Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and lambda repressors

David G. Priest, University of Adelaide
Lun Cui, University of Adelaide
Sandip Kumar, Emory University
David Dunlap, Emory University
Ian B. Dodd, University of Adelaide
Keith E. Shearwin, University of Adelaide

Journal Title: Proceedings of the National Academy of Sciences
Volume: Volume 111, Number 1
Publisher: National Academy of Sciences | 2014-01-07, Pages 349-354
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1073/pnas.1317817111
Permanent URL: https://pid.emory.edu/ark:/25593/s5rtq

Final published version: http://dx.doi.org/10.1073/pnas.1317817111

Accessed November 22, 2019 10:43 AM EST
Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and λ repressors

David G. Priest\textsuperscript{a}, Lun Cui\textsuperscript{a}, Sandip Kumar\textsuperscript{b}, David D. Dunlap\textsuperscript{b}, Ian B. Dodd\textsuperscript{c,1}, and Keith E. Shearwin\textsuperscript{a}

\textsuperscript{a}Discipline of Biochemistry, School of Molecular and Biomedical Science, University of Adelaide, Adelaide, SA 5005, Australia; and \textsuperscript{b}Department of Cell Biology, Emory University, Atlanta, GA 30322

Efficient and specific interactions between proteins bound to the same DNA molecule can be dependent on the length of the DNA tether that connects them. Measurement of the strength of this DNA tethering effect has been largely confined to short separations between sites, and it is not clear how it contributes to long-range DNA looping interactions, such as occur over separations of tens to hundreds of kilobase pairs in vivo. Here, gene regulation experiments using the Lac and \( \lambda \) CI repressors, combined with mathematical modeling, were used to quantitate DNA tethering inside \textit{Escherichia coli} cells over the 250- to 10,000-bp range. Although Lac and CI loop DNA in distinct ways, measurements of the tethering effect were very similar for both proteins. Tethering strength decreased with increasing separation, but even at 5- to 10-kb distances, was able to increase contact probability 10- to 20-fold and drive efficient looping. Tethering in vitro with the Lac repressor was measured for the same 600- to 3,200-bp DNAs using tethered particle motion, a single molecule technique, and was 5- to 45-fold weaker than in vivo over this range. Thus, the enhancement of looping seen previously in vivo at separations below 500 bp extends to large separations, underlining the need to understand how in vivo factors aid DNA looping. Our analysis also suggests how efficient and specific looping could be achieved over very long DNA separations, such as what occurs between enhancers and promoters in eukaryotic cells.

Interactions between proteins bound to separate sites on the same DNA molecule are critical in gene regulation and other DNA processes (1-4). The DNA separation between functionally interacting sites ranges from a few base pairs to hundreds of kilobase pairs, as in the case of some eukaryotic enhancers and their promoters. At short separations, it is clear that the DNA acts as a tether that keeps one site in the vicinity of the other so that the proteins at one site can find the other site in 3D space more efficiently than if they were free in solution (Fig. 1). Tethering is also a way to provide specificity because it aids interaction with linked sites but not unlinked sites. However, as the separation between the sites increases, this tethering effect becomes weaker, and it is not understood how the DNA linkage between widely separated sites, for example, between enhancers and promoters, provides the efficiency and specificity required for proper regulation.

The effect of DNA tethering can be quantified by the factor \( j_{\text{LOOP}} \) (M), the effective molar concentration of one site on the DNA relative to the other, or as the free energy of the DNA looping reaction \( \Delta G_{\text{LOOP}} \) (Fig. 1). These parameters are interconvertible: \( \Delta G_{\text{LOOP}} = -RTj_{\text{LOOP}} \) (kcal/mol; the reference \( j = 1 \text{ M} \)). The formation of a naked DNA loop is in itself an energetically unfavorable reaction (\( \Delta G_{\text{LOOP}} \) is positive) under physiological conditions due to the enthalpic cost of DNA bending and twisting (particularly important for short DNA segments) and the entropic cost of restricted configurational freedom of the DNA (the major limitation for long DNA loops). Thus, protein-mediated DNA looping reactions are driven by thermodynamic linkage to favorable protein–protein and protein–DNA interactions (Fig. 1).

Nevertheless, the DNA tether can help assemble DNA-protein complexes because the effective concentration of the DNA-bound protein at the distant site, \( j_{\text{LOOP}} \), can be greater than the concentration of available free protein (Fig. 1). Despite its critical role in DNA looping interactions, there are few measurements of \( j_{\text{LOOP}} \) in vivo, and these are mostly restricted to short site separations. Many in vivo techniques used to detect DNA looping interactions can quantitate relative contact efficiencies but do not permit measurement of \( j_{\text{LOOP}} \). Chromosome conformation capture (3C) and related methods have revealed a complex network of in vivo interactions, many over megabase pair DNA distances, between DNA sites in the genomes of organisms ranging from bacteria to humans (4, 10, 11). Assays using DNA recombinases have shown, as expected from DNA polymer theory, that specific contact efficiencies decrease with increasing DNA separation between the sites over the 1- to 100-kb range (12, 13). None of these methods reveal the fraction of the time that the sites are in contact.

\( j_{\text{LOOP}} \) for naked DNA in vitro is quite low at long separations. DNA cyclization measurements show that \( j_{\text{LOOP}} \) decreases with increasing tether length beyond \( \sim 500 \text{ bp} \), falling to \( \sim 10 \text{ nM} \) at 4,000 bp (5, 14). Measurements of \( j_{\text{LOOP}} \) for protein-induced looping give a similar picture. Most studies have focused on linear tethers <500 bp, finding variable \( j_{\text{LOOP}} \) values generally below 100 nM (9). Lower \( j_{\text{LOOP}} \) values are seen at longer separations: \( \sim 18 \text{ nM} \) for Cre DNA recombination at 3,000 bp (15), whereas LacI DNA binding cooperativity was undetectable at

\textbf{Significance}

Proteins bound to DNA often interact with proteins bound elsewhere on the same DNA to regulate gene expression. The intervening DNA tethers the proteins near each other, making their interaction efficient and specific, but the importance of this tethering effect is poorly understood at large DNA separations. We quantitated tethering inside bacterial cells, using two different proteins at separations up to 10,000 bp, to show that tethering is strong enough to drive efficient interactions over these distances. The same interactions were \( \sim 10 \text{-fold} \) weaker outside cells, implying that cellular factors enhance tethering. However, tethering was lost at a DNA separation of 500,000 bp inside bacteria, indicating special mechanisms inside eukaryotic cells to provide efficient and specific interactions over such distances.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

\textsuperscript{1}To whom correspondence should be addressed. E-mail: ian.dodd@adelaide.edu.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1317817111/-/DCSupplemental.
Fig. 1. The tethering effect of DNA. The assembly of a protein-DNA complex at a promoter (or other site) can be assisted if part or all of the complex is bound at a separate site on the same DNA molecule. The formation of the complex can occur more efficiently because the DNA tether can cause the effective concentration of the distally bound proteins at the target site to be greater than their concentration in solution (dashed arrows). The formation of the complex may repress the promoter, as studied here, or may activate the promoter (e.g., when the distal site is an enhancer). The complex may be a multimer of the same protein, as shown, or may comprise different protein subunits. (A) Case where the protein-protein interactions (red) are very strong and loop formation is driven by protein-DNA interactions (blue), as for the Lac repressor. (B) Case where the protein-DNA interactions are strong and loop formation is driven by protein-protein interactions, as for the λ CI repressor.

4,000 bp (16). The only long-range tethered particle motion (TPM) study found \( j_{\text{LOOP}} \) of \( \sim 24 \text{nM for CI looping at 2,300 bp} \) (17).

Most existing estimates of \( j_{\text{LOOP}} \) in vivo have been obtained by analysis of DNA loop-dependent repression of transcription by the Lac and λ CI repressors (6–8, 18–20). In such noncatalytic systems, the degree of repression reflects the fraction of the time that the repressor is bound at the promoter, allowing the thermodynamic stability of the looping interaction to be estimated. Analyses of data for LacI looping at short DNA separations (21, 22) have given \( j_{\text{LOOP}} \) values of the order of 1 \( \mu \text{M} \) (6, 8, 19), which is much higher than would be expected based on in vitro estimates and modeling. However, at longer separations, in vivo looping by Lac repressor seems to be weak, with little effect seen at a spacing of 1,500 bp (21) and no effect at 4,300 bp (22), leading these authors to conclude that long-range looping by LacI is inefficient. In contrast, we estimated \( j_{\text{LOOP}} \) of \( \sim 850 \text{nM for λ CI DNA loops} \) of 3.8 and 2.3 kb (7, 23), and single-molecule imaging in live *Escherichia coli* cells of CI-mediated interactions over a 2.3-kb separation gave results indicating a \( j_{\text{LOOP}} \) of \( \sim 240 \text{nM} \) (20).

In theory, \( j_{\text{LOOP}} \) for DNA loops formed by LacI or λ CI should be the same at large separations, where protein conformation effects are minimal. To resolve this apparent discrepancy and to extend the estimates of \( j_{\text{LOOP}} \) for long DNA separations in vivo, we performed systematic repression-based \( j_{\text{LOOP}} \) assays over the 300-bp to 10-kb range using the LacI and λ CI repressors in *E. coli*. To test whether the in vivo enhancement of looping seen at a spacing of 1,500 bp (21) and no effect at 4,300 bp (22), leading these authors to conclude that long-range looping by LacI is inefficient. In contrast, we estimated \( j_{\text{LOOP}} \) of \( \sim 850 \text{nM for λ CI DNA loops} \) of 3.8 and 2.3 kb (7, 23), and single-molecule imaging in live *Escherichia coli* cells of CI-mediated interactions over a 2.3-kb separation gave results indicating a \( j_{\text{LOOP}} \) of \( \sim 240 \text{nM} \) (20).

In this work, we used lacZ reporters driven by the catabolite activator protein (CAP)-independent *PlacUV5* promoter controlled by a proximal lac operator (OP) at the normal O1 position (centered at +11) and an upstream distal lac operator (OD) (Fig. 2D). To characterize the system, we used three natural operators (O1, O2, and O3) at OP, and three operators (Oid, a high affinity operator; O1; and a mutant operator, O′) at OD, with a 300-bp spacing (center to center) between OD and OP. Expression of lacZ was measured in the absence of Lac repressor (L0) or at two different Lac concentrations (L1 and L2), expressed from *Plac*I-lacI or *PlacP*lacI constructs integrated at a separate chromosomal site (Fig. 2C). The results showed, as expected, that the presence of O1 or Oid at OD increased repression substantially, depending on the operator combination and the LacI concentration (Fig. 2B).

By fitting the data to a statistical-mechanical model of LacI repression (Fig. 2C), we were able to estimate a number of in vivo parameters (Fig. 2D). The model is similar to Han et al. (9), with the addition of RNA polymerase (RNAP) binding to *PlacUV5* (24). The cell is a collection of LacI tetramers supplied from the WT lac promoter has been estimated at 18 nM (25) (11 tetramers in a 1-fL volume) and was used for L1. The model does not permit determination of absolute binding constants, and KI was arbitrarily fixed to 1 nM. The relative dissociation constants obtained for the four operators are similar to those found by Garcia and Phillips (25) (0.22, 1, 4.5, and 353 for Kid, KI, K2, and K3, respectively).

Effect of DNA Sequences on LacI Looping. The upstream operator aided repression most strongly at the shortest spacing (242 bp). Although its effect weakened with increasing spacer length, there remained a significant effect of the upstream operator at 5,600 bp at the L1 concentration (Fig. 2A). The operator placed 500 bp away had no effect on repression. The reporter data were fitted using the model and the parameters from Fig. 2D but allowing \( \beta_{\text{LacI/Oid-L1}} \) to vary for each operator spacing (SI Materials and Methods). These \( j_{\text{LOOP}}/L1 \) values (Fig. 3B and Fig. S5) show that the effective concentration of the operators relative to each other ranges from 180-fold the background LacZ units from the *PlacUV5::lacZ* reporter (Fig. 2D).

The fitting gives an estimate of \( j_{\text{LOOP}}/L1 \) that is independent of the fixed values chosen for KI or L1. Assuming L1 = 18 nM gives \( j_{\text{LOOP}} = 1,400 \text{nM for this 300-bp spacing} \), similar to estimates of \( \sim 1 \mu \text{M} \) for short lac operator spacings in vivo (6, 8, 19).

Effect of DNA Sequencing on LacI Looping in Vivo. To measure \( j_{\text{LOOP}} \) over longer distances, we used Oid-O2 reporters with operator spacings ranging from 242 to 5,600 bp. The sequence of spacer DNA can affect the efficiency of short LacI looping (24). The cell is a collection of LacI tetramers supplied from the WT lac promoter has been estimated at 18 nM (25) (11 tetramers in a 1-fL volume) and was used for L1. The model does not permit determination of absolute binding constants, and KI was arbitrarily fixed to 1 nM. The relative dissociation constants obtained for the four operators are similar to those found by Garcia and Phillips (25) (0.22, 1, 4.5, and 353 for Kid, KI, K2, and K3, respectively).

The upstream operator aided repression most strongly at the shortest spacing (242 bp). Although its effect weakened with increasing spacer length, there remained a significant effect of the upstream operator at 5,600 bp at the L1 concentration (Fig. 2A). The operator placed 500 bp away had no effect on repression. The reporter data were fitted using the model and the parameters from Fig. 2D but allowing \( \beta_{\text{LacI/Oid-L1}} \) to vary for each operator spacing (SI Materials and Methods). These \( j_{\text{LOOP}}/L1 \) values (Fig. 3B and Fig. S5) show that the effective concentration of the operators relative to each other ranges from 180-fold the L1 concentration of LacI in solution at 242 bp to 1.8-fold L1 at 5,600 bp. Thus, even at a 5,600-bp separation, a Lac repressor at the distal lac operator is seen by the promoter almost two times more frequently than all of the other Lac repressors in the cell combined. In contrast, the data for the 500-kb spacing was fitted with \( j_{\text{LOOP}}/L1 \) value of 0.09 (Fig. 3C); a Lac repressor bound 500 kb away contacts the proximal operator no more frequently than any of the other \( \sim 11 \) Lac repressors in the cell.

The \( j_{\text{LOOP}}/L1 \) values allow calculation of the fraction of looping for Oid and O2 at each separation and [LacI] (Fig. 3C). LacI looping is sensitive to concentration; at low concentrations, looping is obstructed by the formation of doubly bound species (species 6; Fig. 2C). Maximal looping decreases with increasing spacing, but even at separations of 5,600 bp, Oid-O2 are looped at least 30% of the time over a 100-fold range of [LacI] (0.2–20 nM; Fig. 3C).

Effect of Separation on Long-Range Looping by λ CI Repressor in Vivo. We used λ CI to obtain an independent measurement of the relationship between \( j_{\text{LOOP}} \) and DNA separation in vivo. DNA looping by λ CI can be detected by repression of the λ PRM.
At low concentrations, CI dimers form tetraters on OL1.OL2 and OR1.OR2, activating PRM, and these tetraters can interact to form an octamer-bridged DNA loop (Fig. 4A; free energy $\Delta G_{\text{loop}}$). The loop allows a dimer bound at $\text{OL} \rightarrow$ to help a dimer bind to the weak OR3, forming a trans-tetramer and repressing PRM (Fig. 4A; dimer-dimer interaction free energy $\Delta G_{\text{tet}}$). Thus, repression at moderate CI concentrations requires the distal OL site. With a statistical-mechanical model for CI regulation (7, 23), the free energy for the DNA looping reaction in vivo, $\Delta G_{\text{loop}}$, can be extracted from a comparison of PRM activity $\pm \text{OL}$ at high CI concentrations.

The looping reaction comprises the unfavorable DNA looping energy $\Delta G_{\text{loop}}$ and a favorable protein–protein interaction $\Delta G_{\text{PP}}$ between DNA-bound CI tetraters: $\Delta G_{\text{loop}} = \Delta G_{\text{tet}} + \Delta G_{\text{PP}}$. Estimating $\Delta G_{\text{PP}}$ from a measure of the free energy of CI octamerization in vitro ($-9.1 \text{ kcal/mol}$) (28, 29), it is possible to derive $\Delta G_{\text{loop}}$ from $\Delta G_{\text{tet}}$ (7) (SI Materials and Methods).

We made a series of OR.PRM.lacZ reporters with OR (or an OR sequence) at distances from 150 to 10,000 bp upstream of OR (Fig. 4A, Figs. S1 and S2, and SI Materials and Methods). A chromosomally integrated cl gene (or vector only) provided some 10-fold lower than before (7, 23) (Fig. 4C). At LacI concentrations at which $\Delta G_{\text{loop}}$ is fully occupied, the CI values are also re-

The reporter data were fitted by allowing a different $\Delta G_{\text{tet}}$ for each spacing but holding all other parameters fixed to the values obtained in our previous detailed analysis of OL–OR looping at 2.300-bp spacing (23) (SI Materials and Methods and Fig. S6). However, we could not achieve good fits to the data for the shorter spacers unless we decreased $\Delta G_{\text{tet}}$ to $-3.4 \text{ kcal/mol}$, below our previous estimate of $-2.4 \text{ kcal/mol}$. The 2.300-bp data (23) are still reasonably well fitted using this revised value for $\Delta G_{\text{tet}}$ and this value is in better agreement with the estimate of $\Delta G_{\text{tet}} = -3.2 \text{ kcal/mol}$ obtained by single cell imaging of looping at 2,300 bp by CI in vivo (20).

The resulting $\Delta G_{\text{loop}}$ estimates (Fig. 4C) have large errors, primarily reflecting uncertainty in the value for $\Delta G_{\text{tet}}$, which has the effect of shifting the whole $\Delta G_{\text{loop}}$ vs. separation curve up or down. The lower $\Delta G_{\text{tet}}$ value causes our $\Delta G_{\text{loop}}$ estimates to be some 10-fold lower than before (7, 23) (Fig. 4C). However, our new 2,300-bp $\Delta G_{\text{tet}}$ estimate agrees well with that derived from single cell imaging (20) (Fig. 4C). The $\lambda$ CI values are also remarkably similar to those obtained for LacI with $L1 = 18 \text{ nM}$ (Fig. 4C). We also used our Lac model to analyze existing LacI looping data for 60- to 1,500-bp separations (21) and found a good match with the $\Delta G_{\text{loop}}$ values from our LacI and Cl data (Fig. S8).

Long-Range Looping by lac Repressor in Vitro. Previous studies indicate that LacI looping with short DNA tethers (<500 bp) is less efficient in vitro than in vivo. To test whether this difference holds for longer DNA tethers, we examined LacI looping in single molecules by TPM (9, 30–32), where looping of a DNA molecule attaching a bead to a surface can be detected as a restriction in the Brownian motion of the bead (Fig. 5A). The fragments were the same as the 600-, 900-, 1,200-, and 3,200-bp spacing in vivo constructs, but with OI at OP (Fig. S4, Fig. S5, and SI Materials and Methods).

Looped and unlooped states were followed over time over a range of LacI concentrations. Excursion values from the motion records of all selected beads in one experimental condition were assembled in a histogram from which average looping probabilities were determined (Fig. 5B). As expected, the probability of looping goes through a maximum with respect to [LacI] (Fig. 5C). At LacI concentrations at which OI is fully occupied, the decrease in looping with increasing [LacI] allows estimation of $\Delta G_{\text{loop}}$ independently of the affinity of OI (Fig. 5D), because looping or unlooping is a simple binding competition between bound or free LacI (9).

The 600-bp $\Delta G_{\text{loop}}$ value obtained by TPM is 5.7-fold lower than the LacI and Cl in vivo estimates (Fig. 5E), consistent with previous studies. Interestingly, this difference increases at longer separations: 11-fold at 1,200 bp and 45-fold at 3,200 bp (Fig. 4D). The TPM measurement of $G_{\text{tet}} = 1.7 \text{ kcal/mol}$ for CI looping at 2,300 bp (17) gives a $\Delta G_{\text{loop}}$ value that is also approximately sevenfold lower than the in vivo Cl and LacI estimates (Fig. 4D).

Discussion
DNA Can Foster Long-Range Interactions Even at Large Site Separations. We used two well-characterized bacterial systems where transcription is regulated by DNA looping to obtain measurements of $\Delta G_{\text{loop}}$ in vivo for DNA tether lengths beyond 5 kb. The results for LacI

Fig. 2. Using regulation by LacI looping to measure $\Delta G_{\text{loop}}$ in vivo. (A) Chromosomal PlacUV5.lacZ reporters with a lac operator at the promoter (O$_1$) and 300 bp upstream (O$_2$), with Lac supplied from a chromosomal lac gene with plact$^-$ (low [LacI] = $L1$) or plact$^+$ (high [LacI] = $L2$) (SI Materials and Methods and Figs. S1–S3). (B) Data and model fits for the four O$_1$–O$_2$ combinations tested, assuming $L1 = 18 \text{ nM}$. The numbers on the plots are data (model) in the absence of LacI. Data errors are 95% confidence limits; $n \geq 8$. (C) Statistical mechanical model of regulation by LacI tetramers. $L$ is the concentration of LacI tetramers, $K_s$ are dissociation constants, and $R$ is a dimensionless constant describing RNAP occupation of the promoter. (D) Parameters obtained from data fitting. $K_1$, $K_2$, and $K_3$ are dissociation constants for OI, O1, O2, and O3, respectively. $\text{Zmax}$ is the LacZ activity obtained if the promoter were to be fully occupied by RNAP (average shown; SI Materials and Methods). $\text{bgk}$ is the LacZ units obtained if the promoter is completely repressed. Errors are SDs from 100 data fittings (SI Materials and Methods).
and \( \lambda \) CI were in remarkable agreement, especially considering the differences in looping mechanism, with formation of a LacI loop primarily a DNA–protein interaction (Figs. 4 and 2C) and formation of the CI loop dominated by a protein–protein interaction (Figs. 1B and 4A). The different modeling approaches and the distinct additional parameters needed to extract information of the CI loop dominated by a protein–protein interaction (Figs. 1B and 2C). The different modeling approaches and the distinct additional parameters needed to extract information of the CI loop dominated by a protein–protein interaction (Figs. 1B and 2C).

At separations <300 bp, \( j_{\text{LOOP}} \) was >1 \( \mu \)M and decayed with separation roughly as a power law with exponent −1.2 (Fig. S10). In the 5- to 10-kb separation range, \( j_{\text{LOOP}} \) was still ~20–40 \( \text{nM} \), substantially higher than the 1.7 \( \text{nM} \) concentration of a single molecule within a cell volume of 1 \( \text{fL} \) (1 \( \text{\mu \text{m}^3} \)). Thus, as long as the concentration of a protein in the cell is reasonably low, even a distant operator can strongly increase its effective local concentration. As a result, even at these distances, the fraction of time that the DNA sites spend looped can be considerable. For \( \text{Ol}d-OZ \) at 5,600-bp separation, the fraction looped could be almost 40% (Fig. 3C). For \( \lambda \) OL-OR at 10-kb separation, the fraction looped can be at least 50% at CI concentrations >3 \( \text{WLU} \) (Fig. 4D).

This reasonably strong tethering effect at 5–10 kb disappears at a separation of 500 kb, where the distal operator has no detectable effect on regulation by LacI or CI. Thus, even at large separations, distance matters: a DNA site 10 kb away is much closer than a site 500 kb away. A question remains whether there is a gradual decline in \( j_{\text{LOOP}} \) over the 10- to 500-kb range or whether tethering is lost at separations well short of 500 kb. Simple extrapolation of the in vivo data in Fig. 5E (Fig. S10) suggests that substantial tethering could exist at separations of ~100 kb. Previous measurements of relative contact efficiencies using sites for recombinases in bacteria have revealed a steady drop in relative contact efficiency over the 10- to 90-kb range (13). It may be possible to use our \( \lambda \) CI reporter system to quantitate \( j_{\text{LOOP}} \) at 100-kb separations, because \( j_{\text{LOOP}} \) values as low as 3 \( \text{nM} \) (\( \Delta G_{\text{act}} \) of ~3 kcal/mol) should be measurable (Fig. 4D).

### In Vivo Factors Increase \( j_{\text{LOOP}} \) at Long Separations

The in vitro \( j_{\text{LOOP}} \) values obtained for LacI looping by TPM were ~5- to 45-fold lower and appeared to decay faster with separation than the in vivo values (Fig. SE), fitting a power law with an exponent of −1.5 (Fig. S10). Our TPM estimates are roughly comparable with published in vitro values. Our \( j_{\text{LOOP}} \) = 75 \( \text{nM} \) for the 600-bp LacI loop is higher than the ~10- and ~30-nM values obtained for LacI by TPM for ~100- and ~300-bp loops, respectively, by Han et al. (9), but this is consistent with the expected increase in \( j_{\text{LOOP}} \) over these separations due to relief of enthalpy costs (5). The TPM results of Johnson et al. (26) indicate even poorer looping with short spacers, with \( j_{\text{LOOP}} \) values of 0.3–4 \( \text{nM} \) for different ~100-bp LacI loops. Our 900-bp estimate of \( j_{\text{LOOP}} \) = 54 \( \text{nM} \) is similar to the \( j_{\text{LOOP}} \) = 37 \( \text{nM} \) obtained for an 870-bp loop formed by Cre recombinase (15). At the 3,200-bp spacing, our in vitro \( j_{\text{LOOP}} \) fell to ~8 \( \text{nM} \), which is comparable to 18 \( \text{nM} \) at 3,044 bp for FLP (30) and ~10 \( \text{nM} \) for DNA cyclization at 4,000 bp (14). TPM analysis of CI looping at 2,300 bp gave \( j_{\text{LOOP}} \) = 24 \( \text{nM} \) (17), somewhat higher than our TPM values (Fig. 5E); however, this may be an overestimate because a low value of \( \Delta G_{\text{act}} \) was used.

Thus, we are confident that the in vivo—in vitro difference is real and applies over two orders of magnitude of the separation between sites (100–1,000 bp).

In vivo factors that increase flexibility or compaction of DNA such as DNA supercoiling and nonspecific DNA-binding proteins that bend or bridge DNA, such as the nucleoid protein HU, are thought to enhance short-range DNA looping (8, 33–36). It is not clear whether these factors also act at distances over which DNA bending and twisting are not limiting. However, increased DNA flexibility due to random binding of bend-inducing proteins such as nucleoid proteins or histones (35) should make the DNA more likely to wind back on itself, keeping it more compact and making it less likely that sites far apart on the DNA will be far apart in space. Unconstrained DNA supercoiling is implicated in aiding long-range looping in vivo (13). We note that our in vivo data can be reasonably well fitted by assuming that the DNA is a flexible polymer with an apparent persistence length of 23 nm (66 bp; Fig. S10). Relative FLP recombination rates over separations of 70–15,000 bp in human cells indicated a similar apparent persistence length of 27 nm (12).

Our data do not identify which factors are responsible for improved long-range looping in vivo. However, our work provides a quantitative target for the in vivo/in vitro difference—an ~10-fold effect—and provides an experimental system to measure the effects of candidate factors by addition of factors in vitro or removal of factors in vivo.

### Creating Efficient Long-Range Looping

Given a fixed \( j_{\text{LOOP}} \) between any two DNA sites, how can DNA looping be maximized? The LacI and \( \lambda \) CI proteins represent extremes of a biochemical continuum (Fig. 1). In the case of LacI, the protein–protein interaction...
interactions that connect the two DNA-binding ends of the complex are strong, and the DNA-looping multimer, the tetramer, forms at exceedingly low concentrations (37). For CI, the assembly of dimers into tetramers, and tetramers into the loop-forming octamer is relatively weak so that the DNA looping complex does not form readily in solution (29).

Efficient LacI looping requires the protein concentration to be substantially below $J_{\text{LOOP}}$ and to lie close to the $K_D$s of both DNA sites (9, 26). At very long separations, where $J_{\text{LOOP}}$ is small, efficient looping by LacI is thus likely to be limited by difficulties in achieving reliable low protein concentrations inside cells due to gene expression noise.

Fig. 4. Effect of DNA separation on $\lambda$ CI looping in vivo. (A) Approach for measuring $J_{\text{LOOP}}$ with $\lambda$ CI. Structure of the OL-spacer-OR:PRM:lacZ reporters. CI activates PRM at low concentrations and represses it at high concentrations. Repression is dependent on the presence of OL and the free energy of loop formation between OL and OR, $\Delta G_{\text{oct}}$. (B) Decreased repression of PRM with increasing OL-OR separation. OR:PRM:lacZ reporters ± OL placed various distances upstream were assayed at 3.3 WT lysogenic units (WLU) of CI and in the absence of CI (SI Materials and Methods and Fig. S7). Also shown are previous $J_{\text{LOOP}}$ estimates for in vivo CI looping (square, ref. 7; triangle, ref. 23; red diamond, ref. 20); as well as the $J_{\text{LOOP}}$ values for LacI (Fig. 3C), assuming $L_1 = 18$ nM (blue circles). (D) Predicted fraction of looping for OL and OR at different separations up to 3.3 WLU, using $J_{\text{LOOP}}$ estimates from C.

Fig. 5. Measurement of $J_{\text{LOOP}}$ by TPM. (A) The TPM setup. (B) Examples of excursion vs. time traces for single beads, showing transitions between looped and unlooped states (left) and mean excursion probability distributions (right) for tethers with lac operators separated by 600–3,200 bp at 10 nM LacI. Different LacI loop conformations (two peaks) were distinguishable for the 900-bp loop. (C) The probability of looping vs. LacI concentration for the 900-bp separation. (D) $J_{\text{LOOP}}$ for each DNA was determined from the average fraction of time the DNA is looped ($p_{\text{looped}}$) at each [LacI], using the [LacI] dependence of loop blockage at [LacI] 0.5–20 nM (Oid was filled at 0.1 nM), according to the equation $p_{\text{unlooped}}/p_{\text{looped}} = 2K_1/J_{\text{LOOP}} + 2[LacI]/J_{\text{LOOP}}$ (K1 was fixed at 1 nM). The slope of each $p_{\text{unlooped}}/p_{\text{looped}}$ vs. [LacI] plot equals 2$/J_{\text{LOOP}}$. (E) Comparison of the TPM $J_{\text{LOOP}}$ values with the in vivo values (LacI, blue; CI, brown). The red square shows a previous TPM estimate for $\lambda$ CI (17).
For CI, efficient looping requires the free energy of the protein–protein interaction, ΔG_{PPI}, to compensate for the unfavorable ΔG_{LOOP}. The interaction of two CI tetramers is estimated to provide ~9.1 kcal/mol to drive the looping reaction (28, 29). However, this octamerization reaction alone would fail to give more than 50% looping for DNA separations above ~1,000 bp, where ΔG_{LOOP} > +9.1 kcal/mol (ILOOP < 384 nM). One way to improve CI looping would be to strengthen ΔG_{PPI} by increasing the strength of the interactions between protein subunits. However, such changes would tend to increase the formation of larger complexes in solution, which would block looping.

The efficiency of looping by λ CI is increased by increasing the number of interacting proteins at each site. The additional bridge formed by the interaction of a CI dimer at each site, to form a tetramer, effectively means that ΔG_{PPI} (octamerization) and ΔG_{tet} (tetramerization) combine to counterbalance ΔG_{LOOP}. As suggested by Dröge and Müller-Hill (38), this approach of using a DNA scaffold to assemble at each site a large complex of proteins that individually interact weakly but in combination provide a strong protein–protein interaction seems optimal for creating strong looping at very long separations, such as in eukaryotic genomes. A similar method could also be used to make LacI-style looping more efficient; additional looping proteins at each site could permit a DNA loop to be bridged by two Lac tetramers.

However, strategies that use multiple interacting sites for the same protein to strengthen looping will eventually be limited by the formation of short DNA loops within each set of binding sites. A way around this limitation could be to combine different looping proteins. This kind of approach for making looping efficient has the advantage that specificity of looping could be provided by using a relatively small number of different looping proteins in different combinatorial patterns. Such a strategy may be necessary to provide both efficiency and specificity for the very long-range DNA interactions in eukaryogenic genomes.

Materials and Methods

Reporter and Expression Constructs. The lacZ reporters and Lac and CI expression constructs were chromosomally integrated using the OSIP system and its precursors (23, 29) into MG1655 rphΔlacZΔYA (SI Materials and Methods). Cells were grown at 37 °C in minimal medium (LacI looping strains) or in rich medium (CI looping strains) and assayed by modified LacZ microtiter plate method (7) (SI Materials and Methods).

Mathematical Modeling. Data fitting to extract ΔG_{loop} values from the λ CI reporter data was as previously described (23) with some modifications. Details of this and the LacZ modeling are given in SI Materials and Methods.

TPTM. TPTM experiments were conducted as described previously (30–32) with some modifications (SI Materials and Methods). Dröge and Müller-Hill (38) suggested that using a DNA scaffold for discussion, to acknowledge Suleyman Oluwole and Weng Lee Lim in the Emory Physics Department for encouragement, to provide both efficiency and specificity for the very long-range DNA interactions in eukaryogenic genomes.

Acknowledgments. We thank Laura Finzi, Iain Murchland, Julian Pietzsch, and other members of the Shearwin, Dunlap, and Finzi laboratories, as well as University Fellows for discussion, and Jun-Lin Ouyang for assistance. We acknowledge Dr. Yuwu Hu and Weng Lee Lim in the Emory Physics Department for encouragement. We thank Weng Lee Lim for developing the computationally simulated DNA loop scaffold. This work was supported by grants from the National Institutes of Health (GM 53949), the National Science Foundation (DBI 0545879), the Australian Research Council grant (DP110010824 and DP101040740), the Whilt Elliott Fellowship in Biochemistry and National Health and Medical Research Council (Australia) grant APP1025549 (to I.B.D.), National Institutes of Health grant R01GM084070 (to D.D.D.), and the Center for Pediatric Nanomedicine in the Department of Biomedical Engineering, Georgia Institute of Technology and Children’s Healthcare of Atlanta (D.D.D.).