

ARTICLE

DOI: 10.1038/s41467-017-01873-x

OPEN

# Programmable DNA looping using engineered bivalent dCas9 complexes

Nan Hao <sup>1</sup>, Keith E. Shearwin<sup>1</sup> & Ian B. Dodd<sup>1</sup>

DNA looping is a ubiquitous and critical feature of gene regulation. Although DNA looping can be efficiently detected, tools to readily manipulate DNA looping are limited. Here we develop CRISPR-based DNA looping reagents for creation of programmable DNA loops. Cleavage-defective Cas9 proteins of different specificity are linked by heterodimerization or translational fusion to create bivalent complexes able to link two separate DNA regions. After model-directed optimization, the reagents are validated using a quantitative DNA looping assay in *E. coli*. Looping efficiency is ~15% for a 4.7 kb loop, but is significantly improved by loop multiplexing with additional guides. Bivalent dCas9 complexes are also used to activate endogenous *norVW* genes by rewiring chromosomal DNA to bring distal enhancer elements to the gene promoters. Such reagents should allow manipulation of DNA looping in a variety of cell types, aiding understanding of endogenous loops and enabling creation of new regulatory connections.

<sup>1</sup>Department of Molecular and Cellular Biology, School of Biological Sciences, The University of Adelaide, Adelaide, SA 5005, Australia. Correspondence and requests for materials should be addressed to N.H. (email: [nan.hao@adelaide.edu.au](mailto:nan.hao@adelaide.edu.au))

Transcription is frequently regulated positively or negatively by the interaction of proteins bound to different sites on the same DNA molecule. These regulatory interactions can occur at long range (>1 kb), forming large DNA loops<sup>1–3</sup>. In multicellular organisms, many promoters are regulated by distal enhancer or silencer elements, sometimes separated by megabase distances<sup>4, 5</sup>. Similarly, in bacteria, the activity of  $\sigma^{54}$ -dependent promoters can be activated by bEBPs (bacterial enhancer-binding proteins) bound many kb away from the promoter<sup>2</sup>. Though poorly understood, the efficiency and specificity of these interactions appears to be controlled by other DNA looping elements. In eukaryotes, dedicated tethering elements are used to stimulate specific enhancer–promoter contacts<sup>6</sup>, while insulator elements interact with each other to sequester the enhancer and promoter into separate DNA loops<sup>7</sup>. Genome-wide chromatin capture studies have revealed a vast number of highly specific long-range DNA–DNA interactions, supporting a ubiquitous role of DNA looping in gene regulation<sup>8</sup>. However, the function of most of these DNA loops, and how these highly specific patterns of DNA looping are achieved and modulated, remains unclear.

Engineering of artificial DNA loops provides a means to manipulate endogenous DNA loops to better understand their function and to modulate gene expression, potentially with therapeutic applications, such as forming new enhancer–promoter connections to circumvent genetic deficiencies or inhibiting dysfunctional enhancer–promoter contacts. For example, fusion of the Ldb1 protein that binds to proteins at  $\beta$ -globin promoters to an engineered zinc finger DNA-binding domain able to bind to the  $\beta$ -globin locus control region (LCR) was able to redirect the LCR to activate an alternative promoter<sup>9</sup>. However, the design of zinc finger proteins able to target specific sequences can be laborious, and in the above case the looping reagent required a pre-existing protein–protein interaction. An alternative approach, insertion of binding sites for heterologous DNA-looping proteins, requires alteration of the genome sequence. Thus, a heterologous reagent that can be readily programmed to connect any two endogenous DNA segments would greatly facilitate DNA loop engineering.

Fusions of various protein domains to cleavage-defective Cas9 proteins (dCas9s) from bacterial CRISPR (clustered regularly interspaced short palindromic repeat) systems have been widely used to target specific functionalities to specific genomic DNA sequences, directed by the sequence of the associated single guide RNA (sgRNA)<sup>10</sup>. A bivalent CRISPR reagent that contains two orthogonal dCas9 moieties should thus be capable of programmed binding to two DNA segments at the same time by using two different sgRNAs. Here we design and construct such a reagent and show that it can drive de novo DNA loops in *E. coli* cells. We show that this DNA looping reagent is highly programmable and also multiplexable, readily allowing simultaneous linking of multiple targets to promote efficient DNA looping over a ~12 kb distance.

## Results

**Modeling DNA looping by a bivalent dCas9 complex.** Our first approach was to construct dCas9 heterodimers (Fig. 1a). The use of two orthogonal dCas9s allows the DNA-binding specificity of each dCas9 component of the heterodimer to be independently programmed by its specific cognate sgRNA. A homodimeric design, though simpler, has the disadvantage that half of the homodimers would be charged with sgRNAs of the same sequence specificity and would inhibit DNA looping. We chose the well-studied *Streptococcus pyogenes* (*Sp*) and *Streptococcus thermophilus* CRISPR1 (*St*) dCas9s<sup>11</sup>.

In order to optimize the efficiency of looping, we developed a simple statistical mechanical model for DNA looping by a bivalent dCas9 (Fig. 1a, b). The model indicates that the fractional looping  $F_{loop}$ , that is the fraction of time that the DNA loop is formed, is critically dependent on five parameters: (1) the dimerization of the *Sp* and *St* dCas9 monomers, given by the dimer dissociation constant  $K_{dim}$  (nM); (2) the binding affinity of each dCas9–sgRNA complex for its DNA site, given by a dissociation constant,  $K_{DNA}$  (nM); (3) the fractional saturation of each of the dCas9 proteins with its cognate sgRNA,  $\theta$ ; (4) the total concentrations of each dCas9,  $[C_{tot}]$  (nM); and (5) the DNA loopability factor  $J$  (nM), the effective relative concentration of one target DNA site relative to the other. At distances beyond 250 bp in vivo,  $J$  is primarily dependent on the length of the DNA loop and decreases as the loop gets larger, due to the entropic cost of holding two distant DNA sites together<sup>12–14</sup>.

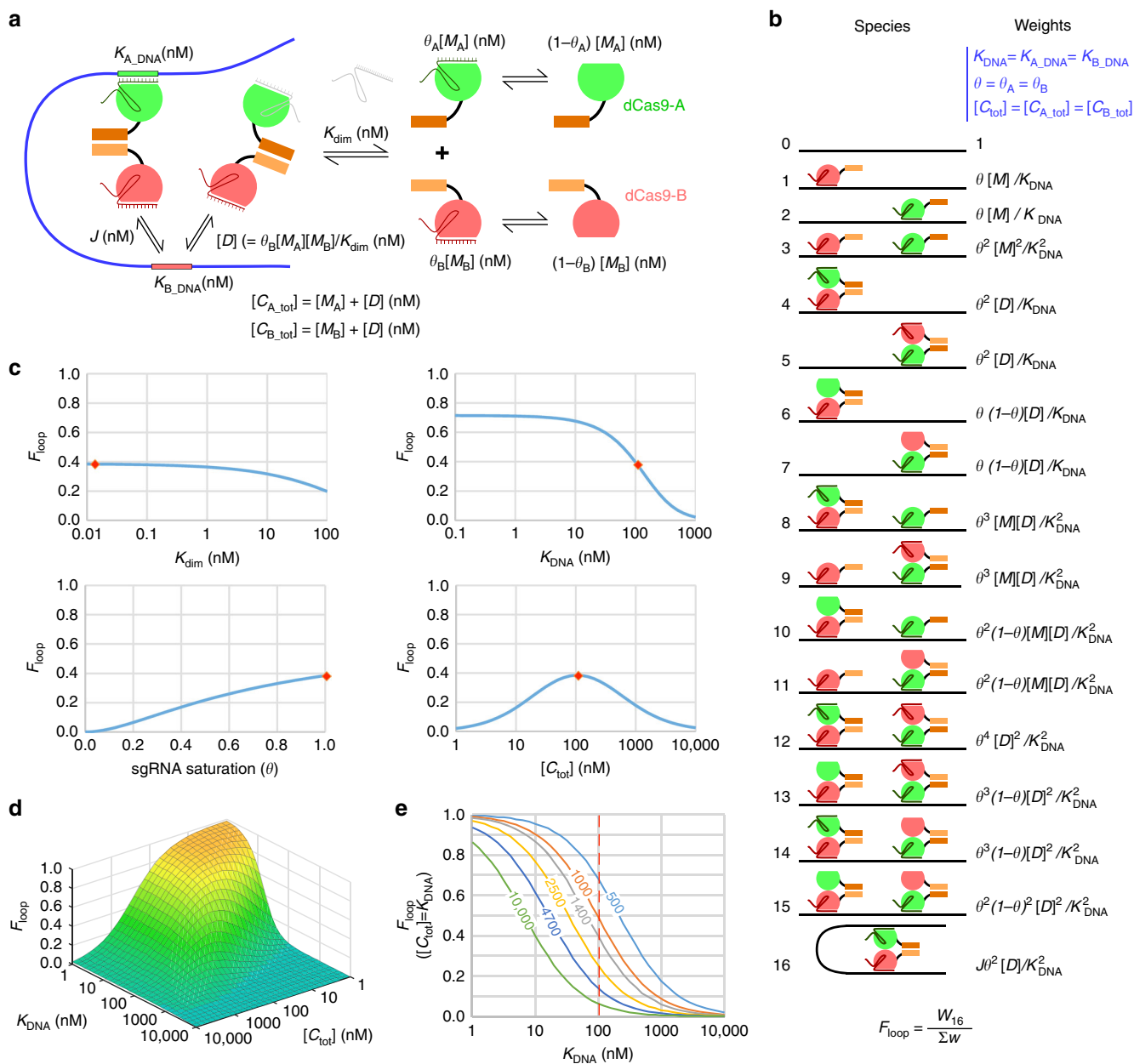
In our model, we made the simplifying assumption that  $K_{DNA}$ ,  $\theta$ , and  $[C_{tot}]$  are the same for both *Sp* and *St* dCas9s (Fig. 1b–e). Looping efficiency depends on all five parameters, such that the exact response of looping to changes in any one parameter depends on the values set for the other four parameters. Nevertheless, with the exception of dCas9 concentration, which has a clear optimum, all the other parameters show a monotonic effect on DNA looping (Fig. 1c).

As expected, DNA looping improves with stronger dimerization between dCas9 monomers, such that the majority of dCas9 complexes exist in a looping competent dimer form rather than looping-incompetent monomers (Fig. 1c). To this end, we fused strong heterodimerizing leucine zippers<sup>15</sup> to the C-termini of *Sp* and *St* dCas9 (dCas9\_Zip). These synthetic leucine zipper tags contain four pairs of electrostatically attractive salt bridges, designed to destabilize homodimer interactions and maximize inter-helical interactions. The strong dimerization constant for this pair,  $K_{dim} = 1.3 \times 10^{-2}$  nM<sup>15</sup>, should be close to optimal for DNA looping (Fig. 1c).

Since only sgRNA-bound dCas9 can bind the DNA target<sup>16, 17</sup>, looping is maximized when the dCas9s are fully saturated with sgRNA (Fig. 1c). To attempt to ensure saturation, we expressed the *Sp* and *St* dCas9\_Zip proteins from chromosomally integrated single-copy constructs, while the sgRNAs were expressed from a multicopy plasmid (*p15a ori*, 15–20 copies per cell) under the control of a strong synthetic promoter.

Our model also predicts that DNA looping increases with higher affinity interactions between the dCas9–sgRNA complex and its target DNA (i.e., low  $K_{DNA}$ ; Fig. 1c). There is some uncertainty about the  $K_{DNA}$  of *Sp* dCas9 in vivo. The *Sp* dCas9–sgRNA complex was shown to interact strongly with a perfectly matched DNA site in vitro, with a  $K_{DNA} \sim 1.3$  nM by bio-layer interferometry<sup>18</sup>. However, a substantially weaker  $K_{DNA}$  of 68.28 nM was obtained using microscale thermophoresis<sup>19</sup>. Furthermore, a  $K_{DNA}$  of 105 nM was obtained by fitting in vivo repression data<sup>20</sup>. We used this in vivo estimate for the plots in Fig. 1c–e.

The model further predicts that there exists an optimal dCas9 concentration for looping (Fig. 1c), as seen for looping by Lac repressor<sup>13, 21</sup>. If the concentration of dCas9 is too low, then its fractional occupation of the target DNA is sub-optimal for looping. Conversely, if the dCas9 concentration is too high, then both target sites will tend to be separately occupied by a dCas9 dimer (Fig. 1b, species 12–15), rather than having a single dCas9 dimer that bridges the two DNA sites. In the case where dCas9 is fully dimerized and saturated with guide RNA, looping is optimal when the dCas9 concentration  $[C_{tot}]$  is equal to its  $K_{DNA}$ . It follows that the lower the absolute values of  $K_{DNA}$  (and  $[C_{tot}]$ ) are, the greater the efficiency of looping (Fig. 1d). To keep the dCas9\_Zips in the concentration range where efficient looping is

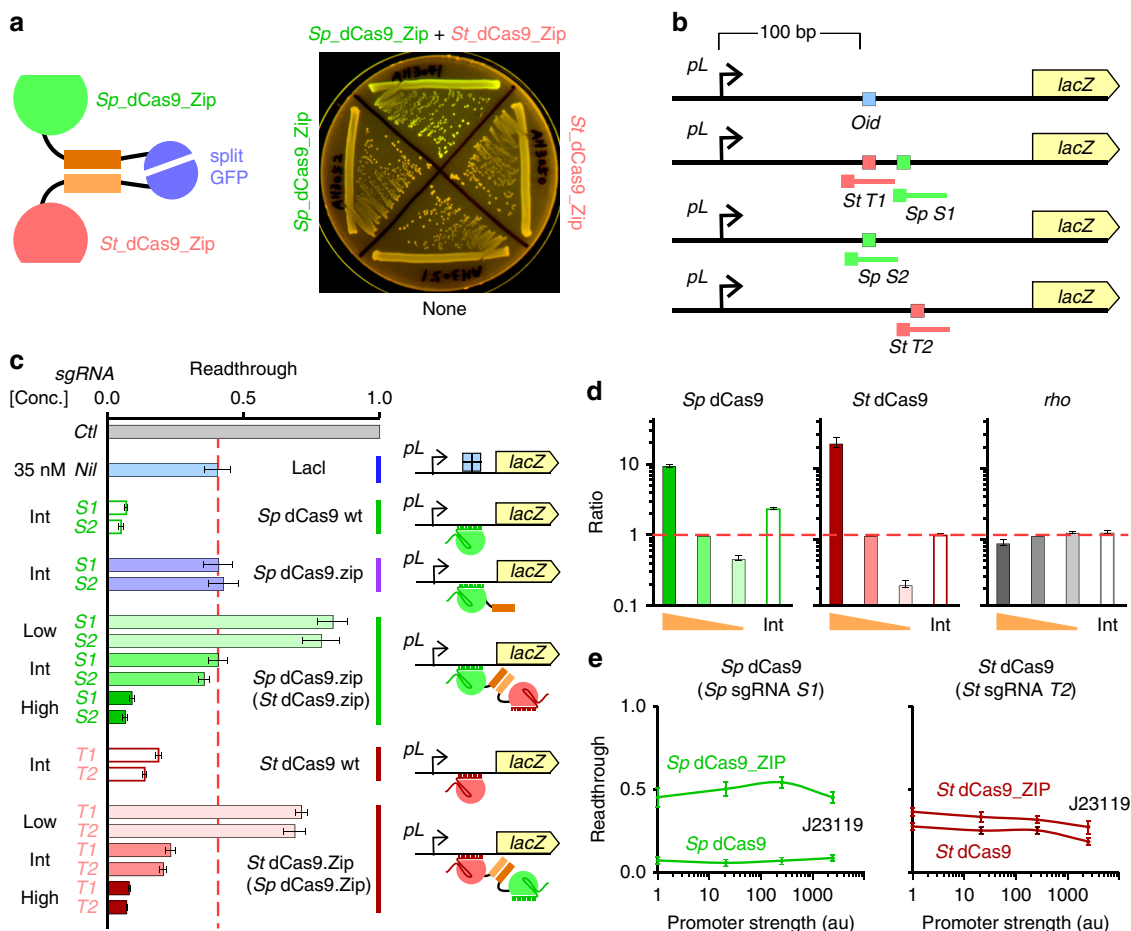


**Fig. 1** Modeling dCas9-mediated DNA looping. **a** Model parameters.  $K_{dim}$  dimerization constant,  $K_{DNA}$  dissociation constant of dCas9 to DNA,  $J$  the effective concentration of one target DNA site relative to the other,  $[M]$  concentration of dCas9 monomer,  $[D]$  concentration of dCas9 dimer,  $[C_{tot}]$  total concentration of one dCas9, and  $\theta$  the fractional saturation of dCas9 with its cognate sgRNA. Subscripts A and B denote two orthogonal dCas9s. **b** Statistical mechanical model of DNA looping by heterodimerizing dCas9s, assuming the same parameters for each dCas9. **c** The effect of varying  $K_{dim}$ ,  $K_{DNA}$ ,  $\theta$ , and  $[C_{tot}]$  on  $F_{loop}$ . In each of the plots, only one parameter is changed while the others are kept constant:  $K_{dim} = 1.3 \times 10^{-2}$  nM<sup>15</sup>,  $K_{DNA} = 105$  nM<sup>20</sup>,  $\theta = 1$ ,  $[C_{tot}] = 1.05 \times 10^2$  nM,  $J = 266$  nM at  $d = 1400$  bp separation ( $J = 1.08 \times 10^9 \times d^{-1.15}$ )<sup>13</sup>. For the case where  $[C_{A\_tot}] = [C_{B\_tot}]$ ,  $[M]$  for each dCas9 was calculated as  $(\sqrt{(1 + 4[C_{tot}]/K_{dim})} - 1) / (2/K_{dim})$  and  $[D]$  as  $[C_{tot}] - [M]$ . **d** At constant  $K_{DNA}$ ,  $\theta$ , and  $J$ ,  $F_{loop}$  increases when  $K_{DNA}$  decreases. At any given  $K_{DNA}$ ,  $F_{loop}$  is highest when  $[C_{tot}] = K_{DNA}$  (assuming complete dimerization). **e** At optimal dCas9 concentration ( $[C_{tot}] = K_{DNA}$ ),  $F_{loop}$  is inversely related to  $K_{DNA}$ . At any given  $K_{DNA}$ , the maximal  $F_{loop}$  decreases as distance separation between two DNA sites increases (i.e., decreasing  $J$ ). At  $[C_{tot}] = K_{DNA} = 1.05 \times 10^2$  nM (marked by the red line), the maximal  $F_{loop}$  is estimated to be 67, 48, 39, 24, 14, and 6% at 0.5, 0.8, 1.4, 2.5, 4.7, and 10 kb separation between the two looping sites

possible, we used three different promoters with roughly a 25:3:1 ratio in their respective promoter strengths<sup>22</sup> to produce a range of dCas9<sub>Zip</sub> concentrations.

Finally, the model shows that the maximal achievable looping efficiency is also dictated by  $J$ . The plots of Fig. 1c assume  $J = 266$  nM, which is the expected value for a 1400 bp loop based on the empirical relationship between  $J$  and loop length obtained from measurements of LacI and  $\lambda$  CI looping in *E. coli* cells<sup>13</sup>. The plots

of Fig. 1e show the expected maximal DNA looping for a range of loop lengths for bivalent dCas9 of different DNA-binding affinities. For a given  $K_{DNA}$ , DNA looping is favored with a smaller separation (higher  $J$  values). For a given loop length (a fixed  $J$ ), lower  $K_{DNA}$  (and  $[C_{tot}]$ ) values improve looping. This is because looping is a competition for binding to the free target site between a dCas9 dimer bound at the other target site, at



**Fig. 2** Characterization of the system components. **a** Split-GFP assay confirming heterodimerization of *Sp* and *St* dCas9\_Zip in vivo. **b** Schematic of the chromosomally integrated roadblocking reporter constructs (see also Supplementary Fig. 1). The relative positions of sgRNA-targeting sites are marked with lines. The PAM proximal end is boxed. **c** Roadblocking readthrough across *lacOid* with 35 nM LacI or dCas9 target sites with various concentrations of wt (unshaded) or Zip fusion (shaded) dCas9s. Both *Sp* and *St* dCas9s were expressed from chromosomally integrated expression modules under the control of low, intermediate (Int), or high strength promoters. Readthrough values are calculated as steady state LacZ units obtained in the presence of roadblock proteins (LacI or dCas9) divided by the activity in their absence. Thus, readthrough = 1 when there is no roadblocking, and readthrough = 0 when there is complete roadblocking. Data are mean  $\pm$  95% confidence intervals ( $n = 9$ ). **d** qRT-PCR showing the relative expression level of the wt (unshaded) and Zip fusion (shaded) dCas9 under the control of low, intermediate, and high strength promoters. The *E. coli gyrA* gene was used as a reference gene for data normalization and the *rho* gene was used as an internal control. Data are mean  $\pm$  standard error ( $n = 3$ ) and are expressed as a ratio relative to the intermediate expression level of dCas9\_Zip. **e** Roadblock readthrough at different sgRNA expression levels. The relative expression levels are 1:21:256:~2500, with the strongest promoter being J23119 (Biobricks, [http://parts.igem.org/Part:BBa\\_J23119](http://parts.igem.org/Part:BBa_J23119))

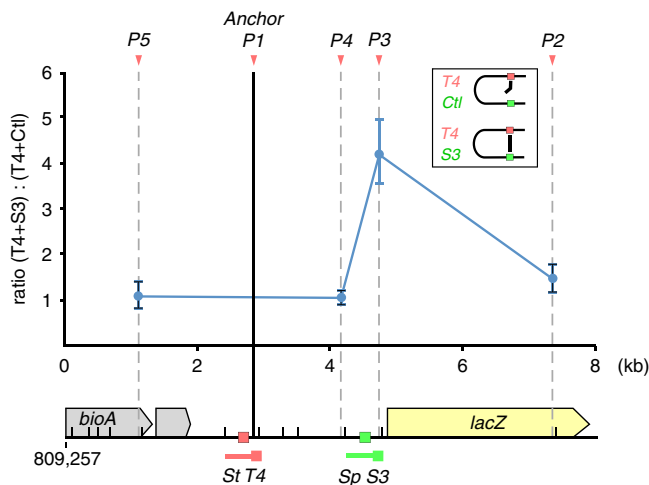
concentration  $J$ , and a dCas9 dimer in solution, at concentration  $[D]$ , with looping favored when  $J > [D]$  (Fig. 1a).

**Testing the system components.** To assess the ability of dCas9\_Zips to heterodimerize in vivo, split-GFP (green fluorescent protein) assays<sup>23</sup> were used. The *Sp* and *St*\_dCas9\_Zips were fused with the C- or N-terminal halves of GFP, respectively (Supplementary Fig. 2). Co-expression of plasmid-borne *Sp* dCas9\_Zip\_GFP\_C and *St* dCas9\_Zip\_GFP\_N, but not any of the individual fusion proteins alone, led to the reconstitution of a functional GFP protein (Fig. 2a), suggesting that the *Sp* and *St* dCas9\_Zips are capable of heterodimerization in vivo.

To confirm DNA binding of the dCas9\_Zips in vivo, transcriptional roadblocking assays<sup>24</sup> were used, where dCas9\_Zips were directed to target the 5' UTR of the *lacZ* reporter gene (Fig. 2b). The principle of this assay is that a DNA-bound complex at a transcribed region can stall elongating RNA polymerase (RNAP), leading to transcription termination and reduced expression of the reporter. Strong roadblocking by dCas9

proteins has been demonstrated in *E. coli*, particularly when the sgRNA targets the non-template strand<sup>25–27</sup>.

Both *Sp* and *St* dCas9\_Zips expressed from the intermediate strength promoter blocked *lacZ* transcription to a level comparable to or below that seen with ~35 nM of Lac repressor and its high-affinity *lacOid* operator (in vivo  $K_{DNA} \sim 0.17$  nM<sup>28</sup>) (Fig. 2c). There was no detectable difference in roadblocking by an *Sp* dCas9\_Zip protein alone or *Sp* dCas9\_Zip in the presence of its heterodimerizing partner protein *St* dCas9\_Zip. However, roadblocking by *Sp* dCas9\_Zip was ~sixfold weaker than that of wt *Sp* dCas9 when expressed from the same intermediate strength promoter, potentially due to its twofold reduced expression level (Fig. 2d). In contrast, roadblocking by *St* dCas9 and its Zip fusion (+heterodimerizing partner) was similar when expressed from the intermediate strength promoter. As expected, roadblocking by the dCas9\_Zips was relatively insensitive to the two different but fully complementary spacer sequences tested, however, roadblocking was strongly affected by the expression levels of dCas9\_Zips. At the high expression level, both *Sp* and *St* dCas9\_Zips were capable of blocking *lacZ* transcription by ~90%, while at the low



**Fig. 3** Assaying dCas9-mediated looping via chromosome conformation capture (3C). Relative contact frequencies between the anchor fragment (solid black line) and four of the BsaWI restriction digest fragments (dotted gray lines) were assessed by qPCR, and plotted as the ratio between cells expressing a pair of specific sgRNAs (T4 + S3) and cells expressing only one specific sgRNA (T4 + Ctl). The sgRNA T4 is the guide RNA that localizes dCas9\_Zip heterodimer to the anchor fragment. Data are mean  $\pm$  standard deviation ( $n = 3$ ). The DNA region of interest is shown below the plot, with the dCas9 target sites highlighted, and the BsaWI restriction sites indicated by short black bars. Orange arrowheads above the plot indicate the qPCR primer sites

expression level, *Sp* and *St* dCas9\_Zips gave only 5–15% and ~30% roadblocking, respectively (Fig. 2c).

In addition, transcriptional roadblocking assays also confirmed that both *Sp* and *St* dCas9\_Zips were saturated with their cognate sgRNAs, as roadblocking was not decreased even when the promoter expressing the sgRNAs was weakened ~2500-fold (Fig. 2e and Supplementary Fig. 2).

**The dual-dCas9 complex is capable of looping DNA in vivo.** To test whether our engineered bivalent dCas9\_Zip complex was capable of looping DNA, chromatin configuration capture (3C) assays were performed. The dCas9\_Zip heterodimers were expressed from an intermediate strength promoter, and paired with either T4 + S3 or T4 + Ctl sgRNAs (Fig. 3). In contrast to T4 + Ctl control, where the dCas9\_Zip heterodimers were only localized to the chromosomal DNA by a single site, targeting dCas9\_Zip to both T4 and S3 sites significantly increased the contact frequency between these two sites, 1.7 kb apart, demonstrating the formation of a dCas9\_Zip-mediated DNA loop between these two sites. However, while the 3C assay is commonly used to demonstrate looping between two regions of DNA<sup>9, 29</sup>, it does not give a quantitative measure of looping efficiency.

To answer this question, we turned to a second in vivo looping assay, based on our previous demonstration that a DNA loop nested inside another DNA loop can assist the outer loop's formation<sup>14, 30</sup>. Loop assistance occurs because looping between the two internal sites reduces the effective distance between the two outside sites, making them more likely to interact with each other (Fig. 4a).

We designed pairs of *Sp* and *St* sgRNAs that targeted two DNA sites about 1.4 kb apart from each other within a 2.1 kb LacI looping reporter (Fig. 4b). The formation of a LacI-mediated loop between the distal *lacOid*-binding site and the promoter-proximal *lacO2* site can be detected by its effect on

promoter repression<sup>13, 28, 30, 31</sup>. In the absence of an upstream operator, LacI binds weakly to the *lacO2* operator, leading to partial repression of the *lacZ* promoter at the low [LacI] used. In the presence of the strong distal *lacOid* operator, LacI bound at that site can interact by DNA looping to simultaneously occupy the *lacO2* site, increasing total *lacO2* occupancy and promoter repression (Supplementary Fig. 3A, B). By comparing promoter activity in the presence or absence of the distal site, it is possible to determine the absolute fraction of LacI looping  $F_{Lac}$ <sup>13</sup>. Loop assistance due to looping between the internal sites increases  $F_{Lac}$  and is revealed by increased LacI-mediated repression of the reporter<sup>14, 30</sup>.

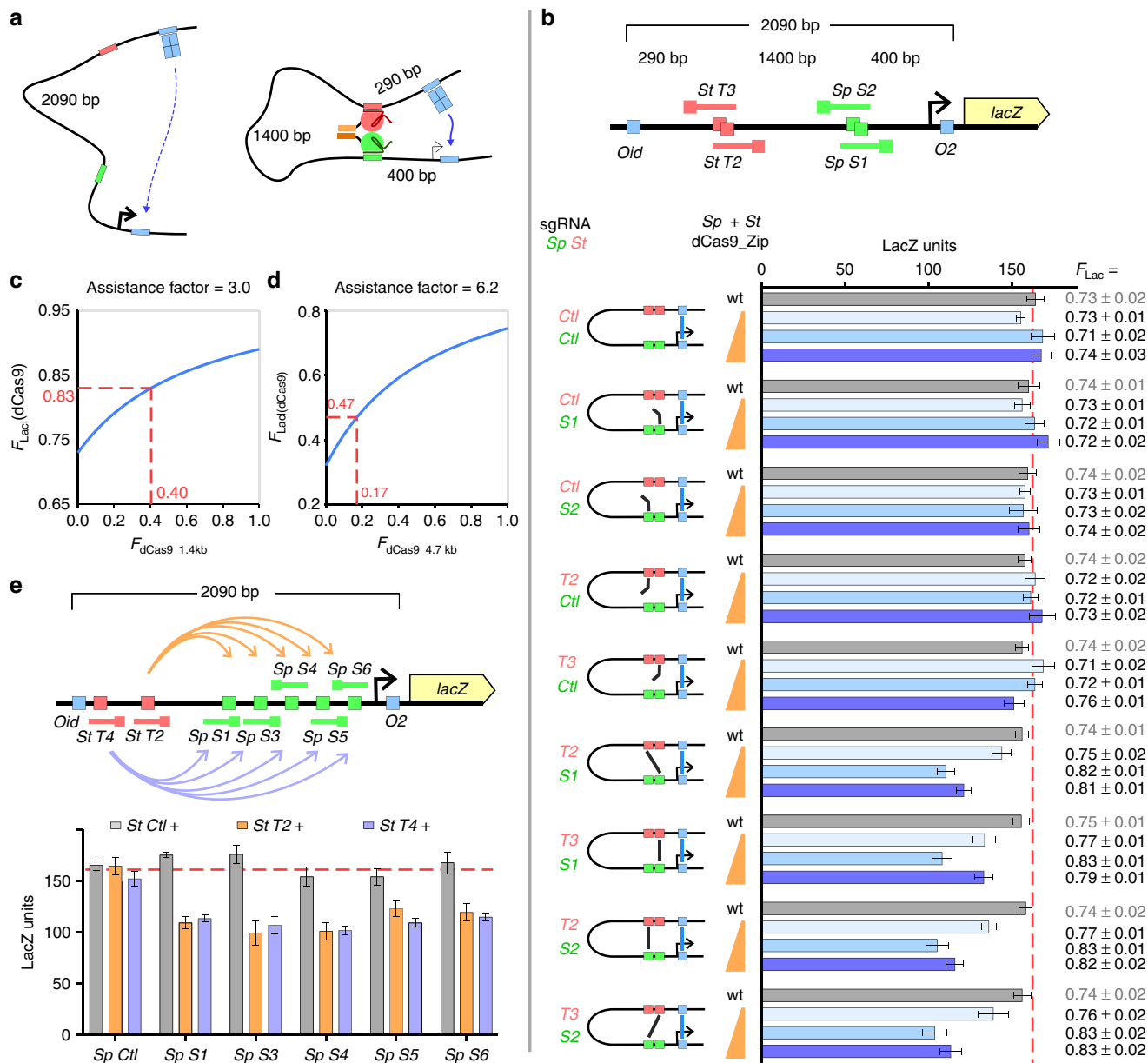
As shown in Fig. 4b, a consistent decreased expression of the reporter was observed for all couplings of the bivalent dCas9\_Zip and specific *Sp* and *St* sgRNAs targeting two internal DNA sites (T2 + S1, T3 + S1, T2 + S2, and T3 + S2). In contrast, targeted wild-type *Sp* and *St* dCas9 proteins did not affect reporter activity. Expression of the bivalent dCas9\_Zip with control sgRNAs, targeting either no internal site or any single internal site, also did not affect the reporter (Fig. 4b). Similarly, a further set of control experiments in which the reporter DNA was altered to remove all but one dCas9 target site (Supplementary Fig. 4), confirmed that two targeted DNA sites were needed for the effect. This dependence on each of the components required for dCas9-mediated looping—the targeted DNA sites, the correct sgRNAs, and the heterodimerizing dCas9-Zips—shows that the decreased reporter activity is due to dCas9 looping.

As expected from our model, the degree of loop assistance changed with different expression levels of the dCas9\_Zips. In all cases, the dCas9 loop was most readily formed with the intermediate expression level, enhancing  $F_{Lac}$  from  $73 \pm 3\%$  (95% confidence intervals,  $n = 9$ ) to  $83 \pm 2\%$ . A model for loop assistance<sup>30</sup> estimates that this improvement in LacI looping is equivalent to ~40% looping by dCas9 alone (Fig. 4c and Supplementary Fig. 3D), which corresponds well with the 40% maximal looping expected from our dimeric dCas9 model (Fig. 1e). Higher expression levels did not increase loop assistance further, despite showing a stronger roadblocking effect in the roadblock assay (Fig. 2c), suggesting that the dCas9\_Zip concentration begins to exceed its optimum (Fig. 1c), which leads to each of the looping sites being occupied by a separate dCas9\_Zip dimer and consequent loop breakage. Loop assistance at the lowest expression level of dCas9\_Zip heterodimer only slightly increased *lac* loop formation to  $77 \pm 2\%$  (Fig. 4b), consistent with a low occupation of target DNA at this concentration (Fig. 2c).

The result also showed that at all three dCas9 expression levels, similar loop assistance was observed when the two halves of a dCas9\_Zip dimer targeted either the same strand of DNA or opposite strands of the DNA (Fig. 4b, e). This flexibility increases the programmability of the system, as PAM sequences on either DNA strand can be utilized as valid target sites.

To further examine this programmability of looping, we designed two *St* sgRNAs that bind ~150 and 290 bp downstream of the distal *lacOid* operator, and paired each of them with either control *Sp* sgRNA or one of five specific *Sp* sgRNAs that target within a ~300 bp window upstream of the *placUV5* promoter (loop sizes for the dCas9 loops ranging from 1400 to 1870 bp). Using the intermediate expression level of the dCas9\_Zips, we showed that DNA looping is extremely robust, with all combinations of specific *Sp* and *St* sgRNAs consistently assisting the formation of a 2.1 kb *lac* loop from ~73 to 82% (Fig. 4e).

**DNA looping by an *Sp*-*St* dCas9 fusion.** An alternative approach to making a bivalent dCas9 is to simply fuse the two

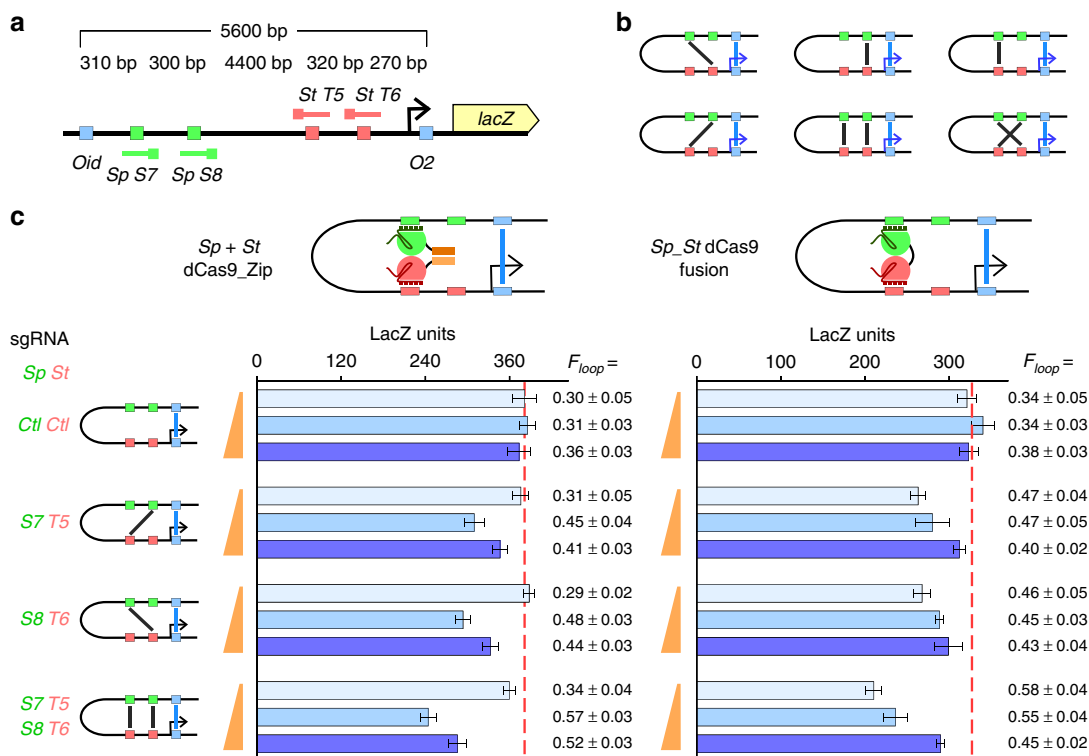


**Fig. 4** Assaying dCas9-mediated looping via loop assistance. **a** Basic principle of the loop assistance assay. **b** Formation of the CRISPR loop assists formation of the LacI loop.  $F_{loop}$  is calculated as  $[(LacZ_{OD-} - bkgd) - (LacZ_{OD+} - bkgd)] / (LacZ_{OD-} - bkgd)$  (Supplementary Fig. 3A-C). The measured average LacZ<sub>OD-</sub> values for wt as well as different concentration of *Sp/St* dCas9\_Zip fusions were  $578 \pm 17$ ,  $549 \pm 18$ ,  $541 \pm 22$ , and  $567 \pm 30$ , respectively. Data are mean  $\pm$  95% confidence intervals ( $n = 9$ ). A *bkgd* value = 16 units (unrepressible LacZ background) is estimated<sup>13</sup>. All  $F_{loop}$  values are mean  $\pm$  standard deviations ( $n = 9$ ). **c, d** The  $F$  for the CRISPR loop alone,  $F_{dCas9}$ , can be calculated from the observed LacI looping in the presence of dCas9,  $F_{LacI(dCas9)}$ , by applying a loop assistance factor equivalent to the distance shortening due to the formation of the internal loop<sup>30</sup>. **e** The formation of the CRISPR loop is relatively insensitive to the spacer sequences used. Data are mean  $\pm$  95% confidence intervals ( $n = 9$ )

dCas9 monomers together as a single polypeptide, in effect making the dimerization infinitely strong (Fig. 1c). To test DNA looping by this alternative design, we joined the two proteins by a 15 amino acid flexible linker, and expressed it from a medium copy plasmid (*CloDF13 ori*, 20–40 copies per cell) under the control of strong, intermediate, and weak promoters. The need for plasmid-based expression and the different expression promoter response compared to the dCas9\_Zips probably reflects difficulties in expressing the 2504 amino acid fusion protein. In the loop assistance assay, DNA looping by this fusion dCas9 was able to increase  $F$  for a 2.1 kb *lac* loop from  $72 \pm 3$  to  $82 \pm 2\%$  (Supplementary Fig. 5), similar to that achieved with the *Sp + St* dCas9\_Zip heterodimer. DNA looping by the fusion protein was

strongest when expressed from the weak *proI* promoter, followed by the intermediate *proA* promoter, and was weakest with the strong *proC* promoter (Supplementary Fig. 5), suggesting that the expression level from both intermediate and strong promoters exceeds the optimum for looping (Fig. 1c).

**Enhancing DNA looping via dCas9-mediated loop multiplexing.** We have demonstrated that both *Sp + St* dCas9\_Zip heterodimers and *Sp\_St* dCas9 fusions were able to form DNA loops within a 2.1 kb LacI loop to assist *lac* loop formation from ~73 to 83% (Fig. 3b and Supplementary Fig. 5). However, loop assistance by the dCas9 complexes was relatively weak, as at this small distance, DNA looping is already quite efficient with LacI



**Fig. 5** Long-range DNA looping can be enhanced via dCas9-mediated loop multiplexing. **a** Schematic representation of the 5.6 kb loop assistance reporter, with the relative positions of the sgRNA-binding sites highlighted. **b** Possible arrangements of CRISPR loops in the presence of four guide RNAs. **c** Single or multiplexed CRISPR loops formed by heterodimerizing dCas9\_Zip complexes or the *Sp\_St* dCas9 fusion can increase LacI looping efficiency. Data are mean  $\pm$  95% confidence intervals ( $n = 9$ ).  $F_{loop}$  values are mean  $\pm$  standard deviations ( $n = 9$ )

alone. Even if the dCas9 loop were to form 100% of the time, the maximal looping expected for the *lac* loop would be 89% (Fig. 4c). To test the effectiveness of our CRISPR looping reagents over a larger distance, we employed an additional chromosomally integrated *lacZ* reporter with a *lacOid* operator 5.6 kb upstream of the *lacO2* operator (Fig. 5a). Thus, looping sizes for sgRNAs S7 + T5, and S8 + T6 are both around 4.7 kb.

As expected, DNA looping by LacI alone becomes much weaker at 5.6 kb, at only  $\sim 34 \pm 5\%$ , consistent with our previous findings<sup>13</sup>. Loop assistance by a single pair of *Sp* and *St* sgRNAs combined with either *Sp + St* dCas9\_Zip heterodimer or *Sp\_St* dCas9 fusion increased LacI looping (Fig. 5c). Similar to the previous results (Fig. 4b and Supplementary Fig. 5), loop assistance was strongest for the *Sp + St* dCas9\_Zip heterodimers expressed from a single-copy intermediate strength promoter, and for the *Sp\_St* dCas9 fusions expressed by a weak promoter from a medium copy number plasmid. When expressed at their optimal concentrations, both *Sp + St* dCas9\_Zip heterodimer and *Sp\_St* dCas9 fusion improved the LacI looping from  $\sim 34$  to 47%. Applying a similar analysis to that used in Fig. 4c, we estimate that the bivalent dCas9 is alone capable of  $\sim 17\%$  looping at a 4.7 kb distance (Fig. 4d and Supplementary Fig. 3D). This is again consistent with the  $\sim 14\%$  maximal looping at this distance predicted from our dimeric dCas9 model (Fig. 1e).

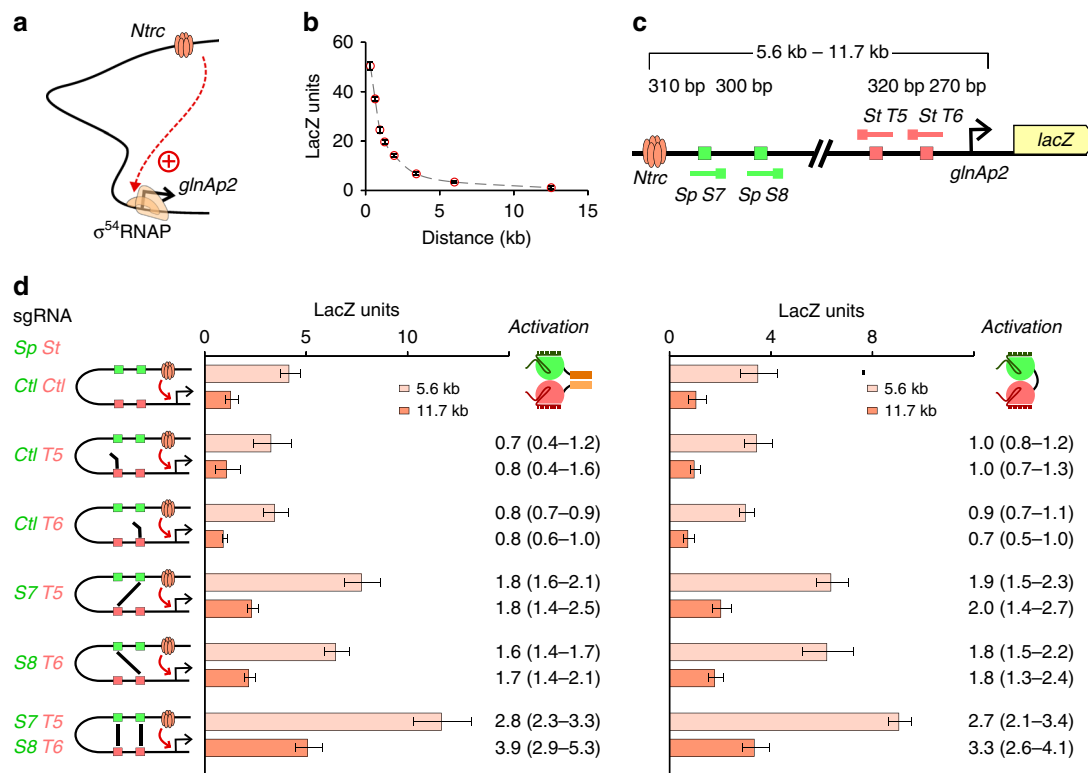
In addition, the ability of the DNA-binding specificity of each dCas9 element to be independently programmed allows for the possibility of increasing overall loop efficiency by creating multiple loops simply by expressing multiple sgRNAs. To test this multiplexability, we simultaneously expressed two sgRNAs for each dCas9, which should allow the dCas9 looping reagent to form four different single loops and two possible double loops (Fig. 5b). Each of these nested looping species should assist the formation of the outside LacI loop. Indeed, expression of all four

sgRNAs in the 5.6 kb reporter was able to increase *lac* looping  $\sim 1.7$ -fold to  $\sim 58\%$  (Fig. 5c), a substantial improvement from the single dCas9-mediated loop assistance.

### The dual-dCas9 loops improves enhancer–promoter contact.

As an independent assay for bivalent dCas9-mediated DNA looping, a classical bacterial promoter–enhancer interaction model<sup>32</sup> was adapted. Here a  $\sigma^{54}$ -dependent *glnAp2* promoter was used to drive the expression of a chromosomally integrated *lacZ* reporter gene. The activity of *glnAp2* is regulated by the NtrB–NtrC bacterial two-component signal transduction system. The NtrB is both a kinase and a phosphatase, which regulates the phosphorylation status of NtrC. Phosphorylated NtrC is a transcriptional activator that binds to enhancer elements and interacts with  $\sigma^{54}$ -RNAP bound at the promoter to catalyze its transition to the open complex<sup>33</sup> (Fig. 6a). The kinase/phosphatase activity of NtrB is normally controlled by the carbon and nitrogen status of the cell. A single mutation of NtrB (A129T) significantly reduces the phosphatase activity of NtrB without affecting its kinase activity, resulting in constitutively active NtrC<sup>34</sup>. Here we used recombineering to introduce the A129T mutation into the endogenous NtrB gene (*glnL*) and showed that the expression of *glnAp2* driving the *lacZ* reporter gene is strictly dependent on the presence of the NtrC enhancer element. As expected, as the separation between the enhancer and the promoter is increased, the effect of the enhancer diminishes (Fig. 6b). When placed 11.7 kb away from the promoter, the effect of the enhancer site is only  $\sim 2\%$  of that seen when it is placed  $\sim 300$  bp away.

However, this enhancer-mediated reporter gene activation can be significantly increased when the enhancer site is brought closer to the promoter using either the *Sp + St* dCas9\_Zip heterodimer or the *Sp\_St* dCas9 fusion (Fig. 6c, d). The effect of bivalent dCas9 is  $\sim$ twofold at 5.6 kb when a pair of specific sgRNAs were



**Fig. 6** dCas9-mediated DNA looping can be used to foster bacterial enhancer–promoter communication. **a** An NtrC hexamer bound at a distal enhancer element can interact with  $\sigma^{54}$ -RNAP bound at the promoter to activate gene transcription. **b** The activation effect of NtrC decreases as the distance between the enhancer and the promoter is increased. Data are mean  $\pm$  95% confidence intervals ( $n = 9$ ) with the background activity in the absence of the enhancer element subtracted. **c** Schematic representation of the 5.6 and 11.7 kb looping reporter used for testing dCas9-mediated promoter–enhancer activation. **d** Single or multiplexed CRISPR loops formed by heterodimerizing dCas9\_Zip complexes or the Sp\_St dCas9 fusion increase *glnAp2* transcription by looping out DNA between enhancer and promoter. Data are mean  $\pm$  95% confidence intervals ( $n = 9$ ). Numbers indicate fold activation relative to control sgRNAs, with 95% confidence limits in brackets

supplied and this effect increased to ~threefold with dual-dCas9-mediated loop multiplexing using four specific sgRNAs (Fig. 6c, d). Similar results were also obtained at a 11.7 kb separation between the enhancer and promoter sites (Fig. 6c, d). In contrast, when only one specific sgRNA is expressed (together with a control sgRNA), binding of bivalent dCas9 to the promoter–proximal site without making a distal site contact did not lead to increased *lacZ* expression, confirming that the increased *glnAp2* transcription is a result of bivalent dCas9-mediated DNA looping.

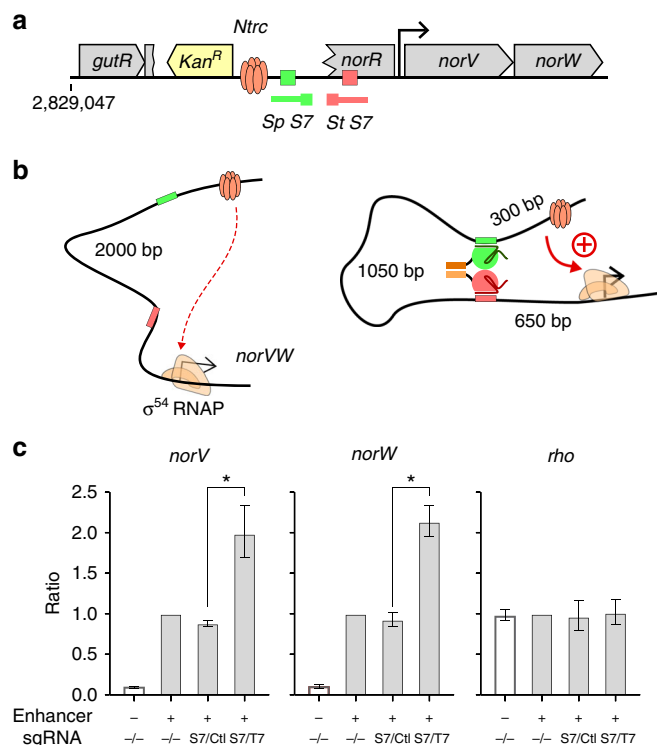
To test whether the CRISPR looping reagents could also be used to regulate a  $\sigma^{54}$ -dependent promoter at its endogenous locus, we inserted an NtrC enhancer element 2 kb upstream of the *norVW* promoter in the NtrB(A129T) strain, and designed a pair of sgRNAs that targeted two DNA sites about 1 kb apart located between the enhancer and the promoter (Fig. 7a). Again, if the bivalent dCas9 complex is capable of looping DNA, it will bring the enhancer site closer to the promoter, and should increase promoter activation (Fig. 7b). The endogenous enhancer for the *norVW* promoter is not expected to be active under the growth conditions used, and insertion of the active NtrC enhancer gave ~10-fold activation of *norVW* transcription (Fig. 4c). More importantly, the formation of the dCas9 loop between the enhancer and promoter further increased the *norV* and *norW* gene expression by ~twofold, suggesting that the bivalent dCas9-based programmable DNA looping reagent is capable of looping DNA in vivo to modulate endogenous gene expression.

## Discussion

In this study, we demonstrated that engineered bivalent dCas9 complexes, consisting of heterodimerizing dCas9s from *S. pyogenes* and *S. thermophilus* or a translational fusion of these two proteins, were capable of driving DNA loop formation in vivo. The efficiency of DNA looping was estimated at ~40 and 17% at 1.4 and 4.7 kb, respectively. DNA looping was further improved by multiplexing (Figs. 4c and 5d). For example, formation of two dCas9-mediated loops, using four sgRNAs, significantly increased a 11.7 kb enhancer–promoter interaction, leading to ~threefold activation of reporter gene expression (Fig. 6d). The improvement in DNA looping with loop multiplexing suggests that higher efficiencies and longer distance DNA looping could be achieved by programming even more target sites.

It may also be possible to improve the efficiency of formation of individual loops by modifying the bivalent dCas9 reagents. Our modeling and data suggest that looping by our dCas9 reagents is unlikely to be limited by a lack of sgRNA charging, poor dimerization, or non-optimal concentration. Rather, the fairly weak  $K_{DNA}$  value of 105 nM for dCas9–sgRNA DNA binding in vivo, obtained from an independent estimate for Sp dCas9<sup>20</sup> and consistent with our data, is likely to be limiting (Fig. 1d, e). Stronger DNA binding by the dCas9–sgRNA elements, together with lower dCas9 concentrations, should produce large improvements in looping efficiency (Fig. 1d). One potential avenue for improving binding is to explore different dCas9\_Zip fusion designs. Another approach would be to test different





**Fig. 7** dCas9-mediated chromosome rewiring of an endogenous gene. **a** Schematic illustration of the *gutRQ-norRVW* loci of the engineered *E. coli*. Recombineering was used to replace the *gutQ-norR* genes with a module containing NtrC enhancer-binding sites and a kanamycin selection cassette (yellow). Gray boxes represent endogenous genes. **b** The principle of the rewiring experiment. Bivalent dCas9-mediated loop assistance is expected to bring the NtrC sites closer to the  $\sigma^{54}$ -controlled *norVV* promoter, leading to improved activation. **c** qRT-PCR showing the relative expression levels of the *norV*, *norW*, and *rho* genes in the wild-type (unshaded) or engineered (shaded) *E. coli* strains, either in the presence or absence of specific sgRNAs. Improved activation of the *norV* and *norW* genes (but not the *rho* control gene) is seen only when both specific guides are present. The *E. coli gyrA* gene was used as a reference gene for data normalization. Data are mean  $\pm$  standard error ( $n=3$ ) and are expressed as a ratio relative to the engineered strain in the absence of sgRNAs. Asterisk indicates  $p < 0.05$  evaluated by a paired one-tailed Student's *t* test on the  $\log_{10}$  transformed data

CRISPR-Cas proteins, e.g., SaCas9, Cpf1, or C2c1<sup>35–37</sup>, that may bind DNA more strongly than the *Sp* and *St* dCas9s. A third approach could be to mutate dCas9 to increase DNA-binding affinity. Structural studies of Cas9 in complex with sgRNA and target DNA indicate that the binding energy of the complex includes both sequence-specific and non-specific contributions<sup>38, 39</sup>. Recently, two independent reports showed that the non-specific Cas9–DNA interactions can be reduced by mutating hydrophilic amino acids that stabilize the non-target DNA strand<sup>40</sup> or that contact the phosphate backbone of the target DNA strand<sup>41</sup>. Thus, it may be possible to take the reverse approach and introduce hydrophilic amino acids to strengthen non-specific DNA interactions, and thus overall DNA binding. For example, the *Sp* Cas9 S845K and L847R mutations, which decrease specificity<sup>40</sup>, may strengthen DNA binding. An increase in off-target binding may not be problematic for looping applications, since increased specificity is provided by the requirement for two DNA targets.

Given the demonstrated cross platform compatibility of CRISPR-Cas techniques<sup>42–45</sup>, we envisage that our CRISPR looping reagents will be readily transferable to other bacterial and eukaryotic systems. The programmability of the CRISPR DNA looping reagents may be readily expanded by substituting either or both the *S. pyogenes* and *S. thermophilus* dCas9\_Zips with other dCas9 proteins with different PAM specificities, as long as these proteins are orthogonal with regard to binding guide RNA and exhibit strong DNA binding. Indeed, while our manuscript was under revision, a related paper<sup>29</sup> has demonstrated the fusion of dCas9 from *S. pyogenes* and *S. aureus* with ligand-inducible heterodimerization tags from the plant phytohormone S-(+)-abscisic acid signaling pathway, and showed that dCas-mediated DNA looping can be used to re-establish  $\beta$ -globin gene expression in K562 human erythroleukemia cells by bringing the LCR to the  $\beta$ -globin promoter. Thus, dCas9-mediated looping works in both prokaryotic and eukaryotic settings.

In addition, bivalent dCas9 reagents have potential applications beyond engineering DNA loops. They could be used to bring together separate DNA molecules in order to improve DNA repair or DNA insertion—for example, after cleavage mediated by active Cas. Also, a bivalent Cas9 reagent will have a higher affinity for its DNA targets compared to a monovalent Cas9, as long as the two target sites are close enough to favor simultaneous binding, since dimerization and DNA looping generates cooperative binding. Thus, a dCas9–Cas9 bivalent reagent where the uncleaved and cleaved target sites are near to each other could provide a simple way to improve the efficiency of on-target DNA cleavage without causing increased off-target cleavage, a major issue for CRISPR-based genome editing<sup>46–48</sup>. The specificity of genomic targeting of transcription-regulating domains using dCas9 fusions<sup>10</sup> is also likely to be improved by use of bivalent dCas9–dCas9 reagents targeted to two nearby DNA sites.

## Methods

**Bacterial strains.** EC100D *mcrA*  $\Delta$ (*mrr-hsdRMS-mcrBC*)  $\phi$ 80*dlacZ* $\Delta$ M15  $\Delta$ *lacX74* *recA1endA1* *araD139*  $\Delta$ (*ara,leu*)7697 *galU galK*  $\lambda^-$  *rpsL nupG pir+* (DHFR) (Epicentre) was used for propagation of *R67K Ori* (*pir*-dependent) plasmids. Other plasmids were propagated in DH5 $\alpha$  *F^-*  $\Phi$ 80*lacZ* $\Delta$ M15  $\Delta$ (*lacZYA-argF*) U169 *recA1endA1* *hsdR17* (*rK*, *mK*) *phoA supE44*  $\lambda^-$  *thi-1 gyrA96 relA1*.

E4643 is BW30270 (MG1655 *rph+*) with the endogenous *lacIZYA* locus (EcoCyc MG1655: 360, 527–366,797) removed by recombineering using a 90bp oligonucleotide that targets the lagging strand of the replication fork, with 45bp of homology either side of the deletion site. The successful recombinant was selected as a white colony on an LB agar plate containing X-gal (20  $\mu$ g mL<sup>-1</sup>), and the sequence across the deletion junction verified<sup>49</sup>. The parental strain for roadblocking assays, 3C assays, and loop assistance assays was AH4213, a derivative of E4643, with a *LacI* expression module consisting of the native *PlacI* promoter and *lacI* gene, integrated at the primary bacteriophage 186 *attB* site (Supplementary Fig. 1B).

The parental strain for enhancer activation assays was *E. coli* strain AH5244, a derivative of E4643 containing the *glnL:A129T* mutation (also known as NRII2302) generated by scarless recombineering<sup>50</sup>. Briefly, the wild type *glnL* gene in E4643 was first disrupted by insertion of a *kanR-P<sub>rhaB</sub>-tse2* module containing a kanamycin resistance gene and a rhamnose inducible *P<sub>rhaB</sub>* promoter driving the *tse2* toxin gene, and selected on LB agar plate supplemented with 2% glucose and kanamycin (20  $\mu$ g mL<sup>-1</sup>). A second round of recombineering was then performed with a 90bp oligonucleotide that harbours *glnL:A129T* mutation, and counter-selected on 1.5% M9 minimal medium (M9MM) agar supplemented with 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, and 0.2% rhamnose. The successful recombinant strain was sequence verified.

**DNA construction.** DNA constructions used commercial DNA synthesis (Integrated DNA Technologies, GenScript), restriction enzyme-based cloning, and isothermal Gibson assembly. All sequences are available on request.

The reporter and expression constructs were made using a plasmid integration system developed from the CRIM plasmid series<sup>51</sup>. As in the CRIM system, transiently expressed phage integrase proteins were used to integrate the plasmids into phage attachment sites in the bacterial chromosome. Integration was at  $\lambda$  *attB* (EcoCyc MG1655 sequence position: 806551),  $\phi$ HK022 *attB* (position 1055419),  $\phi$ P21 *attB* (position 1196214), or the primary  $\phi$ 186 *attB* site (position: 2783828).

PCR was used to screen for correct single-copy integrants<sup>52</sup>. Insertions into the bacterial chromosome by recombineering utilized the pSIM6 plasmid<sup>53</sup>.

**Reporter constructs.** All roadblocking assay reporters (Supplementary Fig. 1) were derived from pIT3-CL<sub>λ</sub>pL<sub>lacOid</sub>lacZ\*(O<sub>2</sub>)<sup>24</sup>, an integratable plasmid carrying a  $\lambda$  phage pL promoter (−73 to +90), a *lacOid* operator (centered at +113 position), and an RNaseIII cleavage site upstream of a *lacO<sub>2</sub><sup>−</sup>lacZ* reporter gene. The *lacZ\** gene carries a weakened ribosome-binding site (nucleotides at positions +289, +302, and +304 were mutated from C, T, C to A, G, and G, respectively). To assay the ability of *Sp* and *St* dCas9 or dCas9\_Zip to bind DNA, the 20 bp *lacOid* operator was replaced by a synthetic DNA sequence containing protospacers for (1) *Sp* sgRNA *S1* and *St* sgRNA *T1*; (2) *Sp* sgRNA *S2*; and (3) *St* sgRNA *T2* on the non-template strand.

All looping reporters (Supplementary Fig. 1) used in the loop assistance experiments were as Priest et al.<sup>13</sup>, described briefly below:

The 300 bp–5.6 kb enhancer reporters and enhancer-minus controls used in the enhancer activation experiments (Supplementary Fig. 1) were derived from those of Priest et al.<sup>13</sup>, by replacing the *PlacUV5* promoter with the *E. coli glnAp2* promoter (−36 to +21), and replacing the *lacOid* operator with an enhancer module comprising the first and the second NtrC sites of *glnA* (−100 to −151).

All *lacZ* reporters were integrated into the host chromosome at the  $\lambda$  *attB* site with chloramphenicol (20  $\mu$ g mL<sup>−1</sup>) selection.

The 11.7 kb NtrC activation reporter (Supplementary Fig. 1) was generated in two steps. First, an enhancer-minus pIT-HF-CL<sub>glnAp2</sub>lacZ\*(O<sub>2</sub>)<sup>−</sup> reporter was integrated into the host chromosome at the  $\lambda$  *attB* site. Second, the NtrC enhancer was inserted into a kanamycin resistance module flanked by transcription terminators<sup>54</sup> in a pUC57 plasmid, and a derived PCR product was inserted into the gap region between the *moaA* and *ybhK* genes (EcoCyc MG1655: 816,719–816,720) by recombineering with kanamycin (20  $\mu$ g mL<sup>−1</sup>) selection.

***Sp* dCas9 and dCas9\_Zip expression constructs.** The catalytically inactive *S. pyogenes* dCas9 was PCR amplified from pdCas9-bacteria (Addgene plasmid # 44249) and assembled together with a *proA* promoter<sup>22</sup> into the pIT3-SH integration plasmid (Supplementary Fig. 1). Heterodimerizing *Sp* dCas9\_Zip was constructed by fusing an engineered 43 amino acid (a.a.) long coiled-coil peptide (EE<sub>12</sub>RR<sub>345</sub>L) containing five leucine zipper heptad repeats<sup>15</sup> at the C terminus of *Sp* dCas9, connected via an 8 a.a. GGGGSGGR linker. To vary the expression levels of *Sp* dCas9\_Zip, the *proA* promoter was replaced with either the *proC* or *proI* promoter<sup>22</sup>. Single-copy integrants of pIT3-SH<sub>proA</sub>*Sp* dCas9, pIT3-SH<sub>proC</sub>*Sp* dCas9\_Zip, pIT3-SH<sub>proA</sub>*Sp* dCas9\_Zip, and pIT3-SH<sub>proI</sub>*Sp* dCas9\_Zip at the  $\phi$ HK022 *attB* site were selected with 20  $\mu$ g mL<sup>−1</sup> spectinomycin and used to provide different concentrations of wild-type or heterodimerizing *Sp* dCas9.

***St* dCas9 and dCas9\_Zip expression constructs.** The catalytically inactive *S. thermophilus* CRISPR1 dCas9 was PCR amplified from DS-ST1casN<sup>−</sup> (Addgene plasmid # 48659) and assembled together with the *proA* promoter<sup>22</sup> into the pIT4-KT integration plasmid (Supplementary Fig. 1). Heterodimerizing *St* dCas9\_Zip was constructed by fusing an engineered 43 a.a. long coiled-coil peptide (RR<sub>12</sub>EE<sub>345</sub>L)<sup>15</sup> at the C terminus of *St* dCas9 with an 8 a.a. GGGGSGGR linker. The (EE<sub>12</sub>RR<sub>345</sub>L) and (RR<sub>12</sub>EE<sub>345</sub>L) leucine zipper pair<sup>15</sup> were designed to form very stable heterodimers, while minimizing homodimer formation.

To vary the expression levels of *St* dCas9\_Zip, the *proA* promoter was replaced with either the *proC* or *proI* promoter<sup>22</sup>. Single-copy integrants of pIT4-KT<sub>proA</sub>*St* dCas9, pIT4-KT<sub>proC</sub>*St* dCas9\_Zip, pIT4-KT<sub>proA</sub>*St* dCas9\_Zip, and pIT4-KT<sub>proI</sub>*St* dCas9\_Zip at the  $\phi$ P21 *attB* site were selected with 20  $\mu$ g mL<sup>−1</sup> kanamycin and used to provide different concentrations of wild-type or heterodimerizing *St* dCas9. The kanamycin resistance gene cassette was subsequently removed using Flp recombinase supplied from the pE-FLP plasmid<sup>52</sup>.

**Split-GFP constructs.** The p*Sp* dCas9\_Zip<sub>gfp\_C</sub> plasmid (Supplementary Fig. 2) was constructed by three-fragment Gibson isothermal assembly consisting of the *proC* promoter *Sp* dCas9\_Zip fragment amplified from pIT3-SH<sub>proC</sub>*Sp* dCas9\_Zip (see above), a gBlock fragment (IDT) carrying the C-terminal half (a.a. 158–238) of a folding reporter GFP (frGFP)<sup>23</sup> with a 6 a.a. TSGGSG linker, and a fragment carrying a *CloDF13 ori* and a spectinomycin resistance gene obtained by PCR amplification from the DS-ST1casN<sup>−</sup> plasmid.

The complementary p*St* dCas9\_Zip<sub>gfp\_C</sub> plasmid (Supplementary Fig. 2) was made by first linearizing the DS-ST1casN<sup>−</sup> plasmid with FspAI/FseI and assembling with a gBlock (IDT) containing an engineered RR<sub>12</sub>EE<sub>345</sub>L leucine zipper<sup>15</sup> and the N-terminal half (a.a. 1–157) of the frGFP joined via a 6 a.a. TSGGSG linker. This construct was then re-digested with KpnI/FseI to replace the *CloDF13 ori* and spectinomycin resistance gene cassette with a *p15a ori* and kanamycin resistance gene cassette amplified from the pUHA-1 plasmid (gift from H. Bujard, Heidelberg University, Germany).

***Sp*\_*St* dCas9 fusion expression construct.** The *Sp*\_*St* dCas9 fusion expression construct p*ProC*\_*Sp*\_*dCas9*\_*St*\_*dCas9* (Supplementary Fig. 2) was derived from p*Sp* dCas9\_Zip<sub>GFP\_C</sub> by replacing the C-terminal Zip-GFP\_C fusion with *St*

dCas9, which had been PCR amplified from DS-ST1casN<sup>−</sup>, using primers designed to join *Sp* and *St* dCas9 by a 15 a.a. (GGGS)<sub>3</sub> flexible linker<sup>55</sup>. To vary the expression levels of the *Sp*\_*St* dCas9 fusion, the *proC* promoter was replaced by either the *proA* or *proI* promoter<sup>22</sup>.

**Dual-sgRNA expression constructs.** The pDual-sgRNA expression construct (Supplementary Fig. 2) was derived from pgRNA-bacteria (Addgene plasmid # 44251). First, a module consisting of an engineered promoter (BioBrick part Bba\_J23119) driving expression of a *St* sgRNA scaffold (the St1m1 variant)<sup>11</sup>, followed by synthetic tandem T3 early and T7 early terminators<sup>54</sup>, was inserted into the AatII/EcoRI sites of pgRNA-bacteria. The inserted *St* sgRNA expression module was placed in a divergent orientation relative to a similar *Sp* sgRNA expression module and was separated from the *Sp* sgRNA expression module via strong M13 central and  $\lambda$  tR2 terminators<sup>54</sup>. Second, the high copy number *ColE1 ori* of pgRNA-bacteria plasmid was replaced by the medium copy number *p15a ori*. To vary the expression levels of the *Sp* and *St* sgRNAs, the strong J23119 promoter was replaced by J23114, J23113, or J23112 promoters, respectively. All spacer sequences used in this study are listed in Supplementary Table 1.

**Multiplex sgRNA expression constructs.** The quadruplicate sgRNA expression construct (Supplementary Fig. 2) was derived from pDual-sgRNA by first introducing a unique NotI site between *p15a ori* and the ampicillin resistance gene cassette. This pDual-sgRNA<sub>NotI</sub> plasmid was linearized at NotI, and re-circularized by using Gibson assembly to incorporate a second *Sp* sgRNA + *St* sgRNA expression module, amplified from pDual-sgRNA.

**Minimal medium LacZ assays.** Microtiter plate-based LacZ assays were carried out as previously described<sup>24</sup>. Briefly, cultures were grown in microtitre plates in M9MM supplemented with 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 0.01 mM (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub>, and 0.4% glycerol. After reaching mid-log phase, 20  $\mu$ l of culture was combined with buffer containing (per well) 30  $\mu$ l of M9MM, 150  $\mu$ l of TZ8 (100 mM Tris-HCl, pH 8.0, 1 mM MgSO<sub>4</sub>, 10 mM KCl), 40  $\mu$ l of ONPG (o-nitrophenyl- $\beta$ -D-galactoside 4 mg mL<sup>−1</sup> in TZ8), 1.9  $\mu$ l of 2-mercaptoethanol, and 0.95  $\mu$ l of polymyxin B (20 mg mL<sup>−1</sup>; Sigma). Assays were performed in triplicate from independent colonies and repeated on at least three different days ( $n = 9$ ).

**Chromosome conformation capture assays.** The 3C assays were performed as previously described<sup>56</sup> with slight modifications. Briefly, 500  $\mu$ l of OD<sub>600</sub> ~0.6 bacterial culture, grown in M9MM, was cross-linked with formaldehyde (1% final v/v) for 30 min at room temperature, followed by 30 min on ice, and then quenched with glycine (125 mM final). Cells were collected by centrifugation, washed twice with 1 mL M9MM, and lysed with 30  $\mu$ l 25 mg mL<sup>−1</sup> lysozyme (Sigma-Aldrich, L6876) at room temperature for 10 min. Ten percent SDS was subsequently added to a final concentration of 0.5% and incubated with the lysed cells for further 10 min. Five microliters of this cell lysate were then added to 47  $\mu$ l of digestion mix (5.3  $\mu$ l 10x New England Biolab (NEB) Buffer 2, 5.3  $\mu$ l 10% Triton X-100, and 35  $\mu$ l MQ water), and incubated at room temperature for further 10 min before 2.4  $\mu$ l of BsaWI (NEB, 10 units per  $\mu$ l) was added. Cell lysate was digested at 60 °C for 1 h. Following digestion, 10% SDS was added to a final concentration of 0.9% and the solution incubated at room temperature for 10 min. The whole digestion reaction was then added to a ligation mix consisting of 74.5  $\mu$ l 10% Triton X-100, 0.13  $\mu$ l 10 mg mL<sup>−1</sup> bovine serum albumin, 5.3  $\mu$ l 10 mM ATP, 76.3  $\mu$ l 10x T4 DNA ligase buffer, 603  $\mu$ l MQ water, and 1  $\mu$ l of 2000 units per  $\mu$ l NEB T4 DNA ligase. Ligation was carried out at 16 °C for 24 h. Following ligation, 0.5 M EDTA was added to a final concentration of 10 mM. Cross-links were reversed via the addition of 2.5  $\mu$ l of 20 mg mL<sup>−1</sup> Proteinase K (Sigma-Aldrich, P6556) and an overnight incubation at 16 °C. DNA was purified via two rounds of phenol-chloroform-isoamyl alcohol extraction (Sigma-Aldrich, P2069) and a single round of ethanol precipitation. The typical yield was 50 ng  $\mu$ l<sup>−1</sup> in 50  $\mu$ l of MQ water. The relative contact frequencies between the anchor fragment and various BsaWI restriction digest fragments were assessed by qPCR. A DNA fragment from the *rho* gene served as a loading control. All experiments were performed in technical triplicates for each of three biological replicates. The primer sequences can be found in Supplementary Table 2.

**RNA isolation and reverse transcription.** Bacterial cells were grown in M9MM to late exponential growth phase cultures (OD<sub>600</sub> ~0.8). An aliquot of 800  $\mu$ l of bacterial culture was then stabilized with RNAProtect Bacteria Reagent (QIAGEN) and total RNA extracted using RNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions. Traces of contaminating DNA were removed from isolated RNA using the Turbo DNA-free kit (Ambion) as per the manufacturer's protocol. RNA concentration and purity, determined by 260/280 and 260/230 ratios, were measured with a NanoDrop 2000 spectrophotometer. An aliquot of 1  $\mu$ g of DNA-free RNA was then reverse-transcribed to cDNA using QuantiTect Reverse Transcription kit (QIAGEN) according to the manufacturer's instructions. The primer sequences as well as primer efficiencies and amplicon sizes are listed below. The *E. coli gyrA* gene was used as a reference gene as its expression did not change across different test strains. All experiments were performed in technical

triplicates for each of three biological replicates. Details of primers used in qPCR experiments are given in Supplementary Table 3.

**Statistics.** For real-time qRT-PCR experiments, statistical differences were evaluated by a paired one-tailed Student's *t* test on the log<sub>10</sub> transformed data, with the level of significance set at *p* < 0.05.

**Data availability.** The data that support the findings of this study are available from the corresponding author on reasonable request.

Received: 8 April 2017 Accepted: 23 October 2017

Published online: 20 November 2017

## References

- Schleif, R. DNA looping. *Annu. Rev. Biochem.* **61**, 199–223 (1992).
- Bush, M. & Dixon, R. The role of bacterial enhancer binding proteins as specialized activators of sigma54-dependent transcription. *Microbiol. Mol. Biol. Rev.* **76**, 497–529 (2012).
- Cui, L., Murchland, I., Shearwin, K. E. & Dodd, I. B. Enhancer-like long-range transcriptional activation by  $\lambda$  CI-mediated DNA looping. *Proc. Natl Acad. Sci. USA* **110**, 2922–2927 (2013).
- Spitz, F. Gene regulation at a distance: from remote enhancers to 3D regulatory ensembles. *Semin. Cell Dev. Biol.* **57**, 57–67 (2016).
- Kadauke, S. & Blobel, G. A. Chromatin loops in gene regulation. *Biochim. Biophys. Acta* **1789**, 17–25 (2009).
- Kwon, D. et al. Enhancer-promoter communication at the Drosophila engrailed locus. *Development* **136**, 3067–3075 (2009).
- Chetverina, D., Aoki, T., Erokhin, M., Georgiev, P. & Schedl, P. Making connections: insulators organize eukaryotic chromosomes into independent cis-regulatory networks. *Bioessays* **36**, 163–172 (2014).
- Denker, A. & de Laat, W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* **30**, 1357–1382 (2016).
- Deng, W. et al. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158**, 849–860 (2014).
- Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17**, 5–15 (2016).
- Esvelt, K. M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10**, 1116–1121 (2013).
- Rippe, K. Making contacts on a nucleic acid polymer. *Trends Biochem. Sci.* **26**, 733–740 (2001).
- Priest, D. G. et al. Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and lambda repressors. *Proc. Natl Acad. Sci. USA* **111**, 349–354 (2014).
- Hao N., Sneppen K., Shearwin K. E. & Dodd I. B. Efficient chromosomal-scale DNA looping in Escherichia coli using multiple DNA-looping elements. *Nucleic Acids Res.* **45**, 5074–5085 (2017).
- Moll, J. R., Ruvinov, S. B., Pastan, I. & Vinson, C. Designed heterodimerizing leucine zippers with a range of pIs and stabilities up to 10<sup>(-15)</sup> M. *Protein Sci.* **10**, 649–655 (2001).
- Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. Structural biology. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
- Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
- Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).
- Dong et al. Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. *Nature* **546**, 436–439 (2017).
- Farasat, I. & Salis, H. M. A biophysical model of CRISPR/Cas9 activity for rational design of genome editing and gene regulation. *PLoS Comput. Biol.* **12**, e1004724 (2016).
- Johnson, S., Linden, M. & Phillips, R. Sequence dependence of transcription factor-mediated DNA looping. *Nucleic Acids Res.* **40**, 7728–7738 (2012).
- Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.* **39**, 1131–1141 (2011).
- Sarkar, M. & Magliery, T. J. Re-engineering a split-GFP reassembly screen to examine RING-domain interactions between BARD1 and BRCA1 mutants observed in cancer patients. *Mol. Biosyst.* **4**, 599–605 (2008).
- Hao, N. et al. Road rules for traffic on DNA-systematic analysis of transcriptional roadblocking in vivo. *Nucleic Acids Res.* **42**, 8861–8872 (2014).
- Bikard, D. et al. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
- Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
- Hao, N., Palmer, A. C., Ahlgren-Berg, A., Shearwin, K. E. & Dodd, I. B. The role of repressor kinetics in relief of transcriptional interference between convergent promoters. *Nucleic Acids Res.* **44**, 6625–6638 (2016).
- Garcia, H. G. & Phillips, R. Quantitative dissection of the simple repression input-output function. *Proc. Natl Acad. Sci. USA* **108**, 12173–12178 (2011).
- Morgan, S. L. et al. Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.* **8**, 15993 (2017).
- Priest, D. G. et al. Quantitation of interactions between two DNA loops demonstrates loop domain insulation in *E. coli* cells. *Proc. Natl Acad. Sci. USA* **111**, E4449–E4457 (2014).
- Muller, J., Oehler, S. & Muller-Hill, B. Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J. Mol. Biol.* **257**, 21–29 (1996).
- Wedel, A., Weiss, D. S., Popham, D., Droge, P. & Kustu, S. A bacterial enhancer functions to tether a transcriptional activator near a promoter. *Science* **248**, 486–490 (1999).
- Popham, D. L., Szeto, D., Keener, J. & Kustu, S. Function of a bacterial activator protein that binds to transcriptional enhancers. *Science* **243**, 629–635 (1989).
- Ninfa, A. J. & Magasanik, B. Covalent modification of the glnG product, NRI, by the glnL product, NRII, regulates the transcription of the glnALG operon in Escherichia coli. *Proc. Natl Acad. Sci. USA* **83**, 5909–5913 (1986).
- Kleinstiver, B. P. et al. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298 (2015).
- Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
- Shmakov, S. et al. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell* **60**, 385–397 (2015).
- Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
- Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
- Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
- Kleinstiver, B. P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
- Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
- Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
- Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355 (2014).
- Wright, A. V., Nunez, J. K. & Doudna, J. A. Biology and applications of CRISPR systems: harnessing Nature's toolbox for genome engineering. *Cell* **164**, 29–44 (2016).
- Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
- Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
- Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
- Cui, L., Murchland, I., Shearwin, K. E. & Dodd, I. B. Enhancer-like long-range transcriptional activation by lambda CI-mediated DNA looping. *Proc. Natl Acad. Sci. USA* **110**, 2922–2927 (2013).
- Khetrpal, V. et al. A set of powerful negative selection systems for unmodified Enterobacteriaceae. *Nucleic Acids Res.* **43**, e83 (2015).
- Haldimann, A. & Wanner, B. L. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.* **183**, 6384–6393 (2001).
- St-Pierre, F. et al. One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.* **2**, 537–541 (2013).
- Datta, S., Costantino, N. & Court, D. L. A set of recombinering plasmids for gram-negative bacteria. *Gene* **379**, 109–115 (2006).
- Cambray, G. et al. Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.* **41**, 5139–5148 (2013).
- Chen, X., Zaro, J. L. & Shen, W. C. Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev.* **65**, 1357–1369 (2013).
- Umbarger, M. A. Chromosome conformation capture assays in bacteria. *Methods* **58**, 212–220 (2012).

## Acknowledgments

We are grateful to members of the Shearwin lab, particularly David Priest, for discussions. This work was supported by the National Health and Medical Research Council via a project Grant (GNT1100653) to I.B.D., K.E.S., and N.H. and the Australian Research Council via a Discovery Early Career Researcher Award to N.H. (DE150100091) and Discovery Grant to K.E.S. (DP160101450).

## Author contributions

N.H., K.E.S., and I.B.D. designed the experiments. N.H. performed the experiments. N.H., K.E.S., and I.B.D. analyzed the experiments. I.B.D. did the modeling. N.H., K.E.S., and I.B.D. wrote the paper.

## Additional information

**Supplementary Information** accompanies this paper at doi:10.1038/s41467-017-01873-x.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017