with increasing distance from O2 (Figure 5A). Strains expressing the dgRNP targeting O2 and one of the other target sites on the template strand were termed as RO2O4-RO2O8. We again measured the β-galactosidase activity of these dgRNA variants in the different strains to monitor the expression level of LacZ after adding IPTG. For all strains, we observed higher levels of LacZ repression compared to strains expressing sgRNP (RO2) (Figure 5B).

Interestingly, despite all dgRNP showing higher repression rate of LacZ compared to the sgRNP, the repression rate was not determined by the distance between two target sites (Figure 5C). For dgRNA-RO2O8, targeting genomic sites spaced more than 4500 bp apart, the repression rate of LacZ was higher than for strains expressing dgRNA-RO2O3, with target sites located less than 500 bp apart (Figure 5B). For the dgRNA-RO2O7, with the target site O7 being 4000 bp away from O2 and less than 500 bp away from O8, the repression rate of LacZ after dCas9 induction was significantly lower than for dgRNP-RO2O8. Different LacZ repression rates here reflected a varying efficiency of the dgRNP in inducing a DNA loop. Since all dgRNAs tested here were anchored to the target site O2, varied efficiency was mainly affected by the other target site. As the other target sites were not overlapping with known regulatory elements of LacZ, we assumed that the distance between the two sites played an important role in altering loop induction efficiency by the dgRNP. Variations of the relative inhibition rate were observed with varied distance between target sites of each strain but showed vague correlation (Figure 5C), indicating that inter-distance could play some role in loop formation ability. The unclear correlation between distance and inhibition rate could be attributed to distal genomic regions organized into spatial proximity. Furthermore, altered local topology of genomic DNA at target sites O4-O8 could affect the accessibility of the dgRNP to the targets, and thus affect the inhibition rate of the dgRNP. Therefore, the efficiency of loop formation mediated by the dgRNP *in vivo* was confounded by the specific local genomic context when varying distance between two target sites.

Discussion

The dgRNA-dCas9 system we introduced here constitutes a new method to induce DNA contacts in living organisms. The system exploits the high programmability of RNA and expands on the functionality of the guide RNA in the CRISPR system by increasing the DNA binding valency. The method is concise in the molecular elements required, enabled by a bi-specific guide RNA, and flexible in targeting arbitrary genomic regions as typical for the CRISPR-dCas9 system. It has been previously reported that the 5′ end of the sgRNA is sensitive to modification (30). In our study, we integrated one of the sgRNAs from the 5′ end to the linker and observed similar binding efficiency of the dgRNA to the target DNA, compared to its sgRNA counterparts. However, we noted a significant reduction in the cleavage efficiency of substrate ‘b’ compared to substrate ‘a’ (Supplementary Figure S2D). These findings are consistent with previous research indicating that 5′ modifications greatly diminish the cleavage rate of Cas9, although they only modestly affect R-loop formation between the RNP and the target DNA. Considering the structural dynamics of Cas9 RNP activation for cleavage, it is plausible that the 5′ modification of the guide RNA may have disrupted the transition of Cas9 from a catalytically inactive checkpoint state to a catalytically competent conformation (32). This transition has been observed to be impeded by mismatches at positions 19 and 20 distal to the PAM, underscoring the significance of the DNA:RNA heteroduplex at the 5′ end of the guide RNA in facilitating the catalytic activation of the complex (33).

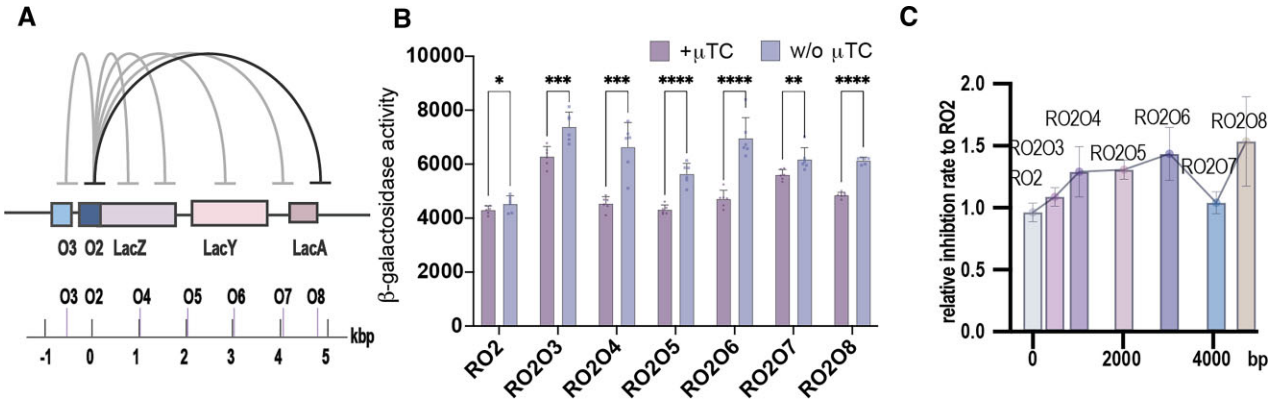


Figure 5. Distance between target sites affects the efficiency of the dgRNP-mediated LacZ inhibition. (**A**) Schematic representations of the DNA target sites location at the LacZYA locus in each *E. coli* strain. Lower panel shows the distance of each target site in relation to O2 in kilobase pairs (kbp). Target sites are termed as O3-O8, with the ascending number suggesting a longer distance from O2. The pair of targeted sites are connected by grey and black curves. (**B**) Measurements of β-galactosidase activity showing varied repression levels of LacZ from strains with different targeting sites. Each plot shows the measurements of the strains expressing the corresponding dgRNAs illustrated in (A) against the strain expressing sgRNP-RO2. Mean value and SD of biological replicates are plotted ($n > 3$). +μTC, after adding anhydro-tetracycline to induce dCas9 expression; w/o μTC, without adding anhydro-tetracycline. (**C**) Calculated inhibition rate of the dgRNP plotted against the genomic distance between target sites. Mean value and SD of biological triplicates were included.

Furthermore, consistent with the observed preference for the catalytically inactive conformation of the dgRNP when interacting with substrate 'b', we detected additional bands above the primary products in the EMSA assay. These additional gel bands were apparent only when the dgRNP was incubated with substrate 'b', while they remained undetectable in the presence of substrate 'a' alone (Figure 2C, Supplementary Figure S2B). The emergence of these gel bands suggests a shifted conformational landscape of the Cas9:dgRNA:DNA complex with substrate 'b', indicative of a more prevalent pre-active checkpoint state compared to the complex formed with substrate 'a'.

The biochemical intricacies of the dgRNP are readily observable and manipulable in controlled *in vitro* settings. However, the dynamic process of its target search and DNA loop induction within the vast genome presents challenges for direct observation. A fundamental disparity between *in vitro* assays and genomic functionality lies in the expansive search space, spanning several mega-base pairs in *E. coli*. To elucidate the functional dynamics encompassing target search, DNA binding, and DNA loop induction of the dgRNA, we propose a hypothetical model. Here, the initial binding to the first DNA target confines the physical search space of the complex, thereby dictating the probability of DNA loop formation (Figure 6A, B). This binding event not only narrows the search space for subsequent guide RNP units but also expedites the discovery of the second target compared to independent searches with sgRNPs. Conversely, it limits the physical accessibility for the remaining guide RNP units.

Another critical difference from *in vitro* binding experiments is the singular presence of the target sequence within *E. coli*, leading to competition among dgRNPs for binding. In scenarios where the second binding site is within reach of an unpaired guide, which is tethered to its pair bound to the genome at the first target site, the likelihood of this dgRNP encountering a second target is heightened (Figure 5B). Prior to binding the second target site, the interaction of the tethered dgRNP with the template strand is weaker, being more similar to that of the sgRNP, making it more susceptible to destabilization by the RNA polymerase. Consequently, the target site could be re-interrogated by the dgRNP. Upon binding to the second target, the resultant DNA loop acts as a topological barrier akin to the LacI dimer, impeding the unwinding machinery, like RNA polymerase, to exert its function. This leads to a robust repression of LacZ expression (Figure 4B), showcasing the functional consequences of DNA loop induction by the dgRNP within the intricate genomic landscape of *E. coli*.

We delved deeper into conditions where the second target sequence lies beyond the reach of the tethered RNP, to determine whether the tethered dgRNP could exert additional effects on the LacZ gene beyond weak repression attributed to unstable genome binding. Through co-expression of two dgRNA, each targeting RO2 or RO3, alongside a guide sequence devoid of targets in the *E. coli* genome, we observed enhanced repression of LacZ compared to co-expression of two sgRNAs (Supplementary Figure S12), albeit weaker than dgRNA alone. The increased repression observed with this dual-dgRNA suggests that tethered dCas9 may influence gene repression by consistently interrogating the nearby genome, thus forming transient DNA loops that act as a roadblock for the transcription machinery. However, this difference might be more straightforwardly elucidated by the introduction of two doubled dCas9 RNPs targeting close-by regions.

While we successfully replicated the loop-mediated repression of the LacZ gene functionally, directly observing DNA loop formation remains a significant challenge. The 3C assay offers a means to resolve DNA contact formation beyond the limits of optical microscopy, as digested genomic fragments must lie within angstrom-scale distances to be religated. However, our attempts to employ 3C to validate the DNA loop formed by the Lac repressor were unsuccessful. We were unable to differentiate the contact frequency between fragments containing the target sequence in the host strain before and after IPTG treatment (Supplementary Figures S10 and S11). The distance between the two operons targeted by the LacI repressor is approximately either 90 bp (O1–O3) or 500 bp (O1–O2), which may be too short to be distinguished from conditions where they were brought into proximity to form a loop, owing to technical noise in the assay. We anticipate that super-resolution microscopy could provide valuable in-

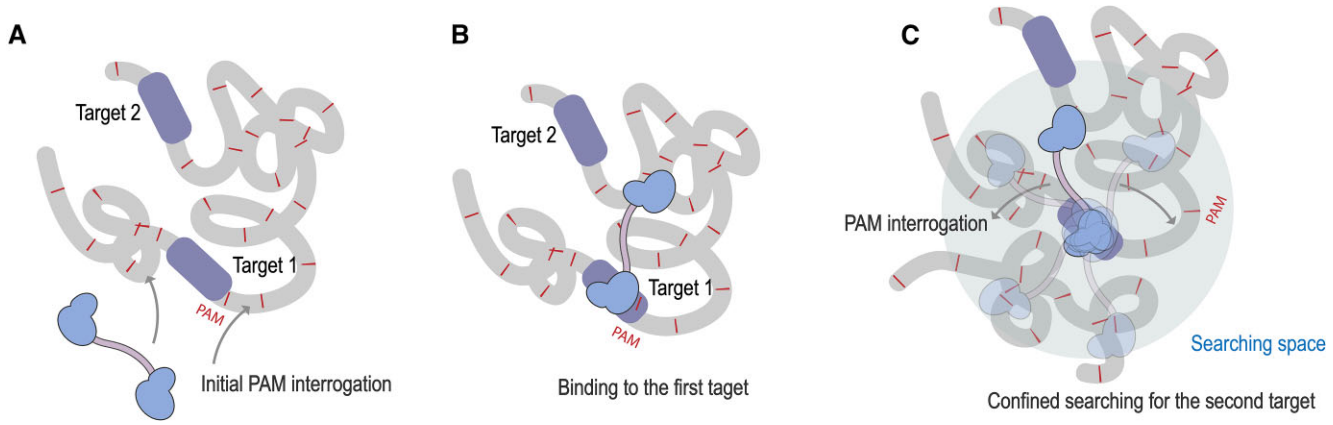


Figure 6. dgRNP target searching in the genomic context with a two-step searching model. (**A**) The dgRNP interrogates the region adjacent to the PAM sequence across the genome. (**B**) The dgRNP binds to the first target site after an initial interrogation and searching step. (**C**) The dgRNP anchored to the first target site searches for the other DNA target within a confined space with the same interrogation mechanism. The searching space is affected by the linker length and the degree of complexity in the local genomic environment.

sights into the study of DNA loops or contacts formed within shorter distances. On top of merely recapitulating the LacI-mediated DNA loop, our research here further demonstrates that dgRNPs can induce potent repression of LacZ by targeting O2 and more distant sites, up to 4700 bp (O8). Such a large distance from the transcription starting site (TSS) of O8 should make it a neglectable target for CRISPRi of the LacZ gene (34), again suggesting the role of induced DNA loop for gene regulation. And noteworthy, the effective distance of the DNA loop from the TSS remains to be investigated.

Another significant factor influencing the function of the dgRNP, aside from the guide sequence, is the design of the linker. In our studies, we observed no significant differences between samples expressing dgRNAs with various linker sizes when targeting O2 and O3, likely due to the short distance between them. However, for future applications, we envision the RNA linker serving as an adjustable handle to tune the searching space of the dgRNA, and thus the probability of loop formation. The high programmability of RNA also opens up possibilities to introduce more structured modules as linkers, which could be more stable and resistant to ribonucleases. For example, we observed that the dgRNP-RO2O8 could sustain LacZ repression for up to 8 h, although with a reduced repression rate after 6 h. This decrease in repression rate could be attributed to the degradation of the exposed RNA linker. Introducing a more stable linker could potentially sustain repression for applications requiring prolonged durations. Additionally, while our current work only explores introducing an RNA linker from end to end, using a pair of interacting domains in the two sgRNA scaffolds could provide a reversible RNA linker to induce DNA looping. This approach could induce DNA loops *via* a completely different mechanism, as the binding to two targets would be independent.

Overall, we have demonstrated the function of a bispecific gRNA in gene regulation by inducing DNA contacts *in vitro* and in *E. coli*. This method should be compatible for various organisms, as well as for *in vitro* delivery as it relies on the conventionally used dCas9 and a sgRNA construct. As it can induce DNA contacts with adjustable efficiency and searching space, it could act as an important tool for the study of the functional implications of a DNA loop in a diverse genomic context. Furthermore, the effects of the DNA loop shown in this work suggest that dgRNA–dCas9 can function as a gene regulatory tool with differentiated mechanisms, broadening the scope for synthetic biology applications.

Data availability

Data and materials are available upon request from the corresponding author.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We would like to thank Yin Lin and Giulio Bernardinelli for advice on RNA production and Sergej Masich for technical assistance.

Funding

Knut and Alice Wallenberg Foundation [KAW 2017.0114, KAW 2017.0276 to B.H.]; European Research Council (ERC) [Acronym: Cell Track GA No. 724 872 to B.H.]; Göran Gustafsson Foundation for Research in Natural Sciences and Medicine (to B.H.); Swedish Research Council [2019-01474 to B.H.]; Academy of Finland [341 908 to B.S.]. Funding for open access charge: Knut and Alice Wallenberg Foundation [KAW 2017.0114].

Conflict of interest statement

None declared.

References

1. Mojica,F.J.M., Díez-Villaseñor,C., García-Martínez,J. and Soria,E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, **60**, 174–182.
2. Bolotin,A., Quinquis,B., Sorokin,A. and Ehrlich,S.D. (2005) Clustered regularly interspaced short palindrome repeats

(CRISPRs) have spacers of extrachromosomal origin. *Microbiology (N Y)*, **151**, 2551–2561.
3. Brouns,S.J.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J.H., Snijders,A.P.L., Dickman,M.J., Makarova,K.S., Koonin,E.V. and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
4. Deltcheva,E., Chylinski,K., Sharma,C.M., Gonzales,K., Chao,Y., Pirzada,Z.A., Eckert,M.R., Vogel,J. and Charpentier,E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, **471**, 602–607.
5. Jinek,M., East,A., Cheng,A., Lin,S., Ma,E. and Doudna,J. (2013) RNA-programmed genome editing in human cells. *eLife*, **2**, e00471.
6. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A., *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
7. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided Human genome engineering via Cas9. *Science*, **339**, 823–826.
8. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable Dual-RNA–Guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
9. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P. and Lim,W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.
10. Maeder,M.L., Linder,S.J., Cascio,V.M., Fu,Y., Ho,Q.H. and Joung,J.K. (2013) CRISPR RNA–guided activation of endogenous human genes. *Nat. Methods*, **10**, 977–979.
11. Hilton,I.B., D'Ippolito,A.M., Vockley,C.M., Thakore,P.I., Crawford,G.E., Reddy,T.E. and Gersbach,C.A. (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.*, **33**, 510–517.
12. Gilbert,L.A., Larson,M.H., Morsut,L., Liu,Z., Brar,G.A., Torres,S.E., Stern-Ginossar,N., Brandman,O., Whitehead,E.H., Doudna,J.A., *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, **154**, 442–451.
13. Thakore,P.I., D'Ippolito,A.M., Song,L., Safi,A., Shivakumar,N.K., Kabadi,A.M., Reddy,T.E., Crawford,G.E. and Gersbach,C.A. (2015) Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods*, **12**, 1143–1149.
14. Konermann,S., Brigham,M.D., Trevino,A.E., Joung,J., Abudayyeh,O.O., Barcena,C., Hsu,P.D., Habib,N., Gootenberg,J.S., Nishimasu,H., *et al.* (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, **517**, 583–588.
15. Zalatan,J.G., Lee,M.E., Almeida,R., Gilbert,L.A., Whitehead,E.H., La Russa,M., Tsai,J.C., Weissman,J.S., Dueber,J.E., Qi,L.S., *et al.* (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, **160**, 339–350.
16. Cox,D.B.T., Gootenberg,J.S., Abudayyeh,O.O., Franklin,B., Kellner,M.J., Joung,J. and Zhang,F. (2017) RNA editing with CRISPR-Cas13. *Science*, **358**, 1019–1027.
17. Strecker,J., Demircioglu,F.E., Li,D., Faure,G., Wilkinson,M.E., Gootenberg,J.S., Abudayyeh,O.O., Nishimasu,H., Macrae,R.K. and Zhang,F. (2022) RNA-activated protein cleavage with a CRISPR-associated endopeptidase. *Science*, **378**, 874–881.
18. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O., *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the Human genome. *Science*, **326**, 289–293.
19. Despang,A., Schöpflin,R., Franke,M., Ali,S., Jerković,I., Paliou,C., Chan,W.-L., Timmermann,B., Wittler,L., Vingron,M., *et al.* (2019) Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.*, **51**, 1263–1271.
20. Kraft,K., Magg,A., Heinrich,V., Riemenschneider,C., Schöpflin,R., Markowski,J., Ibrahim,D.M., Acuna-Hidalgo,R., Despang,A., Andrey,G., *et al.* (2019) Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat. Cell Biol.*, **21**, 305–310.
21. Rao,S.S.P., Huang,S.-C., Glenn St Hilaire,B., Engreitz,J.M., Perez,E.M., Kieffer-Kwon,K.-R., Sanborn,A.L., Johnstone,S.E., Bascom,G.D., Bochkov,I.D., *et al.* (2017) Cohesin loss eliminates all loop domains. *Cell*, **171**, 305–320.
22. Korkmaz,G., Lopes,R., Ugalde,A.P., Nevedomskaya,E., Han,R., Myacheva,K., Zwart,W., Elkon,R. and Agami,R. (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 192–198.
23. Li,K., Liu,Y., Cao,H., Zhang,Y., Gu,Z., Liu,X., Yu,A., Kaphle,P., Dickerson,K.E., Ni,M., *et al.* (2020) Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat. Commun.*, **11**, 485.
24. Morgan,S.L., Mariano,N.C., Bermudez,A., Arruda,N.L., Wu,F., Luo,Y., Shankar,G., Jia,L., Chen,H., Hu,J.-F., *et al.* (2017) Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.*, **8**, 15993.
25. Kim,J.H., Rege,M., Valeri,J., Dunagin,M.C., Metzger,A., Titus,K.R., Gilgenast,T.G., Gong,W., Beagan,J.A., Raj,A., *et al.* (2019) LADL: light-activated dynamic looping for endogenous gene expression control. *Nat. Methods*, **16**, 633–639.
26. Hao,N., Shearwin,K.E. and Dodd,I.B. (2017) Programmable DNA looping using engineered bivalent dCas9 complexes. *Nat. Commun.*, **8**, 1628.
27. Punjani,A., Rubinstein,J.L., Fleet,D.J. and Brubaker,M.A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*, **14**, 290–296.
28. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
29. Schneider,C.A., Rasband,W.S. and Eliceiri,K.W. (2012) NIH image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.
30. Mullally,G., van Aelst,K., Naqvi,M.M., Diffin,F.M., Karvelis,T., Gasiunas,G., Siksnys,V. and Szczelkun,M.D. (2020) 5′ modifications to CRISPR–Cas9 gRNA can change the dynamics and size of R-loops and inhibit DNA cleavage. *Nucleic Acids Res.*, **48**, 6811–6823.
31. Krämer,H., Niemöller,M., Amouyal,M., Revet,B., von Wilcken-Bergmann,B. and Müller-Hill,B. (1987) lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. *EMBO J.*, **6**, 1481–1491.
32. Zhu,X., Clarke,R., Puppala,A.K., Chittori,S., Merk,A., Merrill,B.J., Simonović,M. and Subramaniam,S. (2019) Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat. Struct. Mol. Biol.*, **26**, 679–685.
33. Pacesa,M., Loeff,L., Querques,I., Muckenfuss,L.M., Sawicka,M. and Jinek,M. (2022) R-loop formation and conformational activation mechanisms of Cas9. *Nature*, **609**, 191–196.
34. Smith,J.D., Suresh,S., Schlecht,U., Wu,M., Wagih,O., Peltz,G., Davis,R.W., Steinmetz,L.M., Parts,L. and St.Onge,R.P. (2016) Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. *Genome Biol.*, **17**, 45.

Received: January 17, 2023. **Revised:** May 23, 2024. **Editorial Decision:** June 21, 2024. **Accepted:** June 25, 2024