# Sequence Dependence of Transcription Factor-Mediated DNA Looping: Supplementary Material

Stephanie Johnson, Martin Lindén and Rob Phillips

# Contents

# S1 Modeling the Looping Probability.

We sketch here a statistical mechanical framework that allows us to see how the looping probability that we measure with TPM depends upon various tunable parameters such as the strength of the repressor binding sites, the concentration of transcription factors and the length and flexibility of the intervening DNA. These theoretical developments serve as an important conceptual framework for making the TPM assay a precise measurement scheme for determining properties of the DNA and of DNA-repressor interactions. As we have found concentration titrations to be particularly useful tools for understanding these DNA-repressor interactions, and have used them extensively in the work detailed in the main text, these titration curves will serve as the primary windows through which we will view TPM experiments from a theoretical viewpoint. Some of the results of our theoretical analysis are also presented in the main text.

We first give a brief introduction to the statistical mechanical model that characterizes the TPM experiments, in more detail than in the main text, and examine how the three parameters of the model (effective binding constants and J-factor) affect the looping probability. We then extend this model to examine how the titration curves change when additional, experimentally important complications are added: inactive fractions of repressor, dimers, and low repressor concentrations relative to DNA concentration. The results of this section are summarized in Fig. S1, which illustrates how the looping titration curves are altered as a result of intentional parameter tuning and unintentional deviations from the ideal case. Throughout this section we also make reference to experimental results, presented in this Supporting Information or in the main text, that touch on these intentional and unintentional parameter changes.

## S1.1 Tuning the simple titration curve.

Here we analyze the shape of the ideal titration curve in some depth, in order to answer a variety of questions. How do we expect the shape of the looping probability as a function of concentration, $p_{\text{loop}}([R])$, to change if we replace one of the operators with a repressor binding site of a different affinity? What happens if we change the J-factor of the DNA by, for example, changing the distance between the operators (and therefore the length of the loop)? Do two identical operators produce different titration curves than a pair of very different strengths?

We begin by summarizing the simple model we derived previously in [1], presented also in the main text. In the specific case we are considering here of a TPM experiment to study looping by a protein such as the Lac repressor, the probability of the looped state can be expressed in terms of the Boltzmann weights of each of the five states available to the system: nothing bound to the DNA, one head of a repressor bound at one operator, one head of a repressor bound at the other operator, two repressors bound with one attached to each operator, or one repressor with the two heads bound to the two operators (the looped state). These states and their corresponding weights as derived in [1] are diagrammed in Fig. S2(A). For concreteness here through Section S1.5 we will label the operators $O_{id}$ and $O_1$, representing the strong, synthetic "ideal" operator and the strongest naturally occurring operator $O_1$. These operators form the example case to which all others are compared in Fig. S1 and will be used as such throughout this section, though obviously these results apply equally well to other choices of operators.

For a system that satisfies the equilibrium conditions, the looping probability is given by the statistical weight of the looped state divided by the sum of all the states in Fig. S2(A), or

$$p_{\text{loop}}([R]) = \frac{\frac{1}{2}\frac{[R]J_{\text{loop}}}{K_1 K_{\text{id}}}}{1 + \frac{[R]}{K_1} + \frac{[R]}{K_{\text{id}}} + \frac{[R]^2}{K_1 K_{\text{id}}} + \frac{1}{2}\frac{[R]J_{\text{loop}}}{K_1 K_{\text{id}}}}, \tag{S1}$$

where $[R]$ is the repressor concentration, $K_{\text{id}}$ and $K_1$ are the dissociation constants for the Lac repressor
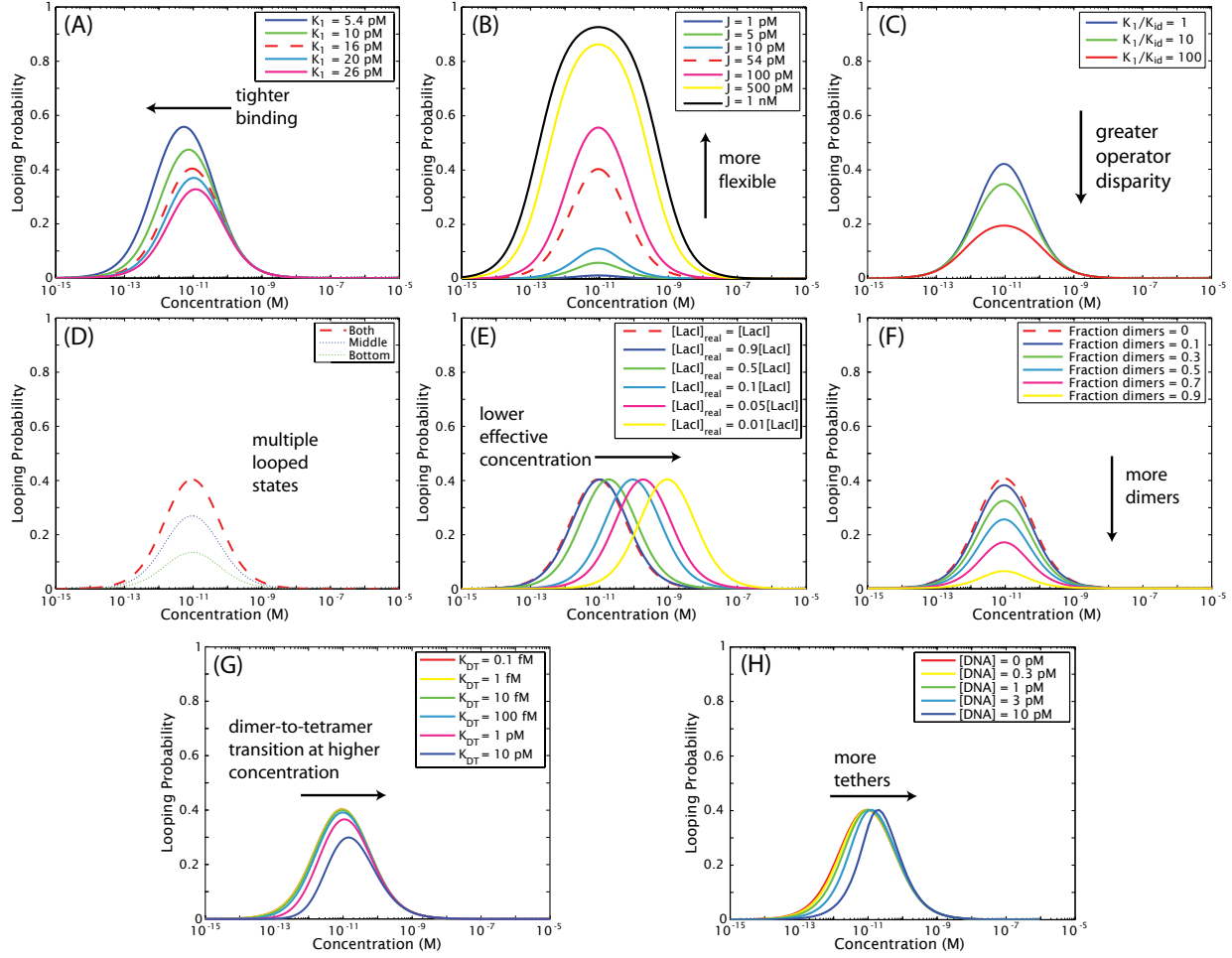
Figure S1: Effect of key parameters on the concentration dependence of the looping probability. Unless otherwise indicated, $K_{id} = 5.4$ pM, $K_1 = 16$ pM, and $J_{loop} = 54$ pM. These values for $K_{id}$ and $K_1$ are comparable to values found in the literature for two of the known binding sites for the Lac repressor [2, 3]. (Note, however, that these are not the values we measured in this work, due to different experimental conditions such as salt concentration.) Curves with these default parameters are shown as dashed red lines for comparison across panels. Panels (A), (B), and (D) are shown in the main text and are presented here for comparison. (A) Effect of changing the strength of one of the operators. (B) Effect of changing the flexibility of the DNA in the loop. (C) Effect of changing the ratio $K_1/K_{id}$ when the concentration $(K_1 K_{id})^{1/2}$ at which looping is maximal is kept the same. (D) Extension of the simple model to the case of two experimentally distinguishable looped states, which we model as having different J-factors. Here the bottom state is one-third that of the default 54 pM, and the middle is two-thirds that of the default (and the dashed red line shows the sum of the probabilities of the two states). (E) Effect of a discrepancy between the presumed concentration of repressor and the actual concentration. (F) Effect of a constant fraction of repressors that cannot dimerize and therefore cannot loop. (G) Effect of taking into account the dimer-to-tetramer dissociation at various low concentrations, for varying values of the tetramer-to-dimer dissociation constant $K_{DT}$. (H) Effect of competition for Lac repressor binding between different tethers at low repressor concentrations. The tether density [DNA] is defined as the number of tethers divided by total volume. Unlike those in (G) and (H), the curves in panels (E) and (F) are indistinguishable in TPM experiments, since the parameters are rescalings of those in the simple model.
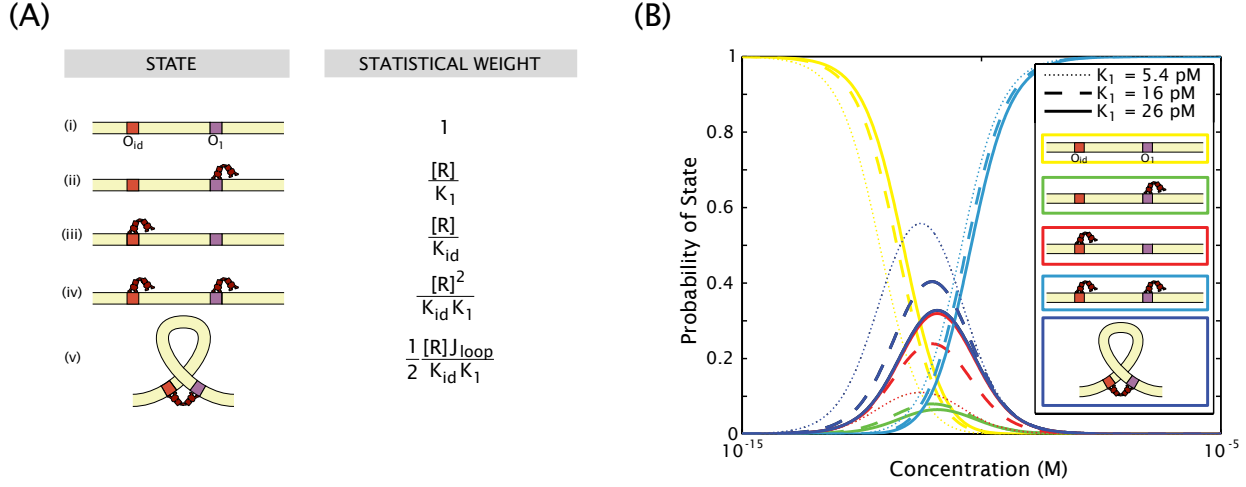
**(A)**



**(B)**



Figure S2: States and weights of the simple model. (A) Schematized states and Boltzmann weights for the simple model (first derived in [1]). $[R]$ is the repressor concentration, $K_1$ and $K_{id}$ are the dissociation constants for the repressor binding to operators $O_1$ or $O_{id}$ respectively, and $J_{loop}$ is the looping J-factor of the DNA between the operators. The looping probability is given by the weight of state (v) divided by the sum of all five states. (B) Probabilities of all the states of the system as $K_1$ is changed. This figure shows three of the curves of Fig. S1(A) but includes not only the looping probability but also the probabilities of the four other states schematized in (A) here. The different colors correspond to the five states as indicated in the legend. Dotted, dashed, or solid lines correspond to $K_1 = 5.4$ pM (in which case $K_1 = K_{id}$), 16 pM, or 26 pM respectively. For example, the solid dark blue and red curves which overlap indicate that when $K_1 = 26$ pM, there is an almost equal probability at all concentrations that the system will be in the looped configuration (dark blue) or that $K_1$ will be bound (red); whereas when $K_1 = 5.4$ pM, where the looped state is maximal (dark blue dotted line) the single-operator-bound states have very low probabilities (green and red dotted lines, which here overlap completely because in this case $K_1 = K_{id}$). These curves confirm the intuition that looping is low at low concentrations because the predominant state is the one with no repressors bound at either operator; however at high concentrations looping is also low, because the predominant state is the one in which two repressors are bound to the two operators.

4

from the $O_{id}$ and $O_1$ operators respectively, and $J_{\text{loop}}$ is the sum of the individual J-factors of all the possible DNA loop topologies. For the simplest case, the looping J-factor is related to the free energy cost of bending the DNA into a loop through

$$J_{\text{loop}} = 1 \text{ M } e^{-\beta \Delta F_{\text{loop}}}, \tag{S2}$$

where $\Delta F_{\text{loop}}$ is the free energy of forming a loop, and $\beta$ is $1/k_B T$. The units are concentration: the J-factor can be thought of as the concentration of one binding site in the vicinity of the other [4, 5]. It appears in Eq. (S1) with a factor $1/2$, which is a combinatorial factor that reflects the symmetry of the Lac protein and the binding sites [1]. Combining the different loop topologies together in a single state, as we have done here, is appropriate for the situation where they cannot be distinguished experimentally. The generalization to several distinctive looping states is discussed below.

In the main text we explore the dependence of the concentration at which looping is maximal, $[R]_{\text{max}}$, and the looping probability at this concentration, $p_{\text{loop}}([R]_{\text{max}})$, on the dissociation constants and J-factor (Eqs. (2) and (3)). One interesting feature of the titration curves not discussed in the main text is the behavior of the outer tails of the curves. As shown in Fig. S1(A), changing the operator strengths changes the behavior at low concentrations, but not at high concentrations, where curves with different operators (but identical $J$-factors) fall off in the same fashion. The behavior at high and low concentrations can be read off directly from Eq. (S1). At high concentrations, the doubly-bound state dominates, which has a weight of $[R]^2/K_1 K_{\text{id}}$; therefore in the high concentration limit,

$$\lim_{[R] \gg J_{\text{loop}}, K_1, K_{\text{id}}} p_{\text{loop}}([R]) \approx \frac{J_{\text{loop}}}{2[R]}. \tag{S3}$$

This shows that the binding constants drop out of the problem at high concentrations. In the limit of low concentrations, the state with no repressors bound dominates, which has a weight of 1; that is, in the low concentration limit,

$$\lim_{[R] \ll J_{\text{loop}}, K_1, K_{\text{id}}} p_{\text{loop}}([R]) \approx \frac{J_{\text{loop}}[R]}{2K_1 K_{\text{id}}} = \frac{J_{\text{loop}}[R]}{2[R]_{\text{max}}^2}. \tag{S4}$$

These results reflect the different states that compete with the looped states in the two limits. At high concentrations, the looped state is out-competed by the doubly occupied state, and since the weights of both states have the same dependence on operator strength, the outcome only depends on the $J$-factor. At low concentrations, on the other hand, the looped state is out-competed by the unoccupied state (with weight unity), and the outcome therefore depends on all parameters. (See also Fig. S2(B) above.)

By way of contrast, changing the $J$-factor moves both tails in a symmetric fashion, as illustrated in Fig. S1(B). This behavior is dictated by the symmetry property: since the peak position is independent of $J$, changes in $J$ have to influence the high and low concentration parts of the curve equally. Likewise, both the high-concentration and low-concentration behaviors of the curve in Eqs. (S3) and S4 depend equally on the J-factor. From an experimental perspective this makes data at concentrations near and below $[R]_{\text{max}}$ crucial for measuring dissociation constants. On the other hand, if one is only interested in the $J$-factor, data at high concentrations is sufficient.

Since the high and low concentration limits (and therefore the width of the titration curve), as well as the peak position, depend on the operator strengths only through $[R]_{\text{max}}$, we can ask how changing the relative strengths of the operators in a way that leaves $[R]_{\text{max}}$ unchanged affects the looping probability. As shown in Fig. S1(C), if we change the operators in a way that leaves $[R]_{\text{max}}$ unaffected, only the peak height changes, not the peak position or the width of the titration curve. The peak looping probability, given by

Eq. (3) in the main text, can be rearranged in a way that separates out the dependence on difference in operator strength from the other factors. Specifically, if we define $\alpha = K_1/K_{id}$, Eq. (3) can be written as

$$p_{\text{loop}}([R]_{\max}) = \frac{\frac{J}{2[R]_{\max}}}{\frac{J}{2[R]_{\max}} + (\alpha^{1/4} + \alpha^{-1/4})^2}. \tag{S5}$$

Since $\alpha^{1/4} + \alpha^{-1/4} \geq 2$, with equality only if $K_1 = K_{id}$, this tells us that equal operators are "best" for looping, in the sense that for given peak position and flexibility, equal operators maximize the looping probability, as shown in Fig. S1(C). Further intuition about this behavior comes from considering the competition of states illustrated in Fig. S2(B). The looping probability near the peak is dominated by a competition between the looped state, and the singly occupied state with the repressor bound to the strongest operator. Changing the relative operator strength while keeping $K_1 K_{id}$, and therefore $[R]_{\max}$, constant selectively strengthens that singly occupied state, and therefore poisons looping near the peak. However, in the limits of high and low concentrations, the looped state is instead out-competed by the doubly occupied and unoccupied state respectively, whose weights relative to the looped state only depend on the average binding strength $[R]_{\max} = (K_1 K_{id})^{1/2}$, so these regimes are not affected by changes in the relative operator strengths.

## S1.2 The case of multiple looped states.

Interestingly, looping by the Lac repressor is more subtle than the simple model described so far. Several studies have previously reported two looped states for LacI in the case of DNAs with two operators [1, 6, 7, 8, 9, 10, 11], and we observe these two states in the work presented here as well (see Figs. 2(F) and 4 in the main text, and Sec. S7 below). These two looped states have been attributed to flexibility in the tetramerization domain of the repressor and/or to superpositions of four DNA loop topologies [6, 7, 9, 12, 13, 14]. As noted there, in the case of either underlying physical origin, the two looped states can be compared by noting that they have the same dissociation constants but different effective J-factors (see Fig. 2(F) in the main text). So the looping probabilities of these two experimentally distinguishable states can be modeled as

$$p_{\text{loop},1} = \frac{\frac{1}{2}\frac{RJ_{\text{loop},1}}{K_1 K_{id}}}{1 + \frac{R}{K_1} + \frac{R}{K_{id}} + \frac{R^2}{K_1 K_{id}} + \frac{1}{2}\frac{RJ_{\text{loop},1}}{K_1 K_{id}} + \frac{1}{2}\frac{RJ_{\text{loop},2}}{K_1 K_{id}}} \tag{S6}$$

$$p_{\text{loop},2} = \frac{\frac{1}{2}\frac{RJ_{\text{loop},2}}{K_1 K_{id}}}{1 + \frac{R}{K_1} + \frac{R}{K_{id}} + \frac{R^2}{K_1 K_{id}} + \frac{1}{2}\frac{RJ_{\text{loop},1}}{K_1 K_{id}} + \frac{1}{2}\frac{RJ_{\text{loop},2}}{K_1 K_{id}}}. \tag{S7}$$

Note that Eqs. (S6) and (S7) sum to Eq. (S1) with $J_{\text{loop}} = J_{\text{loop},1} + J_{\text{loop},2}$. Also note that the ratio of Eqs. (S6) and (S7) is just $J_{\text{loop},1}/J_{\text{loop},2}$, which means that the titration curves of the two states should have identical shape up to an overall scaling factor. This scaling can be seen in Fig. S1(D), where Eqs. (S6), (S7), and (S1) are plotted with $J_{\text{loop}} = 54$ pM, $J_{\text{loop},1} = J_{\text{loop}}/3$ (labeled the "bottom" state), and $J_{\text{loop},2} = 2J_{\text{loop}}/3$ (labeled "middle"), and in Fig. 2(F) in the main text.

## S1.3 Effect of an "inactive fraction" of repressor.

One question that arises in thinking about actual TPM experiments is how should we expect $p_{\text{loop}}([R])$ to change if the concentration of repressor that is presumed to have been pipetted into the TPM chamber isn't the "real" concentration of repressor that contributes to the observed looping? The most obvious way

in which this could happen is by measurement errors when determining the repressor stock concentration. Another possibility is that some repressor molecules cannot bind DNA, for example due to misfolding that affects both DNA binding sites. Unless they interact in some way with the functional repressor (e.g., via crowding effects at high concentrations) the effect is the same as in the first case: the concentration of active repressors is lowered. A third case is tested experimentally below (Sec. S5.3): some protein may bind nonspecifically to the TPM chamber walls and thereby not participate in observable looping. Note that in Section S1.4, we will consider the case in which dimers poison looping, by binding one operator but not forming a loop. Here, inactive repressors cannot bind DNA and so simply contribute to a discrepancy between the real and presumed concentrations.

We can model these cases of inaccurate concentration by a fraction $f$, such that $[R]$ is the concentration we believe we flow in, but the concentration of active repressors that contribute to looping is instead $f[R]$. Substituting $[R] \to f[R]$ in Eq. (S1) leads to

$$p_{\text{loop, inactive fraction}}([R]) = \frac{\frac{1}{2} \frac{f[R] J_{\text{loop}}}{K_1 K_{\text{id}}}}{1 + \frac{f[R]}{K_1} + \frac{f[R]}{K_{\text{id}}} + \frac{(f[R])^2}{K_1 K_{\text{id}}} + \frac{1}{2} \frac{f[R] J_{\text{loop}}}{K_1 K_{\text{id}}}}, \tag{S8}$$

which can be rewritten in the form

$$p_{\text{loop, inactive fraction}}([R]) = \frac{\frac{1}{2} \frac{[R](J_{\text{loop}}/f)}{(K_1/f)(K_{\text{id}}/f)}}{1 + \frac{[R]}{(K_1/f)} + \frac{[R]}{(K_{\text{id}}/f)} + \frac{([R])^2}{(K_1/f)(K_{\text{id}}/f)} + \frac{1}{2} \frac{[R](J_{\text{loop}}/f)}{(K_1/f)(K_{\text{id}}/f)}}. \tag{S9}$$

This means that uncertainty about the overall repressor concentration affects all parameters equally, by literally distorting the basic yardstick of the TPM titration assay. The effect on titration curves is a simple horizontal shift, as shown in Fig. S1(E). The experimental implication is that one cannot detect an inactive fraction from any distortion of the titration curve. On the other hand, the parameters are rescaled equally, so ratios of fitted parameters are insensitive to this kind of experimental uncertainty. Section S4.5 discusses the impact of this potential source of error on the fitted parameters for the experimental results presented in this work, and as mentioned above, Section S5.3 describes an experimental control to test for one of the potential sources of a concentration discrepancy, namely, loss of protein to the chamber walls.

## S1.4  Effect of the presence of dimers in solution.

The Lac repressor is a dimer of dimers, with each dimer of the wild-type tetramer forming a single DNA binding domain [15]. Therefore only tetramers can loop DNA, having two DNA binding domains in the same molecule; but dimers can bind individual operators. If a dimer binds one of the operators of a DNA molecule, that DNA cannot form a loop even if a tetramer binds the other site; thus dimers "poison" looping.

There are three conceivable scenarios that would lead to dimers in a solution of otherwise wild-type tetrameric repressors. First, at very low repressor concentrations the tetramer is thought to dissociate into its component dimers [16], a reaction governed by an equilibrium constant that we will call $K_{DT}$. A second possible scenario is one in which a fraction of repressors is damaged in some way due to the purification, storage, or thawing process, leading to an inability of some repressors to form tetramers. This will lead to a fraction of dimers that is constant with the total repressor concentration. Third, a fraction of monomers could be damaged such that when incorporated into a tetramer they result in a head that is unable to bind DNA. In this case, a fraction of tetramers would be "dimers" in the sense that one head can bind DNA but the protein cannot loop the DNA; however this would also result in a population of tetramers that cannot
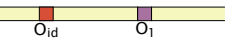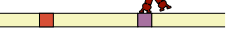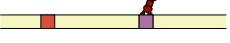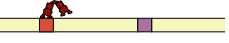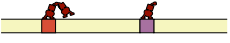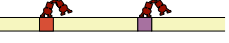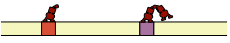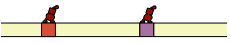
| STATE | STATISTICAL WEIGHT | STATE | STATISTICAL WEIGHT |
|---|---|---|---|
| (i) $O_{id}$ $O_1$ | $1$ | (vi) | $\dfrac{1}{2}\dfrac{[D]}{K_{id}}$ |
| (ii) | $\dfrac{[T]}{K_1}$ | (vii) | $\dfrac{1}{2}\dfrac{[D]}{K_1}$ |
| (iii) | $\dfrac{[T]}{K_{id}}$ | (viii) | $\dfrac{[T][D]}{K_{id}K_1}$ |
| (iv) | $\dfrac{[T]^2}{K_{id}K_1}$ | (ix) | $\dfrac{[T][D]}{K_{id}K_1}$ |
| (v) | $\dfrac{1}{2}\dfrac{[T]J_{loop}}{K_{id}K_1}$ | (x) | $\dfrac{1}{4}\dfrac{[D]^2}{K_{id}K_1}$ |

Figure S3: Schematized states and Boltzmann weights for a model that includes dimers. Eq. (S13) is obtained by observing that since $[R] = [T] + [D]/2$, states (ii) and (vii) combine to make the third term of Eq. (S13), states (iii) and (vi) to make the second term, and states (iv), (viii), (ix), and (x) to make the fourth term. Therefore the presence of dimers affects only the looped state—all other states are insensitive to whether a dimer or a tetramer is bound at each site.

bind DNA at all. We will first discuss $p_{\text{loop}}([R])$ for the case where we consider the tetramer-to-dimer dissociation at low concentrations, and then comment on the case of a constant fraction of dimers. The third case will not be considered but is well within the scope of scenarios that can be captured by the class of models presented here.

The most recent estimate of $K_{DT}$ for the Lac repressor from biochemical data claims an upper bound in the femtomolar range [16]. This estimate is obtained in part from the fact that no dimers have been observed at any concentrations used in biochemical experiments, which typically do not examine concentrations below about 1 pM (since $K_{DT}$ is the concentration at which half of the repressors in solution are dimers, $K_{DT}$ could be at most in the femtomolar or tens of femtomolar range, in order for the fraction of dimers to be essentially zero at picomolar repressor concentrations). Single-molecule techniques can, however, measure looping at concentrations below 1 pM, and in fact these concentrations are crucial to the determination of the shape of the looping probability versus concentration curve for some choices of operators and J-factors (see Fig. S1(A,B)). It is therefore critical to determine the effect of $K_{DT}$ on the looping probability.

In contrast to the simple model, a model that takes into account dimers must have ten states, five the same as those of the simple model and five that allow dimers or combinations of dimers and tetramers to bind to the DNA. These ten states and their associated weights are diagrammed in Fig. S3. To calculate the statistical weights of the dimer-containing states, we will assume that dimers and tetramers have the same operator dissociation constants. This assumption is reasonable given experimental evidence that at least some forms of LacI mutants that cannot tetramerize retain the same dissociation constants as wild-type repressor [17, 18, 19]. On the other hand, this assumption is not critical to the calculation and the more general case is a simple extension of the calculation presented here.

We also assume that the binding of a tetramer or dimer to a DNA tether does not affect the equilibrium between tetramers and dimers in solution. This assumption is similar to that discussed in Section S1.5 below, regarding the independence of different tethers in the same flow chamber. Since it is believed that tetramers dissociate into dimers at low total repressor concentrations, this is not obviously true; at low concentrations, one might indeed expect that single binding and association/dissociation events affect the binding and dimerization equilibria. However, the results of Sec. S1.5 shows that low repressor

concentration only affects the simple model for certain values of operator binding strengths and DNA J-factors. Analogously, we expect the calculation in this section to be reasonable for some parameter values, but note that the approach to low concentrations discussed in Sec. S1.5 could be extended to include dimerization effects as well.

Finally, we define the total repressor concentration $[R]$ such that

$$[R] = \frac{[D]}{2} + [T],$$ (S10)

where $[T]$ and $[D]$ are the concentrations of tetramers and dimers, respectively. This definition arises as follows: experimentally we measure the absorbance of purified repressor at 280 nm, and use the monomer extinction coefficient to obtain a mass concentration of monomers. We then divide by the molecular weight of a tetramer, which is 4 times the molecular weight of a monomer (since the Lac repressor is a homotetramer), to obtain what we call the concentration of repressor that we flow onto the slide. Therefore we can say that this concentration, $[R]$, is

$$[R] = \frac{\text{num. monomers}}{4 \times \text{vol}}.$$ (S11)

Since a tetramer is four monomers and a dimer two monomers, we have

$$[R] = \frac{2D}{4 \times \text{vol}} + \frac{4T}{4 \times \text{vol}},$$ (S12)

which simplifies to the equation above.

With these definitions and assumptions, we find that the partition function for the states schematized in Fig. S3 can be written as

$$Z_{\text{dimers}} = 1 + \frac{[R]}{K_{\text{id}}} + \frac{[R]}{K_1} + \frac{[R]^2}{K_1 K_{\text{id}}} + \frac{[T] J_{\text{loop}}}{2 K_1 K_{\text{id}}}.$$ (S13)

Note that the only state that has changed in the dimers model, compared to the simple model, is the looped state. This makes sense because the looped state is the only one in which it matters if a tetramer or a dimer is bound to the operators.

Finally, we make use of the definition of the tetramer-to-dimer dissociation constant for the reaction $T \leftrightarrows 2D$, which is

$$K_{DT} = \frac{[D]^2}{[T]}.$$ (S14)

This allows us to eliminate $[T]$ from Eq. (S13), leaving us with our final result in terms of $[R]$ and $K_{DT}$:

$$p_{\text{loop, dimers}} = \frac{\frac{[R] J_{\text{loop}}}{2 K_1 K_{\text{id}}} \left( 1 + \frac{K_{DT}}{8[R]} - \frac{1}{8} \left[ \left( \frac{K_{DT}}{[R]} \right)^2 + 16 \frac{K_{DT}}{[R]} \right]^{1/2} \right)}{1 + \frac{[R]}{K_{\text{id}}} + \frac{[R]}{K_1} + \frac{[R]^2}{K_1 K_{\text{id}}} + \frac{[R] J_{\text{loop}}}{2 K_1 K_{\text{id}}} \left( 1 + \frac{K_{DT}}{8[R]} - \frac{1}{8} \left[ \left( \frac{K_{DT}}{[R]} \right)^2 + 16 \frac{K_{DT}}{[R]} \right]^{1/2} \right)}.$$ (S15)

We recover the simple model in the limit that $K_{DT}$ is zero, that is, when tetramers never dissociate into dimers. Fig. S1(G) illustrates how the looping probability changes as $K_{DT}$ approaches the $K_d$'s for the operators. The presence of dimers has two main effects. First, tetramer dissociation breaks the symmetry of the titration curve (since it occurs only at low repressor concentrations), and therefore is an effect that can, at least in principle, be detected in TPM experiments, in contrast to an inactive fraction, which

rescales all the parameters. Second, since tetramer dissociation only occurs at low concentrations, it will contribute mainly an uncertainty factor in the determination of binding constants. The J-factor should be less affected, since it is strongly influenced by the high concentration data where dimerization is not an issue (see the discussion of Eqs. (S3) and (S4) above, and the numerical arguments discussed in Sec. S4.5 below).

We turn briefly to the case that involves the presence of a constant fraction of dimers at all concentrations. We note that as in the previous derivation involving $K_{DT}$, the only state that is affected by the binding of a dimer versus a tetramer is the looped state. The weights of all other states depend only on the total repressor concentration $[R]$. We can therefore start with Eq. (S13), but then define the concentration of repressors in tetrameric form, $[T]$, to be

$$[T] = (1 - \nu)[R], \tag{S16}$$

where $\nu$ is the fraction that are dimers. Then because $\frac{[D]}{2} + [T] = [R]$, we must define $\nu$ as

$$\nu = \frac{[D]}{2[R]}, \tag{S17}$$

so that $[D]/2 + [T] = 2\nu[R]/2 + (1 - \nu)[R] = [R]$. We can now use this expression for $[T]$ in Eq. (S13), so that when we form the looping probability we obtain

$$p_{\text{loop, const. dimers}} = \frac{\frac{(1-\nu)[R]J}{2K_1 K_{id}}}{1 + \frac{[R]}{K_{id}} + \frac{[R]}{K_1} + \frac{[R]^2}{K_1 K_{id}} + \frac{(1-\nu)[R]J}{2K_1 K_{id}}}. \tag{S18}$$

As with the case of an inactive fraction, this model involves only a rescaling of the parameters $K_1$, $K_{\text{id}}$, and $J_{\text{loop}}$, by a factor $1/(1 - \nu)$, and therefore cannot be distinguished from the simple model in the absence of additional information. Fig. S1(F) above shows how a constant fraction of dimers affects the looping probability as a function of concentration.

## S1.5   Low concentrations

A major assumption behind the simple titration curves featured throughout this paper is that the conformations of different tethers are independent, so that we only have to consider one tether. However, a test chamber contains many tethers, and one of the ways they might interact is if binding of repressor molecules at some tethers significantly decreases the number of repressors available for binding to other tethers. Intuitively, we expect this to be an issue at low concentrations only, where the total number of repressors is comparable to, or smaller than, the total number of tethers. In that case, the number of repressors available for binding might become significantly less than the total number of repressors, which lowers the looping probability. On the other hand, low concentrations also increase the probability of the unoccupied state. These two trends compete, and in the following, we will use a simple mean field analysis to estimate how these competing effects play out at low repressor concentrations.

The starting point for this analysis is the observation that the repressor molecules in the test chamber are either bound to an operator, or free in solution and available for binding. (Here, as in [1], we assume the binding of repressors to non-operator DNA is negligible, based on the relative magnitudes of the non-operator DNA concentration in the chamber and the association constant of repressor to non-operator DNA.) If we define a tether concentration [DNA] as the total number of tethers divided by the total volume, and $\langle n \rangle$ as the average number of bound repressors per tether (that is, $n$ can be 0, 1, or 2, since

each tether can have zero, one, or two repressors bound, and $\langle n \rangle$ is the average across all tethers in the sample), we can divide the total repressor concentration into a free and a bound part according to

$$[R] = [R]_{\text{free}} + [\text{DNA}] \langle n \rangle , \tag{S19}$$

where $[R]_{\text{free}}$ is the average concentration of free (unbound) repressors. Next, we make the approximation that the tethers are in equilibrium with the average repressor concentration $[R]_{\text{free}}$. This means that we neglect both temporal and spatial fluctuations in repressor concentration, similar in spirit to simple mean-field theories of spin systems [20], and simply substitute $[R] \to [R]_{\text{free}}$ in the weights of the simple theory of Fig. S2(A). We can then use these weights to write an approximate expression for $\langle n \rangle$, by noting that $\langle n \rangle$ is the probability of having no repressors bound times zero, plus the probability of having one repressor bound times one, plus the probability of having two repressors bound times two. That is,

$$\langle n \rangle = \frac{\left( \frac{[R]_{\text{free}}}{K_1} + \frac{[R]_{\text{free}}}{K_{\text{id}}} + \frac{J[R]_{\text{free}}}{2K_1 K_{\text{id}}} \right) + 2 \frac{[R]_{\text{free}}^2}{K_1 K_{\text{id}}}}{1 + \frac{[R]_{\text{free}}}{K_1} + \frac{[R]_{\text{free}}}{K_{\text{id}}} + \frac{J[R]_{\text{free}}}{2K_1 K_{\text{id}}} + \frac{[R]_{\text{free}}^2}{K_1 K_{\text{id}}}}. \tag{S20}$$

We now have two equations for the two unknowns, $[R]_{\text{free}}$ and $\langle n \rangle$. These can be solved numerically, but it is also instructive to study the behavior at high and low repressor concentrations analytically. Since $\langle n \rangle \leq 2$, and $[\text{DNA}]$ is constant, the expected high concentration limit $[R]_{\text{free}} \approx [R]$, i.e., the simple model, can be read off from Eq. (S19). The low concentration limit of Eq. (S20), which we get by retaining only the linear term in the numerator and the constant term in the denominator, is

$$\langle n \rangle \to \frac{[R]_{\text{free}}}{c_T}, \text{ with } \frac{1}{c_T} = \frac{K_1 + K_{\text{id}} + J/2}{K_1 K_{\text{id}}}. \tag{S21}$$

If we substitute this back into Eq. (S19), we can solve for the fraction of free repressors at low concentration,

$$\frac{[R]_{\text{free}}}{[R]} \to \left( 1 + [\text{DNA}]/c_T \right)^{-1}. \tag{S22}$$

At intermediate concentrations, this ratio interpolates smoothly between the high and low concentration limits. The low concentration limit is interesting for two reasons. First, we note that the ratio $[R]_{\text{free}}/[R]$ in Eq. (S22) becomes independent of $[R]$. This means that the low concentration part of the titration curve is simply shifted to the right, as illustrated in Fig. S1(H). Second, the magnitude of the maximum shift depends on the tether properties, through the characteristic concentration $c_T$: only if this concentration is low compared to the tether concentration $[\text{DNA}]$ will the simple model fail at low concentrations. A high J-factor, low peak concentration $[R]_{\text{max}} = (K_1 K_{\text{id}})^{1/2}$, and large variation in operator strength (for a given peak concentration) leads to a low $c_T$. The intuition here is that these features all lead to an increased tendency to have repressors bound at low concentrations, which lowers the number of repressors in solution. Section S4.5.1 describes the applicability of these low-concentration considerations to the data presented in this work.

## S1.6  Calculating relative J-factors.

In Fig. 2(E) and Table 1 in the main text, we use a concentration titration and a fit to our model to measure the J-factors for E8 and TA in absolute units. However, if we are interested only in the relative flexibilities of two sequences, or if the J-factor for one construct is known and we wish to find the J-factor for a construct with the same operators but a different loop (as in Fig. 3(C) and (F) in the main text), our

model predicts that we can compute the ratio of the looping J-factors of two sequences based solely on a single pair of looping probabilities.

We can do so by fixing the concentration of repressor and measuring the looping probabilities of the two DNAs, and then computing the ratio

$$\frac{p_{\text{unloop}}/p_{\text{loop}}}{p'_{\text{unloop}}/p'_{\text{loop}}} = \frac{J'_{\text{loop}}}{J_{\text{loop}}}, \tag{S23}$$

where $p_{\text{unloop}}$ for a sequence is $1 - p_{\text{loop}}$. This result is obtained by starting with the probability of being in the *unlooped* state,

$$p_{\text{unloop}} = \frac{1 + \frac{R}{K_i} + \frac{R}{K_{ii}} + \frac{R^2}{K_i K_{ii}}}{1 + \frac{R}{K_i} + \frac{R}{K_{ii}} + \frac{R^2}{K_i K_{ii}} + \frac{1}{2}\frac{R J_{\text{loop}}}{K_i K_{ii}}}. \tag{S24}$$

We then construct the ratio of unlooping-to-looping probability for a single DNA. The ratio of $p_{\text{unloop}}$ to $p_{\text{loop}}$ is given by

$$\frac{p_{\text{unloop}}}{p_{\text{loop}}} = \left( \frac{1 + \frac{R}{K_i} + \frac{R}{K_{ii}} + \frac{R^2}{K_i K_{ii}}}{1 + \frac{R}{K_i} + \frac{R}{K_{ii}} + \frac{R^2}{K_i K_{ii}} + \frac{1}{2}\frac{R J_{\text{loop}}}{K_i K_{ii}}} \right) \left( \frac{1 + \frac{R}{K_i} + \frac{R}{K_{ii}} + \frac{R^2}{K_i K_{ii}} + \frac{1}{2}\frac{R J_{\text{loop}}}{K_i K_{ii}}}{\frac{1}{2}\frac{R J_{\text{loop}}}{K_i K_{ii}}} \right) \tag{S25}$$

which simplifies to

$$\frac{p_{\text{unloop}}}{p_{\text{loop}}} = \frac{1 + \frac{R}{K_i} + \frac{R}{K_{ii}} + \frac{R^2}{K_i K_{ii}}}{\frac{1}{2}\frac{R J_{\text{loop}}}{K_i K_{ii}}}. \tag{S26}$$

We form the same ratio $p'_{\text{unloop}}/p'_{\text{loop}}$ for a second DNA. When we divide these ratios of unlooped-to-looped probabilities for the two DNAs, we obtain

$$\frac{p_{\text{unloop}}/p_{\text{loop}}}{p'_{\text{unloop}}/p'_{\text{loop}}} = \left( \frac{1 + \frac{R}{K_i} + \frac{R}{K_{ii}} + \frac{R^2}{K_i K_i}}{\frac{1}{2}\frac{R J_{\text{loop}}}{K_i K_{ii}}} \right) \left( \frac{\frac{1}{2}\frac{R J'_{\text{loop}}}{K_i K_{ii}}}{1 + \frac{R}{K_i} + \frac{R}{K_{ii}} + \frac{R^2}{K_i K_{ii}}} \right). \tag{S27}$$

Note that the concentration dependence (as well as operator dependence) cancels. Fig. S8(A) below illustrates this key claim, that the ratio of J-factors computed from a pair of looping probabilities is independent of repressor concentration. We consider this concentration independence to Eq. (1) to be an important test of our model and also a reasonable basis for using single concentrations to measure relative J-factors. (We note, however, that as is clear from Fig. S8(A), some concentrations result in smaller measurement errors, and would therefore be better choices for measuring relative J-factors.)

Section S4.6 below discusses how Eq. (S23) was used in this work to calculate the J-factors presented in Fig. 3(C) in the main text.

## S2    Compiling looping predictions for Figure 4 in the main text.

Cyclization free energies according to the Shimada and Yamakawa approximation [5] were taken from [1, 12]. Predictions of the looping free energy of various looped conformations from [13, 14] were digitized with Enguage Digitizer (http://digitizer.sourceforge.net/). From [13], looping free energies of the V-like conformations P1 and A1 (called "vp" and "va" in Fig. 4 in this work) and the extended conformation P1E ("e" in Fig. 4) are taken from Fig. 4 of [13]. The extended conformation free energy of [13] contains a contribution that describes the cost of opening the Lac repressor tetramer to the extended conformation.

This term was estimated with an interval, reflected by the upper and lower edges of the gray polygon in Figure 4 in this work. From [14], looping free energies of the V-like conformation LB and extended conformation SL are taken from Fig 3(B) of that work. 20.5 bp was subtracted from the DNA length values used in that figure in order to convert it to the loop length convention used in [1, 12, 13], and in this work. Ref. [14] also discusses loop conformations similar to that labeled "vp" for [13]; these results were excluded for clarity, as they (and the "vp" conformation for [13] shown in the figure) contribute very little to the total J-factor.

As noted in the main text, detailed comparisons between these theoretical predictions and with our data may not be possible. However, we note the following observations: first, the most striking feature of our data not consistently captured by all of the models summarized here is that the two looped states we observe can have comparable J-factors at some lengths—that is, the curves for $J_B$ and $J_M$ intersect, at least without the promoter. This is surprising if one postulates that one of the two looped states corresponds to a V-like protein conformation and the other to a more extended protein conformation, as in [7], as computational analyses usually find the extended protein conformation to be so favorable in terms of the DNA mechanics as to generate a J-factor orders of magnitude larger than any V-like conformations at the lengths we examined (*e.g.* [14]). Second, the "B" and "M" states are out-of-phase with each other, much as the "v" and "e" states of [14] or the "A1" and "A2" or "P1" and "P2" states of [1, 12], but not the "va" and "e" states of [13]. Finally, as noted in Sec. S7 below, we observe direct interconversion between the two looped states, which would suggest that they differ in protein conformation and not loop topology; however it is possible given the 4-second Gaussian filter applied to our data that we are smoothing out short transitions to the unlooped state. We are therefore pursuing a hidden markov model-based kinetic analysis of our data which we hope will clarify whether or not direct interconversions between the looped states occur.

## S3    DNAs.

The preparation of the DNAs used in this work are described in the Materials and Methods section of the main text. Table S1 gives the sequences of the operators used in this paper, and for completeness the remaining naturally occurring operator $O_3$. Figures S4 and S5 shows the E8 and TA sequences that form the loops in the constructs used in this work.

| Name | Sequence |
|------|----------|
| $O_1$ | AATTGTGAGCGGATAACAATT |
| $O_2$ | GGTTGTTACTCGCTCACATTT |
| $O_3$ | GGCAGTGAGCGCAACGCAATT |
| $O_{id}$ | AATTGTGAGCGCTCACAATT |

Table S1: Sequences of the three naturally occurring Lac repressor operators $O_1$, $O_2$, and $O_3$, and of the synthetic $O_{id}$ ("Oideal") operator. All sequences are 5´ to 3´ and are from [21]. Note that $O_{id}$ is perfectly symmetric about its midpoint, whereas the naturally occurring, weaker operators are only pseudo-symmetric, with $O_1$ being the strongest, $O_2$ weaker, and $O_3$ the weakest. In this work, the loop is to the 3´ end of the appropriate ($O_{id}/O_1/O_2$) operator sequence shown here; the $O_1$ that is constant in all constructs (nearest to the bead) has the loop 5´ to the sequence given here.

```
E89 : GGCCG--- --- ---GCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGAC --- --- --- --- --- --- C
E90 : GGCCG--- --- --TGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGAC --- --- --- --- --- --- C
E91 : GGCCG--- --- -CTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGAC --- --- --- --- --- --- C
E92 : GGCCG--- --- -CTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACG-- --- --- --- --- --- C
E93 : GGCCG--- --- -CTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGT- --- --- --- --- --- C
E94 : GGCCG--- --- --- --TGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCC -- --- --- --- C
E95 : GGCCGG -----GCTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGT- --- --- --- --- --- C
E96 : GGCCG ----AGGCTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGT- --- --- --- --- --- C
E97 : GGCCG ----AGGCTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTC--- --- --- --- --- C
E98 : GGCCG ---AGGCTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCC-- --- --- --- --- C
E99 : GGCCG ---GAGGCTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCC-- --- --- --- --- C
E100: GGCCG ---GAGGCTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCG- --- --- --- --- C
E101: GGCCG --- --- --- GCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCCAGCCG--- --- C
E102: GGCCG --- --- --- GCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCCAGCCGA-- --- C
E103: GGCCG --- --- --- GCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCCAGCCGAC- --- C
E104: GGCCG --- --- --- GCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCCAGCCGACG--- C
E105: GGCCG --- --- --TGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCCAGCCGACG--- C
E106: GGCCG --- --- -CTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCCAGCCGACG--- C
E107: GGCCG --- --- GCTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCCAGCCGACG--- C
E108: GGCCG --- -- GGCTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCCAGCCGACG--- C
E109: GGCCG --- -AGGCTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCCAGCCGACG--- C
E116: GGCCGGCGGAGGCTGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACGTCCGCCAGCCGACGACGC


T89 : --- --- --- -GGCCG--- --- -GGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGT-- --- --- --- --- C
T90 : --- --- --- -GGCCG--- ---TGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGT-- --- --- --- --- C
T91 : --- --- --- -GGCCG--- --TTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGT-- --- --- --- --- C
T92 : --- --- --- -GGCCG--- --TTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC- --- --- --- --- C
T93 : --- --- --- -GGCCG--- --- -GGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCT- --- --- --- C
T94 : --- --- --- -GGCCG--- --- -GGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTA--- --- --- C
T95 : --- --- --- -GGCCG--- -ATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTC-- --- --- --- C
T96 : --- --- --- -GGCCG---AATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTC-- --- --- --- C
T97 : --- --- --- -GGCCG--TAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTC-- --- --- --- C
T98 : --- --- --- -GGCCG-TAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCT- --- --- --- C
T99 : --- --- --- -GGCCG-TAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTA--- --- --- C
T100: --- --- --- -GGCCG--TAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAG-- --- --- C
T101: --- --- --- -GGCCG--TAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGG- --- ---C
T102: --- --- --- -GGCCG--TAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGGC --- ---C
T103: --- --- --- -GGCCG--TAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGGCA-- --- C
T104: --- --- --- -GGCCG--TTAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGGCA-- --- C
T105: --- --- --- -GGCCG-TTAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGGCAC- --- C
T106: --- --- --- -GGCCG-TTAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGGCACG --- C
T107: --- --- --- -GGCCG-TTAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGGCACGT-- C
T108: --- --- --- -GGCCG-TTAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGGCACGTG- C
T109: --- --- --- -GGCCG-TTAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGGCACGTGTC
T116: CCGGTGCTAAGGCCGCTTAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGGCAC----C


601TA: ctggagatacCCGGTGCTAAGGCCGCTTAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAGG
CACGTGT-aagatatatacatcctgtgcatgta
```

Figure S4: Sequences of the no-promoter E8 and TA constructs used in this work (Fig. 3(A)-(C)). All sequences are 5´ to 3´ and are listed such that the $O_{id}$ operator (or $O_1$ or $O_2$ operator in the case of the E894 sequence) is immediately 5´ of these sequences, and $O_1$ is immediately 3´. Bolded sequence labels indicate constructs examined by cyclization in [22]; the rest were designed and created by the authors for this work (see Materials and Methods section in the main text). In the top section containing the E8 sequences, dashes indicate bases missing relative to the 116 bp E8 sequence listed at the bottom of that section. In the bottom section containing the TA sequences, dashes indicate bases missing relative to the full 154-bp 601TA sequence (provided to us by Jon Widom; see also [23]) listed below the TA sequences; in that 601TA sequence, the dash indicates where a C has been inserted at the end of all of the TA sequences used in both cyclization and in the looping work presented here. Upper-case letters in the full 601TA sequence indicate the region from which all TA sequences in this work were derived. The 601TA sequence is so named because of the TA dinucleotide steps which occur every 10 bp and which are thought to confer its affinity for nucleosome formation [24]; these TA steps have been highlighted in red. Note that the E8 sequence also has several TA steps spaced 10 bp apart; however this pattern does not repeat across the entire sequence as it does in the 601TA sequence, nor does the E8 sequence have other characteristics of the 601TA sequence such as GC pairs between the TA pairs which are also supposed to be important for its particular properties.

14

```
E92:  -- --- --- --- --- --- --- --- --- --- --- -CTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E93:  -- --- --- --- --- --- --- --- --- --- --- CCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E94:  -- --- --- --- --- --- --- --- --- --- --GCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E95:  -- --- --- --- --- --- --- --- --- --- -CGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E96:  -- --- --- --- --- --- --- --- --- --- TCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E97:  -- --- --- --- --- --- --- --- --- --ATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E101: -- --- --- --- --- --- --- --- -ATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E103: -- --- --- --- --- --- --- --TTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E105: -- --- --- --- --- --- ---TTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E114: -- --- --- --- TAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E115: -- --- --- --GTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E116: -- --- --- -CGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E117: -- --- --- GCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E118: -- --- --TGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E119: -- --- -CTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E120: -- --- GCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E121: -- --GGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E122: -- -CGGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E123: -- CCGGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC
E124: G-CCGGCTGCGTAGAACTACTTTTATTTATCGCCTCCACGGTGCTGATCCCCTGTGCTGTTGGCCGTGTTATCTCGAGTTAGTACGACC


T92:  -- --- --- --- --- --- --- --- --- --- --- -ACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T93:  -- --- --- --- --- --- --- --- --- --- --- AACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T94:  -- --- --- --- --- --- --- --- --- --- --AAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T95:  -- --- --- --- --- --- --- --- --- --- -TAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T96:  -- --- --- --- --- --- --- --- --- --- TTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T97:  -- --- --- --- --- --- --- --- --- --CTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T101: -- --- --- --- --- --- --- --- -ACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T103: -- --- --- --- --- --- --- --GCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T105: -- --- --- --- --- --- ---TAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T106: -- --- --- --- --- --- --CTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T107: -- --- --- --- --- --- -TCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T108: -- --- --- --- --- --- CTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T114: -- --- --- --- AGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T115: -- --- --- --TAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T116: -- --- --- -GTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T117: -- --- --- CGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T118: -- --- -- TCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T119: -- --- -GTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T120: -- ---GGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T121: -- --GGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T122: -- -CGGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T123: -- CCGGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
T124: G-CCGGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTC
```

Figure S5: Sequences of the with-promoter E8 and TA constructs used in this work (Fig. 3(D)-(F)). All sequences are 5´ to 3´ and are listed such that the $O_{id}$ operator is immediately 5´ of these sequences, and $O_2$ is immediately 3´ (but $O_2$ is the reverse complement of the sequence given in Table S1). The *lac*UV5 promoter is to the 3´ end, before the $O_2$ operator; its sequence is TTTACAATTAATGCTTCCGGCTCGTATAATGTGTGG. As in Fig. S4, TA steps have been highlighted in red. Dashes indicated bases missing from the 89 bp no-promoter equivalents shown in the previous figure.

# S4  Data analysis.

## S4.1  Determining the looping probability for each trajectory.

For the constructs of Fig. 2(D-E) in the main text (the constructs with 94-bp loops, which have primarily only the "middle" looped state), data for each tether were histogrammed separately and fit to one (all looped or unlooped), two (unlooped and one looped state), or, rarely, three (two looped states and an unlooped state) Gaussians. The looping probability was determined as the area under Gaussian(s) corresponding to the looped state(s) divided by the sum of the areas under all the Gaussians. This was done on a tether-by-tether basis; the mean looping probabilities and standard errors on these means for a population of tethers were then plotted in Fig. 2(D-E) (and likewise in similar figures in this SI).

However, for the predominantly three-state DNAs in Figs. 2(F) and 3, the two looped states were often not well described by Gaussians (see Figs. S11 and S12 for examples). We therefore investigated a thresholding approach to calculating each bead's looping probability. Threshholding was performed subsequent to the Gaussian-fitting method described above. The intersection points of the fitted Gaussians were used to identify initial threshold values, which were adjusted manually as needed, such that the thresholds split the trajectories into the unlooped state and any looped state(s). A threshold above which data were excluded was set to the mean RMS of the tether in the absence of repressor plus three times the standard deviation of the no-repressor RMS; data above this point were usually due to tracking errors or free beads in solution temporarily entering the field of view. An empirically determined lower bound was set at 80 nm to exclude sticking events. The looping probability was then determined as the number of data points between the thresholds that delineated looped state(s), divided by the total number of points in the trajectory below the topmost threshold and above the sticking-state threshold. For traces with well-separated and well-populated states, the looping probabilities calculated by this thresholding method were comparable to those calculated by the Gaussian fitting method; where they differed, we believe the thresholding method to better represent the behavior of the trajectory. Therefore all looping probabilities for the length series presented in Fig. 3 in the main text were obtained by this thresholding method.

## S4.2  Minimum number of beads and minimum observation time.

In order to make our measurements of dissociation constants and J-factors as precise as possible, we considered how many trajectories needed to be included in each population mean looping probability, and how long each trajectory needed to be, in order to obtain reproducible mean looping probabilities. We will briefly summarize our methods and conclusions here.

The minimum number of trajectories needed to measure the mean looping probability of a population of tethers under a given set of conditions depends both on intrinsic tether-to-tether variation in looping probability, and on the minimum observation time discussed below. With regards to the latter: consider, for example, a case where it takes 2000 seconds to obtain an accurate measure of a tether's looping probability. Only 20 trajectories may be needed to sample the intrinsic spread in looping probabilities in a population of tethers; but if trajectories that only last 1000 seconds are also included, those trajectories will increase the spread of the data and more than 20 tethers will be needed to accurately measure the mean.

Our approach to choosing a minimum number of trajectories is depicted in Figure S6. After histogramming the set of looping probabilities for a population of tethers under certain conditions of repressor concentration, etc. (Fig. S6(A)), we chose, with replacement, progressively larger subsets of these looping probabilities and recalculated the mean looping probability and the standard error of this mean, repeating this procedure $10^4$ times per subset size (Fig. S6(B,C)). If very few tethers are used to calculate the mean, the mean fluctuates wildly between resamplings (indicated by the vertical spread in the blue points); but
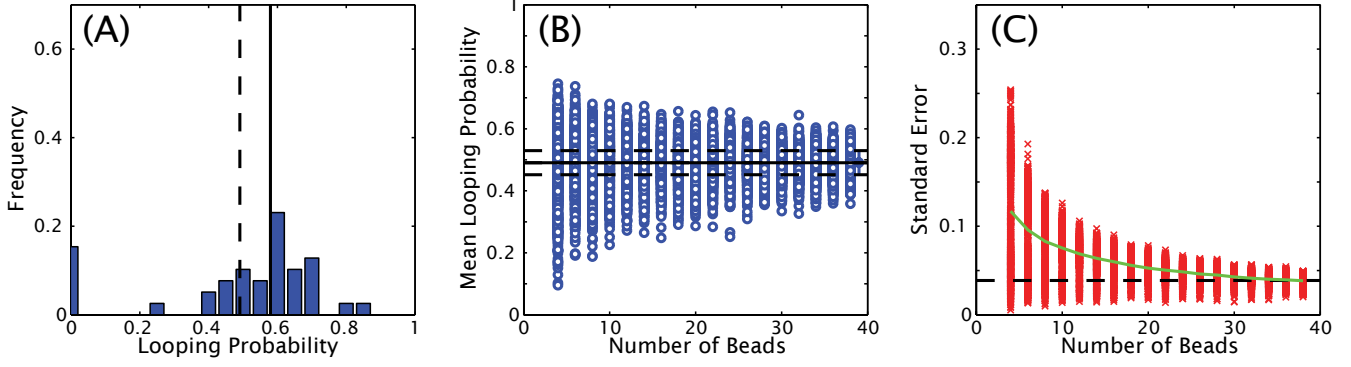
Figure S6: Distribution of looping probabilities obtained with Oid-E894-O1 at 10 pM repressor, as representative of the issues discussed in Sections S4.2 and S4.3. (A) Histogram of looping probabilities of all trajectories that lasted at least 3000 seconds, normalized by the total number of tethers included in the distribution. Black vertical dashed line is the mean of the distribution; solid black vertical line is the mean not including tethers that have zero looping probability (the "nonloopers" discussed in Sec. S4.3). (B) Mean looping probability as a function of the number of tethers included in the mean, resampled with replacement $10^4$ times per number included (that is, there are $10^4$ blue points per x-value). Solid horizontal black line is the mean of the distribution that includes all 39 tethers in the distribution in (A) (so it is the same as the vertical dashed line in (A)); dashed lines here indicate the mean plus or minus the standard error of the distribution with all tethers. (C) Standard errors of the resampled distributions whose means are plotted in (B); that is, there is a red "x" for every blue circle. Horizontal dashed line indicates the standard error with all tethers included; green curve is the standard deviation of the $10^4$ blue points at each x-value tested (so it is a measure of the vertical spread of the blue points in (B)).

as the number of tethers included in the mean surpasses 20, the spread in the blue points remains constant, suggesting we need about 25 beads to accurately calculate the true mean. Similarly, as the number of trajectories included exceeds 20-25, the standard error of the distribution decreases until it reaches a constant value, which is the error associated with intrinsic tether-to-tether variation in looping probability and will not decrease with more data. We found this number of 20-25 tethers to be consistent across repressor concentrations and DNA constructs, and so all reported means and standard errors in this work are obtained from sample sizes of at least 20 tethers.

As mentioned above, we asked not only how many trajectories were needed, but also how long each trajectory needed to be in order to be included in the analysis. As discussed in the next section, we considered several schemes for calculating the mean looping probability for a population of tethers. Most of these methods, and the analysis in [1], involved weighting each tether's looping probability equally in the calculation of the population mean (the other strategy, discussed below, is to weight each tether's looping probability by the observation time). For the schemes in which each tether was weighted equally, it is important to include in the population mean only those that have been observed long enough to obtain an accurate measure of their looping probability.

We took two approaches to analyzing the minimum observation time. The first is analogous to the approach used to determine the minimum number of trajectories: the mean and standard error for sets of looping probabilities with successively stricter minimum observation requirements were calculated, to determine at what cutoff the mean and standard error of the distributions converges to a constant. We found these calculated means and standard errors to be surprisingly insensitive to minimum observation time, except at very low repressor concentrations where looping events are rare. We therefore considered a second approach, in which we considered each tether's trajectory individually, and asked how its individual mean looping probability varied as data were removed from the end of the trajectory. Here again we found the looping probabilities for individual tethers to be surprisingly insensitive to observation time; where

17

they did vary, an ideal minimal observation time (above which a tether's looping probability generally remained constant) was about 3000 seconds. Therefore all tethers included in the analyses in this work lasted at least 3000 seconds.

## S4.3   Calculating the average looping probability.

The histogram of looping probabilities in Fig. S6(A) illustrates an aspect of our data that is not captured by the simple mean and standard error of a distribution, namely, the clustering of the looping probabilities of most trajectories around a single peak but a substantial fraction of trajectories that have zero looping probability. We see this bi-modal behavior, with a fraction of trajectories with no looping activity, at all concentrations where the rest of the distribution is sufficiently nonzero, for all DNAs derived from the low-copy pZS25 plasmid used in this work (that is, E8 and TA of all lengths and with various operators). However we do not see this nonlooping population with DNAs derived from the high-copy pUC19 plasmid used in [1], at least for DNA constructs that are twice as long as the 450 bp constructs discussed in this study. Moreover we observe the same fraction of these "nonloopers" with the different repressors batches discussed in Sec. S5.1 below. We therefore suspect that these nonloopers are caused by multiple DNAs tethering one bead, defects in a DNA tether, or some other DNA-specific factor.

As a result, we investigated several methods of describing the average behavior of the distribution of looping probabilities that involved different ways of handling the nonlooper population. If we assume the nonloopers to be DNA-specific but concentration independent, then for the concentration titrations of Fig. 2(D-F), we can calculate the average number of nonloopers at concentrations where most trajectories have significantly nonzero looping probabilities, and subtract this average fraction from the distributions at all concentrations. The problem with this method is the relatively small number of concentrations at which the nonloopers are sufficiently well-separated from the rest of the distribution from which to calculate the average. We found a more robust method to be to remove *all* trajectories with zero looping probability for those samples where the nonloopers were well-separated from the rest of the distribution; and then to use the average number of these nonloopers (about 10% for most constructs) for distributions with significant weight near zero. We find that subtracting nonloopers using this second approach results in a mean parameter that best represents the distribution of looping probabilities; and moreover, it results in fitted operator dissociation constants that agree well with values found in the literature—while including nonloopers does not (see Table S2). Therefore this second method was used for all data reported in this work. (We note here two cases in which the approach had to be modified: first, in the case of O2-E894-O1, no concentration resulted in a mean looping probability significantly above zero to completely separate the nonloopers. We therefore calculated the average fraction of nonloopers between 50 pM and 500 pM, where the zero bin was clearly in the tail of the distribution, and subtracted this average from all concentrations. Second, for the length series shown in Fig. 3 in the main text, where we do not have multiple concentrations for each construct, we applied the same approach but as a function of loop length: all nonloopers were subtracted from those lengths with clearly separated nonloopers, and an average number derived from these clear cases were subtracted from the rest.)

We note here an alternate approach to calculating the population mean (applicable when nonloopers have been kept or excluded): weighting each tether's looping probability by the amount of time it was observed, so that longer trajectories contribute more to the calculation of the mean, and to calculate the errors on each mean by bootstrapping the distribution (similar to the process described in Sec. S4.6 below). If all tethers are drawn from the same population, that is, if all tethers behave the same way, then observing many tethers for shorter amounts of time should be equivalent to observing only a few tethers for longer times. This method could be particularly relevant at low repressor concentrations, which may require

very long observation times to measure the equilibrium looping probability due to the limiting amount of repressor. However, we found that that weighting by observation time, versus weighting all tethers equally, has no statistically significant effect on the calculation of the mean.

## S4.4  Fitting concentration curves.

One of the assumptions behind Eq. (1) is that binding and looping are independent, i.e., that the binding constants do not depend on the DNA outside the operator sites, and that the looping J-factor does not depend on the operator sites. Hence, we should be able to model all our data for the TA and E8 sequences with three binding constants, $K_1$, $K_2$, and $K_{id}$, and two J-factors, $J_{\text{loop,TA}}$ and $J_{\text{loop,E8}}$.

To see that this is indeed possible, we fit Eq. (S1) to looping data in two different ways. "Individual fits" involved independent parameters for each data set (we only used one binding constant if both operator sequences were the same), while "global fits" involved fitting several data sets simultaneously with the five parameters mentioned in the previous paragraph. That is, an individual fit to Oid-E894-O1 had three free parameters, $K_{id}$, $K_1$, and $J_{\text{loop, E8}}$; but a global fit to the three E8-containing data sets has only four, because it requires that all three data sets be fit with the same J-factors and the same $K_1$, reflecting the reality of the DNA constructs.

Fitting was performed in Matlab with a weighted nonlinear least squares method, which means that we minimized

$$\chi^2 = \min_{\theta} \sum_{k=1}^{n} \left( \frac{\overline{p}_{\text{loop},k} - p_k(\theta)}{\sigma_k} \right)^2 \tag{S28}$$

with respect to the free parameters $\theta$ (which would include dissociation constants and J-factors). In Eq. (S28), $n$ is the number of data points to fit (i.e., concentrations and, in the case of the global fits, DNAs as well), $\overline{p}_{\text{loop},k}$ is the mean looping probability for one DNA at one concentration, $\sigma_k$ is the standard error of $\overline{p}_{\text{loop},k}$, and $p_k(\theta)$ is the theoretical prediction (using Eq. (S1)) for data point $k$.

Fig. S7 shows the results of individual versus global fits for all of the constructs in Fig. 2(D-E) in the main text. Fit parameters, plus or minus the standard errors described in the next paragraph, are given in Table S2. The difference in looping probability titration curves from the individual and global fits are within the experimental uncertainty almost everywhere, which indicates that the different data sets are indeed consistent with a single set of parameters.

We estimated the standard errors of the fit parameters using a bootstrap method [25], in which we constructed $10^4$ resampled data sets (with replacement) at each concentration, computed the mean and standard error of the looping probability for each of these resampled sets, and redid both the individual and the global fits for each set. We then estimated the standard error of the fit parameters by the standard deviation of the bootstrap parameters.

As noted in the main text, we privilege the global fits over the individual fits, for several reasons: they better reflect the physical reality of the DNA constructs, they better constrain the parameter values in many cases, and they match more closely with values in the literature obtained through traditional bulk biochemical assays (see Table S2). For similar reasons, we consider the subtraction of some fraction of nonloopers to be justified; as mentioned in the previous section and in Sec. S5.1 below, we suspect these nonloopers to be experimental artifacts derived from the DNAs we use.

The Oid-E107-O1 concentration curve shown in Fig. 2(F) in the main text was fit separately from the other concentration curves, assuming the values of $K_{id}$ and $K_1$ obtained from the global fit to the three E894 data sets and the TA94 data set (minus nonloopers). The looping probabilities for the two looped states were fit simultaneously, like the global fits above, but to Eqs. (S6) and (S7) in Sec. S1.2, assuming the two looped states have the same $K_d$'s and differ only in J-factor. The errors were computed according
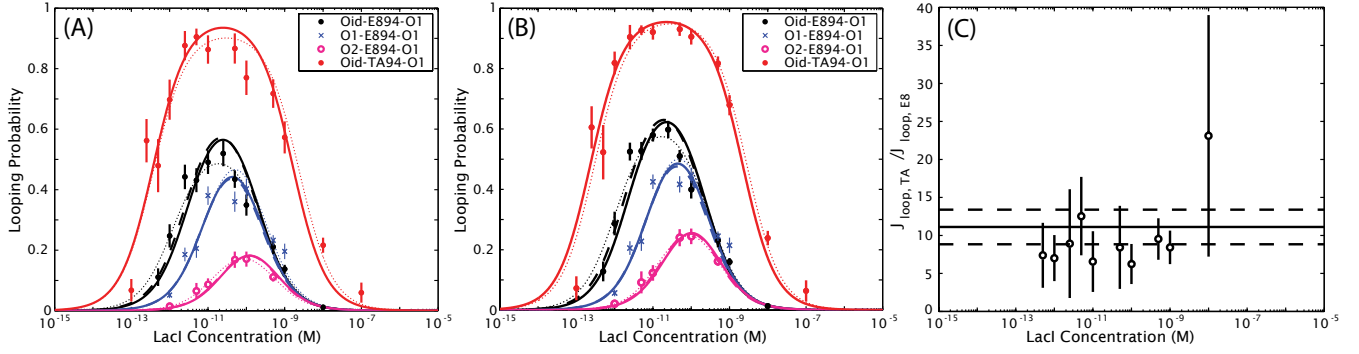
19

Figure S7: Comparison of individual and global fits to E894 and TA94 data, and of the effect of excluding some fraction of "nonloopers" (see Sec. S4.3). (A) Looping probabilities and individual and global fits for all data (that is, nonloopers included) for the constructs shown in Fig. 2(D-E) in the main text. (B) Same as (A), but with some nonloopers excluded, according to the scheme (see Sec. S4.3 for details) where all nonloopers were discarded for concentrations where they are clearly separated from the rest of the distribution, and a constant fraction of nonloopers were discarded for the other concentrations. The data in this panel are the same as in Fig. 2(D) and (E) in the main text. In (A) and (B) here, the dotted lines represent individual fits to the data set of the corresponding color; the dashed lines correspond to a global fit with the three E8 data sets only; and the solid lines to a global fit to all 4 data sets simultaneously. Fit parameters are given in Table S2. For all data sets aside from Oid-E894-O1, the global fits with and without the TA data are essentially indistinguishable from the individual fits. (C) Ratio of J-factors as a function of concentration including nonloopers, shown for completeness with Fig. S8(A) below. See the caption of that figure for details.

to a similar bootstrapping method as that used above: data at each concentration were resampled with replacement $10^4$ times, and then the $10^4$ $K_d$'s obtained from the global fits to the other concentration curves were used to refit the $10^4$ new Oid-E107-O1 data sets. As above, we estimated the standard error of the fitted J-factors for the two states by the standard deviation of the bootstrap parameters.

## S4.5 Additional potential sources of error: dimers at low concentration, the active fraction of repressor, and low repressor concentrations.

Though we find the model in Eq. (S1) to be sufficient for fitting the concentration titrations in this work, yielding operator dissociation constants that are consistent with values cited in the literature obtained by bulk methods (see Table 1 in the main text and Table S2 above), we also asked if the modifications to the model presented in Section S1 above can identify potential sources of error in our measurements. In particular, we wish to address whether it is necessary to take into account the dimer-to-tetramer transition at low repressor concentrations (Sec. S1.4), a discrepancy between the assumed repressor concentration and the actual concentration (Sec. S1.3), or an excess of tethers relative to repressors at low repressor concentrations (Sec. S1.5). By examining the effects of these potential sources of error on the values we measure, we can ask if it is possible from our data to determine upper bounds on $K_{DT}$, the dissociation constant for the tetramer-to-dimer transition, or $f$, the fraction of repressor that contributes to looping. We note, however, that as derived in Section S1, the dimer-to-tetramer transition and limiting repressors at low concentrations will affect only our measured $K_d$'s, and not the J-factor, since low-concentration data are crucial for measuring $K_d$'s but not $J_{\text{loop}}$.

We chose a numerical approach to address the questions of the dimer-to-tetramer transition at low concentrations, and a potential concentration uncertainty. The issue of an excess of tethers over repressors at low concentrations will be addressed in the next section. We first chose a range of reasonable values for

20

| Data | $K_{id}$ | $K_1$ | $K_2$ | $J_{\text{loop, E8}}$ | $J_{\text{loop, TA}}$ |
|---|---|---|---|---|---|
| Oid-E894-O1 | 3 ($\pm$ 1) pM | 130 ($\pm$ 30) pM | - | 320 ($\pm$ 50) pM | - |
| O1-E894-O1 | - | 50 ($\pm$ 4) pM | - | 350 ($\pm$ 30) pM | - |
| O2-E894-O1* | - | 20 (8, 130) pM | 300 (80, 1500) pM | 200 (100, 600) pM | - |
| Oid-TA94-O1* | 5 (2, 50) pM | 200 ($\pm$ 100) pM | - | - | 4400 ($\pm$ 700) pM |
| Global Fit, E8 only | 12 ($\pm$ 2) pM | 42 ($\pm$ 3) pM | 310 ($\pm$ 70) pM | 260 ($\pm$ 20) pM | - |
| Global Fit, E8 & TA | 14 ($\pm$ 4) pM | 44 ($\pm$ 4) pM | 340 ($\pm$ 70) pM | 280 ($\pm$ 20) pM | 3000 ($\pm$ 600) pM |
| Oid-E894-O1 | 3 ($\pm$ 1) pM | 90 ($\pm$ 20) pM | - | 350 ($\pm$ 40) pM | - |
| O1-E894-O1 | - | 47 ($\pm$ 4) pM | - | 380 ($\pm$ 30) pM | - |
| O2-E894-O1 | - | 26 (11, 125) pM | 300 ($\pm$ 200) pM | 320 ($\pm$ 90) pM | - |
| Oid-TA94-O1 | 10 (5, 46) pM | 80 ($\pm$ 40) pM | - | - | 5500 ($\pm$ 600) pM |
| Global Fit, E8 | 9 ($\pm$ 1) pM | 42 ($\pm$ 3) pM | 210 ($\pm$ 40) pM | 300 ($\pm$ 20) pM | - |
| Global Fit, E8 & TA | 12 ($\pm$ 3) pM | 44 ($\pm$ 3) pM | 240 ($\pm$ 50) pM | 330 ($\pm$ 30) pM | 4200 ($\pm$ 600) pM |
| Literature values | 8.3 $\pm$ 1.7 pM [2] | 37 $\pm$ 5 pM [26, 27, 28] | 350 $\pm$ 130 pM [26] | - | - |

Table S2: Fit parameters for individual and global fits, with and without nonloopers subtracted. "Global fit, E8 only" includes only Oid-E894-O1, O1-E894-O1, and O2-E894-O1; "Global fit, E8 & TA" includes these three data sets and Oid-TA94-O1 as well. The top section is with all data; the bottom is with nonloopers subracted and is the same as Table 1 of the main text (shown here for comparison). The last row here is the same as the last row of Table 1 as well. An asterisk (*) indicates that the distributions of fit parameters obtained from bootstrapped data were multimodal. In most cases, the best fit parameter plus or minus an error that is the standard deviation of the fit parameters to boostrapped data is reported, as described in the text; however, for those cases in which the standard deviation includes negative parameters values, a 95% confidence interval is reported in parenthesis instead.

$K_{DT}$ and for $f$, and then used these values along with the $K_d$'s and E894 J-factor from our global fit (row six of Table 1 in the main text) in the modified models of Eqs. (S9) and (S15) to generate looping probabilities at a range of concentrations comparable to those used in the TPM assays. We then fit the original model of Eq. (S1) to these simulated data, and asked how closely the fitted parameter values matched to the "true" values that were the inputs to the simulated data. The results are shown in Table S3.

In the first case, that of considering the tetramer-to-dimer transition at low repressor concentrations, we find that the fit parameters to the data generated from the model with $K_{DT}$ (Eq. (S15)), but fit to the model without $K_{DT}$ (Eq. (S1)), were within error of the fit parameters for real data until $K_{DT}$ exceeded 50 to 100 fM. Therefore given the uncertainty in our experimental data, we can put an upper bound on $K_{DT}$ in the tens of femtomolar range. This is in good agreement with recent estimates of $K_{DT}$ from other techniques (see [16]), which put an upper bound on $K_{DT}$ in the femtomolar range. Again we note that, as concluded in Sec. S1.4 above, if the true value of $K_{DT}$ is above 50-100 fM but we do not take it into account in our fits, we would obtain systematic errors in the fitted values for the dissociation constants but not the J-factor. Therefore regardless of the actual value of $K_{DT}$, our measurement of the J-factor remains the same.

On the other hand, in the second case, that of an inaccuracy in the assumed repressor concentration, we find, as predicted in Sec. S1.3 above, that both the dissociation constants and the J-factor are affected. However, we find that the assumed value of the repressor concentration could vary by up to about 10% and we would obtain fit parameters within error of those we now have (or within error of current literature values for the dissociation constants); and that within this range our measurement of the J-factor would not change within experimental error.

Since the model that includes $f$ is not an independent model but involves a rescaling of the parameters of the original model in Eq. (S1), it cannot be used to fit a concentration curve unless one of the parameters ($f$, $K_i$, $K_{ii}$, or $J_{\text{loop}}$) is known from another source. In principle, however, we should be able to fit both the model with $K_{DT}$ and without $K_{DT}$ to any concentration curve. A fit of the model that includes $K_{DT}$ to the three E8 data sets of Fig. 2(D) in the main text, plus the TA data set of Fig. 2(E), yields $K_{id} = 8 \pm 2$

| $K_{DT}$ or $f$ | $J_{\text{loop, E8}}$ | $K_{id}$ | $K_1$ |
|---|---|---|---|
| 1 fM | 300 pM | 9 pM | 41 pM |
| 5 fM | 300 pM | 10 pM | 41 pM |
| 10 fM | 300 pM | 10 pM | 40 pM |
| 50 fM | 300 pM | 11 pM | 38 pM |
| 100 fM | 300 pM | 12 pM | 37 pM |
| 500 fM | 290 pM | 18 pM | 27 pM |
| 1 pM | 290 pM | 24 pM | 24 pM |
| 5 pM | 300 pM | 29 pM | 29 pM |
| 10 pM | 300 pM | 32 pM | 32 pM |
| 97% | 310 pM | 9 pM | 43 pM |
| 95% | 321 pM | 9 pM | 44 pM |
| 93% | 330 pM | 10 pM | 45 pM |
| 90% | 340 pM | 10 pM | 47 pM |
| 80% | 380 pM | 11 pM | 50 pM |
| 70% | 430 pM | 13 pM | 60 pM |
| (experimental) | 330 ($\pm$ 30) pM | 12 ($\pm$ 3) pM | 44 ($\pm$ 3) pM |

Table S3: Fit parameters of the original model (Eq. (S1)) to data generated by a model that takes into account the dimer-to-tetramer transition at low concentrations, or a potential inaccuracy in repressor concentration. The top section gives the fit parameters for data generated by Eq. (S15), with varying values of $K_{DT}$; the middle section, for data generated by Eq. (S9), with varying values of $f$; and the last section gives the fit parameters to real data (row six of Table 1 in the main text), which were used as the inputs to generate the data. When $K_{DT}$ exceeds 50 to 100 fM, or $f$ is smaller than at least 90%, the fit parameters to the generated data cease to be within error of the fit parameters to real data.

pM, $K_1 = 37 \pm 6$ pM, $K_2 = 210 \pm 30$ pM, $J_{\text{loop, E8}} = 300 \pm 20$ pM, $J_{\text{loop, TA}} = 4600 \pm 500$ pM, and $K_{DT} = 0.8$ (0.2, 20) pM. Here we are reporting a 95% confidence interval for the error on $K_{DT}$ because the fit was not well constrained when $K_{DT}$ was included. As noted above, the J-factors do not change appreciably compared to the fit that does not include the dimer-to-tetramer transition; and in fact the fitted $K_{DT}$ is low enough that the dissociation values do not change significantly either. We therefore conclude, as above, that $K_{DT}$ is no larger than 100's of femtomolar and is low enough to be irrelevant to our analysis.

### S4.5.1 Data analysis at low repressor concentration.

As noted in Sec. S1.5, the statistical mechanical model to which we fit concentration curves depends on the assumption that tethers are independent, that is, that the binding or unbinding of a repressor from one tether does not affect the binding or unbinding of repressors on other tethers. This assumption rests in turn on the assumption that repressors are always in excess of the number of tethers, so that the removal of one repressor from the solution when it binds to an operator does not change the effective concentration of repressors that the other tethers "see." This assumption is valid at most concentrations that we use; however, it could be called into question at very low repressor concentrations. To estimate when the assumption of an excess of repressors over tethers fails, in this section we estimate the number of tethers per chamber and then compare to the numbers of repressors per chamber as a function of concentration.

Our hand-made TPM chambers have volumes of about 40 $\mu$L, which means that at the lowest repressor concentrations we use, there are on the order of 240,000 (at 10 fM) to 24 million (at 1 pM) repressors per chamber. To estimate a typical number of tethers per chamber, we note that we usually see fewer than 50 tethers in a field of view, each of which is about $3 \times 10^9$ nm$^2$ in area, corresponding to roughly 0.3 nL in volume, given the double-sided tape's thickness of 100 $\mu$m. This means that in a 40 $\mu$L chamber, there will be on the order of 7 million tethers. Even if the estimate of tether density is an overestimate (given that 50 tethers per field of view is very high), 1 pM repressor still seems to be the lower bound on

concentrations we can use with our model for some choices of operators and J-factors, below which the assumption of an excess of repressors over tethers breaks down. In particular, according to Eq. (S21) and the parameter values given in Table 1, the Oid-TA94-O1 construct will be most sensitive to low repressor concentration effects. This perhaps contributes to the large spread in looping probabilities at 250 fM and 500 fM for Oid-TA94-O1, as evidenced by the much larger error bars on these data points than others.

## S4.6 Calculating J-factors without concentration curves for each construct.

This section describes the use of the results presented in Sec. S1.6 above to calculate absolute J-factors for the DNAs in the length series presented in Fig. 3(C) in the main text. First, however, the validity of Eq. (S23) was examined by using it to calculate a ratio of J-factors as a function of concentration for the Oid-E894-O1 and Oid-TA94-O1 data in Fig. 2(E) in the main text. The resulting J-factor ratios are plotted in Fig. S8(A), along with a solid horizontal line that indicates the ratio of J-factors obtained from a global fit to the E8 and TA data. The errors on the ratios were calculated using a bootstrap method similar to that described for fitting concentration curves in Sec. S4.4 above: the distribution of looping probabilities for each DNA at each concentration was resampled with replacement $10^4$ times. After each resampling, a new mean looping probability was computed, and then a new ratio of J-factors computed from these new means, for each E8-TA concentration pair. This resulted in a distribution of $10^4$ new J-factor ratios. The error was taken to be the standard deviation of this distribution. A similar procedure was used to compute the horizontal dashed lines in Fig. S8(A), which represent the error on the ratio of J-factors obtained from the global fit. As described in Section S4.4 above, these global fits were performed on bootstrapped data (resampled with replacement $10^4$ times), which yielded distributions of $10^4$ J-factor values for Oid-E894-O1 and Oid-TA94-O1. From each of the $10^4$ rounds of fitting, a new ratio of the fitted J-factors for TA and E8 were computed. The error plotted as horizontal dashed lines in Fig. S8(A) is the standard deviation of these $10^4$ ratios.

We now turn to the use of Eq. (S23) to compute J-factors for the constructs plotted in Fig. 3(C) in the main text, for which we do not have concentration titrations. The absolute J-factors in that figure were obtained from ratios with the 94-bp E8 construct whose J-factor is known from the concentration titrations of Fig. 2(D)-(E); that is, Eq. (S23) was used with DNA $i$ being one of the E8 or TA constructs of variable loop length whose looping probability is given in Fig. 3(A), and DNA $ii$ being the 94-bp E8 construct used in the concentration titrations of Fig. 2(D) and (E). However, instead of using the measured looping probability for Oid-E804-O1, we used the looping probability predicted for 100 pM based on the global fit to the three E8 data sets plus the TA data. We thereby obtained the ratio of the J-factor for the sequences in Fig. 3(A) to the J-factor for Oid-E894-O1 given by the global fit; and since this latter J-factor is known, we could then calculate J-factors for the other E8 and TA lengths.

To estimate the errors on these length-series J-factors, we bootstrapped the 4 sets of data used in the global fit for the Oid-E894-O1 J-factor, as well as the looping probabilities for each length-series constructs in Fig. 3(A). This resulted in $10^4$ new sets of predicted looping probabilities for Oid-E894-O1 at 100 pM, and $10^4$ new looping probabilities for each loop lengths in Fig. 3. For each of the $10^4$ new values, we computed absolute J-factors for the E8 and TA length series, as described in the previous paragraph, and then took the standard deviation of these new absolute J-factors. This standard deviation became the errors plotted in Fig. 3(C). We note that it is equally possible to use the looping probability for Oid-TA94-O1 as the reference instead of Oid-E894-O1; however, doing so results in much larger errors on the calculated J-factors, probably because the looping probability for the TA-containing sequence is very close to 1.

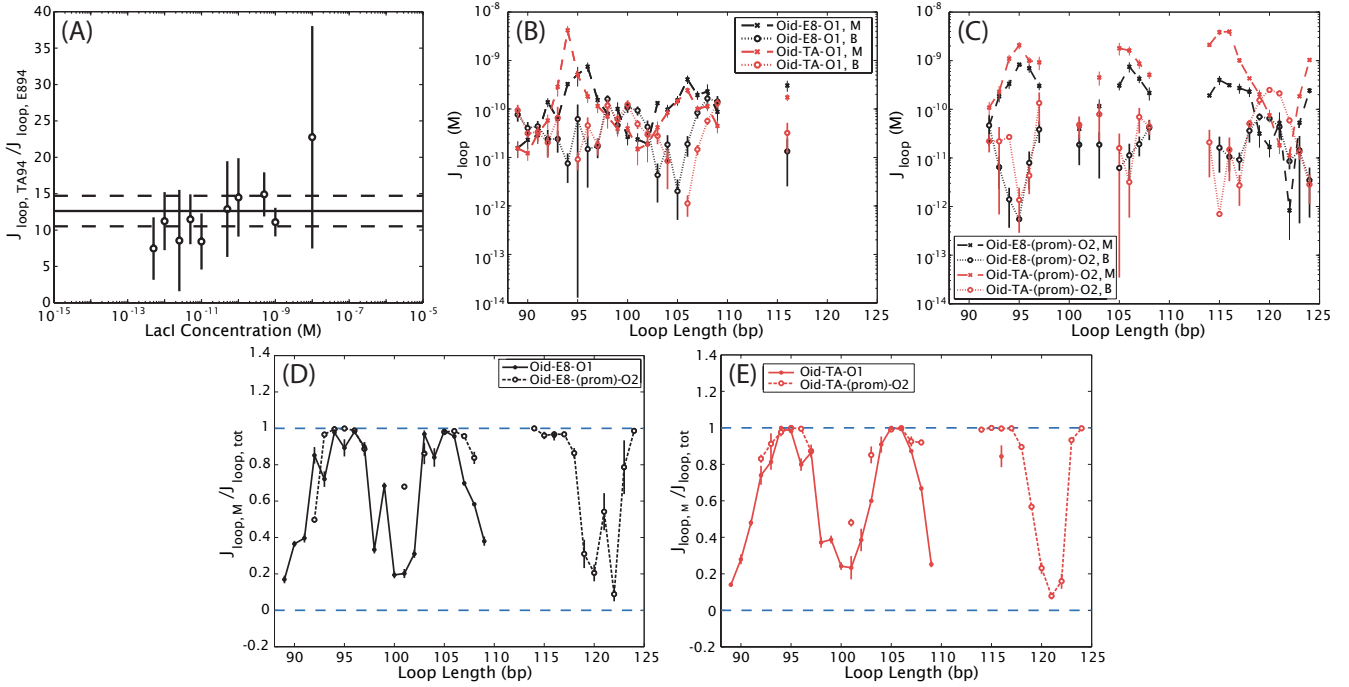The J-factors shown in Fig. 3(C) in the main text are for both looped states combined. It is also

Figure S8: Relative J-factors and the J-factors for the two looped states, with and without promoter. (A) Ratio of J-factors for the Oid-TA94-O1 and Oid-E894-O1 data shown in Fig. 2(E) in the main text, as a function of concentration, calculated from Eq. (S23); the errors are bootstrapped as described in the text here. The solid horizontal line corresponds to the ratio of J-factors obtained from the global fit to the concentration titrations in Fig 2(E) (that is, the ratio of the $J_{\mathrm{loop,TA}}$ to the $J_{\mathrm{loop,E8}}$ given in row six of Table 1 in the main text); the dashed lines are this ratio plus or minus a bootstrapped error, described in the text here. Note that it is necessary if Eq. (S23) is to hold that the ratio of $J$'s is constant across all concentrations: that is, it does not matter at what concentration we measure the looping probabilities for the two sequences, we will still obtain the same ratio of J-factors. It is clear, however, that some concentrations are better choices than others—for example, where both constructs show looping probabilities not too close to 0 or 1. (B) J-factors for the two looped states separately for the no-promoter constructs shown in Fig. 3(A)-(C) in the main text. Because of the results in (A) here, we can use Eq. (S23) to compute absolute J-factors from J-factor ratios, when one of the two components of the ratio is known. The result of using Eq. (S23) to calculate J-factors for the E8 and TA constructs of varying lengths (without a concentration curve for each construct) is shown in Fig. 3(C) in the main text; here we show J-factors for the two looped states of these constructs separately. The computation of these J-factors and associated errors are described in the text here. The E8 J-factors shown here are the same as are plotted in Fig. 4 in the main text. (C) J-factors for the two looped states separately for the with-promoter constructs shown in Fig. 3(D)-(F) in the main text. Here we cannot use Eq. (S23) because these constructs have a different combination of operators than any of the constructs used in the concentration titrations of Fig. 2 in the main text. However, we can use the fitted parameters for $K_2$ and $K_{\mathrm{id}}$ obtained from those data to calculate total J-factors for the with-promoter constructs, as shown in Fig. 3(F) in the main text, or J-factors for the two states separately (described in the text here). (D) The fraction of the total J-factor that is contributed by the J-factor of the middle state for E8, with and without the promoter. If only the middle state is occupied at a certain loop length, this fraction is 1; if only the bottom state is occupied, this fraction is 0. Where this ratio is not already close to 1, the addition of the promoter shifts it closer to 1 at most lengths, meaning that the promoter increases the probability that the middle looped state, rather than the bottom looped state, will form. (E) Same as (D) but for the TA constructs.

24

possible to calculate the J-factors separately for each of the two looped states (see Section S1.2 above). To do so we note that the sum of J-factors of the two looped states is the total J-factor calculated in the previous paragraph; and that the ratio of the J-factors of the two looped states is simply the ratio of their looping probabilities (which can be seen if Eq. (S7) is divided by Eq. (S6)). Fig. S8(B) shows the result of this calculation. The errors in Fig. S8(B) follow the usual formulation: for each DNA construct, the pair of looping probabilities corresponding to the two states was resampled with replacement $10^4$ times (that is, the resampling was done in such a way that the looping probability for the bottom looped state for one bead was not severed from the looping probability for the middle state for that particular bead), and $10^4$ new J-factors calculated in the same way as each state's mean J-factor. The error was then the standard deviation of these $10^4$ J-factors.

In the case of the promoter-containing constructs of Fig. 3(D)-(F) in the main text, the change of operator precluded the use of the above procedure for calculating either total or separate looping probabilities. However the dissociation constants for both $O_{id}$ and $O_2$ are known from the concentration curves of Fig. 2(D), and so the total J-factors for each of the constructs in Fig. 3(D) could be calculated directly from their looping probabilities by solving Eq. (S1) for $J_{\text{loop}}$. Then the errors were computed using a bootstrapping procedure similar to that described above, utilizing the $10^4$ bootstrapped fit values for $K_2$ and $K_{\text{id}}$ from the concentration curve fits. Finally the J-factors for the two looped states separately were computed by taking advantage, as in the no-promoter case, of the fact that the ratio of the J-factors for the two states is the ratio of their looping probabilities; and the errors were again bootstrapped. Fig. S8(C) shows these with-promoter J-factors.

## S5    Additional experimental controls.

### S5.1    Protein from different in-house purifications give the same results.

We initially collected data with purified repressor that was kindly shipped to us from the Kathy Matthews lab at Rice University. However, the fitted $K_d$'s from these results were not consistent with literature values; moreover, we could not obtain consistent results with repressor shipped at different times from different purifications in the Matthews lab. Upon a suggestion from Kathy Matthews that shipping the protein on dry ice may damage the protein, we purified two batches of protein in our lab, according to the Matthews lab protocol and after extensive help from their lab. As shown in Fig. S9(A), we were able to obtain consistent results and reasonable parameter values with protein purified in-house. Except where noted in Fig. S9(A), all data in this paper were obtained with the "SJLacI" batch.

### S5.2    Smaller beads result in similar looping probabilities.

A common concern with single-molecule experiments such as TPM that use large particles as reporters of molecular dynamics is that the reporters affect the observed dynamics. In particular in our TPM experiments, the 490-nm diameter bead is attached to a surface by a roughly 450-bp, or 150 nm, DNA tether, so it is reasonable to ask what the impact of excluded volume effects from the bead may be on the observed looping. Segall and coworkers (see [29]) explored these effects from a theoretical standpoint and found in the regime applicable to our experiments that halving the bead diameter would halve the force experienced by the tethered bead. We expect this force to affect primarily the measured J-factor: a smaller force would allow more looping and thereby increase the measured J-factor.

We measured the looping probability of the Oid-E894-O1 construct with 270-nm diameter beads in addition to the 490-nm diameter beads used in the rest of this work, and found that the measured looping
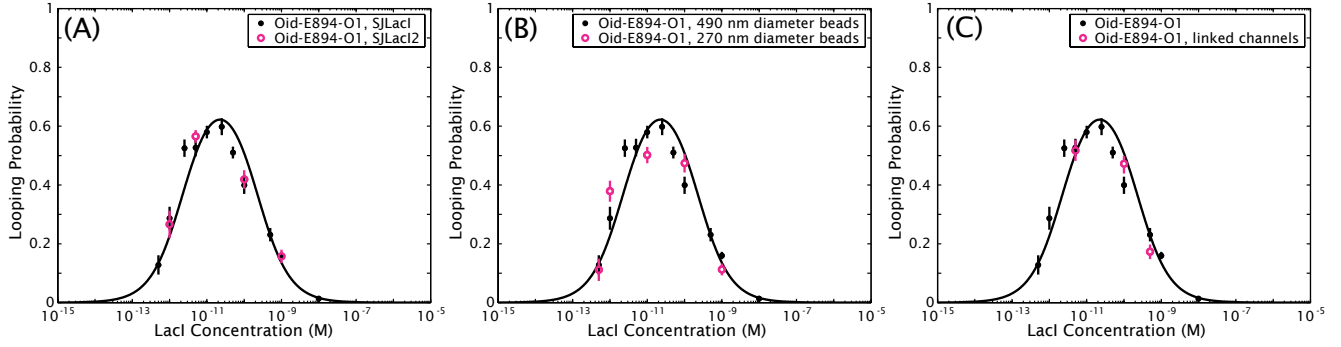
Figure S9: Effect of repressor purification batch, bead size and chamber surfaces on the measured looping probability. Different repressor batches (A), halving the diameter of the reporter bead (B), or linking two chambers to double chamber surface area (C) does not affect the looping probability of Oid-E894-O1 in a way consistent with theoretical predictions for varying protein activity, bead size, or protein loss to channel walls (see text for details). In all panels, the black data are the Oid-E894-O1 data (with nonloopers subtracted) shown in Fig. 2(D) of the main text, and the black curve is the solid black (global) fit of Fig. 2(E). The magenta data in all panels have likewise had some nonloopers subtracted as described in Sec. S4.3.

probability varied significantly at three out of five measured concentrations (Fig. S9(A)). However the trend in these discrepant points are not consistent with an increased effective J-factor. The 270 nm beads are difficult to image and track with the brightfield microscopy employed in this work, and we attribute discrepancies between the looping probabilities with 270-nm beads compared to 490-nm beads to be due to tracking difficulties, and not to an effect of bead size on the measured looping probability. (See also recent work by Milstein and coworkers [30], who found that looping and unlooping rates varied by only a factor of 2 between 800 nm and 50 nm reporter beads.)

## S5.3   No detectable loss of protein to chamber walls.

Our ability to use our standard model (Eq. (S1)) to obtain $K_d$'s and J-factors from concentration titrations depends on our knowing the absolute concentration of repressor in the TPM chamber. As described in Section S1.3 above, if the actual concentration of repressor in our sample is less than we assume, the parameters we measure will be scaled by some constant relative to their true values. In particular, we will measure effective $K_d$'s that are weaker than they should be, and J-factors that appear larger. We therefore asked if loss of protein to the walls of the TPM chamber could be affecting our measurement, as one potential source of a discrepancy between the assumed and actual concentration of repressor in our experiments. To do so, we made a chamber as described in [1], except no DNA was added to the 250 $\mu$L 3P introduced into the chamber after the anti-digoxigenin was washed out, and no beads were added. We then attached the output of this empty chamber to the input of a chamber prepared as usual. Repressor was introduced into the chamber with DNA via the empty chamber, and data were taken on the DNA-containing chamber.

The results of this "linked-channel" experiment, in which the surface area to which repressor could adsorb is effectively doubled, are shown in Fig. S9(B). If protein is adsorbing to channel walls, when the surface area is increased, data at lower concentrations than the maximum of looping (*e.g.*, the 5 pM data in Fig. S9(B)) should have a lower looping probability than in the normal single-channel experiment, and data at higher concentrations than the maximum (*e.g.*, 100 pM and 500 pM) should have higher looping probabilities. This is not what we observe, and so we cannot conclude that increasing the surface area of

the sample leads to a detectable change in looping probability. It should be noted, however, that due to the error on each measured looping probability, the effective repressor concentration in the linked chamber experiment would have to be significantly reduced, compared to the usual single-channel concentration, to be detectable (see Fig. S1(E) and Sec. S4.5 above).

## S6   RMS of the unlooped and looped states as a function of concentration and of loop length.

We hope that the data presented in this work will aid in future attempts to model the interactions of the Lac repressor with DNA, and to that end in this section we comment on the tether lengths of the unlooped and two looped states that we observe with TPM. In this work we have focused on using tether length as an indicator of the state (looped or unlooped) of the system, which in turn enables us to calculate looping probabilities; however we recognize that tether lengths contain additional information about the underlying conformation of the tethered DNA. For example, in Fig. S10(A) we discuss an effect that may be indicative of the bending of the operator DNA by bound Lac repressor, seen in the crystal structure of [15].

The tether lengths we observe in populations of otherwise identical tethered DNAs vary noticeably (see the black horizontal dashed lines in Figs. S11 and S12(A-B)), which we suspect arises at least in part from variations in bead diameter: the manufacturer reports a coefficient of variation of 1% in the base polystyrene particle, which should correspond roughly to a standard deviation in bead diameter of about 5 nm. Since what we measure as the effective tether length is the contour length of the DNA *plus* the radius of the bead, any variability bead-to-bead in diameter will manifest as effective tether lengths that vary from bead to bead for otherwise identical DNAs. What we report on the y-axes in Fig. S10 is the average tether length *relative to the "no lac" length*; that is, the average difference between a bead's unlooped or looped state(s) and the length of that particular tether recorded before repressor has been introduced into the TPM flow chamber. This allows us to resolve small but detectable changes in tether length in the presence versus the absence of repressor which would otherwise be obscured by the larger bead-to-bead variation in diameter. The improvement in resolution that we obtain by this method is perhaps one reason why we see evidence for operator bending where previous TPM experiments with the Lac repressor has not [7].

## S7   Representative Traces.

Figures S11 and S12 give representative examples of RMS motion versus time for the various constructs used in this work. Fig. S11, which shows examples from different repressor concentrations, illustrates several points mentioned elsewhere in this work: the presence of two experimentally distinguishable looped states at some loop lengths and concentrations (see also Fig. S12(A)); the difficulty of using a Gaussian-fitting method to obtain looping probabilities for some of these three-state trajectories (see Sec. S4.1 and Fig. S12(B)); and the shortening of the unlooped state observed at high repressor concentrations, discussed in Sec. S6. We note here also the difference in looped and unlooped state lifetimes at low and high repressor concentrations, even when a low concentration and a high concentration result in almost equal looping probabilities (compare Oid-E894-O1 at 1 pM and at 500 pM): at low concentrations, long dwells in the unlooped state are interspersed with bursts of looping transitions (or long dwells in the looped state, as for the Oid-TA94-O1 data at 500 fM), whereas there are more transitions and shorter dwell times at higher concentrations. We attribute the larger error bars on low concentration data for some constructs
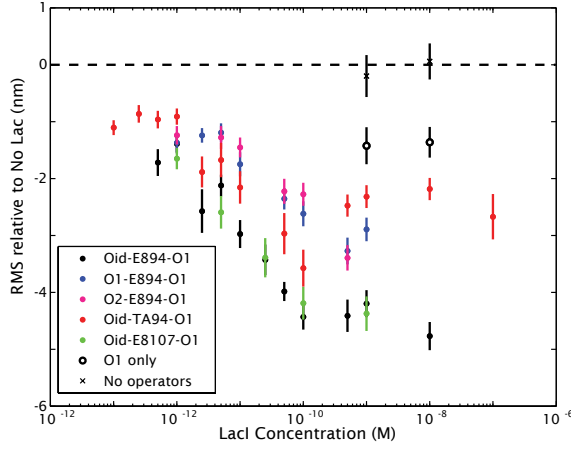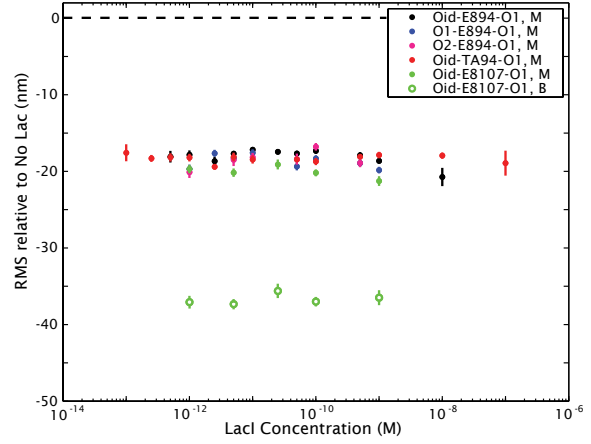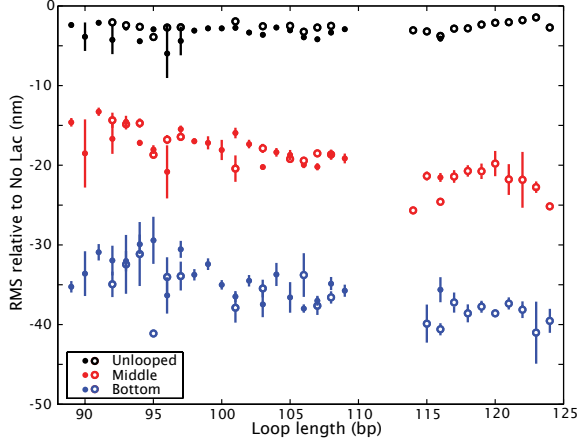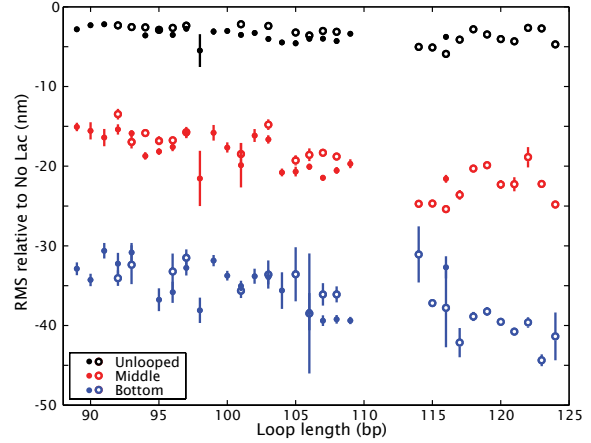
Figure S10: The effect of increasing repressor concentration or loop length on observed tether length. As noted in the text, due to initial tether length variability, we here report average RMS motion relative to the RMS in the absence of repressor. The tether length of each state was determined by the thresholding method described in Section S4.1 (even for constructs whose looping probabilities were determined by Gaussian fitting); the average tether length of the state was taken to be a weighted average of RMS values within the threshold limits. The average tether length of the DNAs with 94 bp loops in Fig. 2(D-E) in the absence of repressor is about 147 nm; the average tether length of the DNA with a 107 bp loop in Fig. 2(F) is comparable (150 nm). (A) The average relative RMS of the unlooped state as a function of concentration, for the five DNAs in Fig. 2(D-F), plus two DNAs which are missing one ("No Oid," which is (noOid)-E872-O1) or both ("No operators," (noOid)-E872-(noO1)) of the operators. These DNAs with missing operators are slightly shorter than the others in this figure, with average tether lengths in the absence of repressor of about 140 nm. For the two-operator DNAs, as the amount of repressor increases, the RMS of the unlooped state decreases. We suspect this shortening is due to the bending of the operator DNA by a bound repressor; in the crystal structure in [15], a repressor bound to the $O_{id}$ operator produces a 45° bend. At low repressor concentrations, the unlooped state that we observe in TPM is mostly composed of the state where nothing is bound to the DNA (state (i) of Fig. S2(A)), and so little operator bending is observed. However, at high concentrations, the unlooped state is composed mostly of the doubly-occupied state (state (iv) of Fig. S2(A)), and so the tether length shortens significantly. The addition of repressor has no effect on DNA with no operators at high repressor concentrations, suggesting that the shortening we observe is not due to nonspecific binding of the repressor to non-operator DNA. Note that the observed reduction in tether length for two operators will not necessarily be double the shortening of tether length for a single operator, depending on the phasing of the two bend angles. (B) Average relative RMS for the middle and bottom looped states as a function of concentration, for the constructs of Fig. 2(D-F). Unlike the unlooped state, the looped states are invariant with concentration. (C) Average relative RMS for the unlooped and looped states as a function of increasing loop length, for the E8-containing DNAs whose looping probabilities are shown in Fig. 3(A) (no promoter, closed circles) and Fig. 3(D) (with promoter, open circles) in the main text. Note the shortening of the unlooped state that we attribute to operator bending, since these data are taken at 100 pM repressor. The contour lengths of the with-promoter DNAs are slightly shorter than the no-promoter DNAs, with an average RMS in the absence of repressor of about 143 bp at a 94 bp loop length. (D) Same as (C) but for the TA-containing DNAs.

28

to finite observation time combined with few transitions between states, which leads to larger tether-to-tether variability in looping probability (see also Sec. S4.5.1). Finally we note that it appears that the two looped states can directly interconvert, which suggests that the two states differ in repressor conformation instead of operator binding orientation; however these trajectories have been downsampled to reduce file size (in addition to being Gaussian filtered), and so it is possible that very short transitions to the unlooped state between looped state interconversions are not visible.

Figure S12 shows representative examples of DNA constructs with only the bottom state (to complement the middle-state-only and both-states examples in Fig. S11), trajectories without clearly separated states, and distributions of looping probabilities. In Fig. S12(B), the top two trajectories illustrate a behavior observed in a minority of Oid-E894-O1 and O1-E894-O1 tethers: instead of the clear two states seen for most tethers with these constructs (see representative examples in Fig. S11), the looped and unlooped states have such similar RMS motions that they overlap in the histograms to the right of each trace. It is unclear what causes this behavior. The middle two trajectories in Fig. S12(B) illustrate a similar behavior that is more prevalent in the $O_2$-containing construct, which we attribute to the shorter looped-state lifetimes with this weaker operator compared to constructs with $O_1$ or $O_{id}$. The bottom two trajectories in (B) illustrate the kinds of Gaussian fits to poorly separated looped states that motivated a thresholding method, instead of a Gaussian-fitting method, for calculating the looping probabilities of the three-state DNAs in Figs. 2(F) and 3 in the main text (see Sec. S4.1). We were also interested in assessing whether thresholding the length series data of Fig. 3 would reduce the spread in the distributions of looping probabilities obtained from populations of otherwise identical tethers, shown in Fig. S12(C). All of the distributions of looping probabilities for the middle-state-only constructs used in the concentration titrations of Fig. 2(D-E) in the main text, for most of the with-promoter constructs of Fig. 3(D-F), and for some of the bottom-state only or both-states no-promoter constructs of Figs. 2(F) and 3(A-C), showed a clustered set of looping probabilities with a clear peak, as with the Oid-E8103-O1 example here. However, many of the no-promoter constructs in Fig. 3(A-C) in the main text showed broad distributions of looping probabilities, as with the Oid-TA106-O1 example here, and for a minority of constructs the distribution was so broad as to include almost all probabilities from 0 to 1, as with Oid-E109-O1. In some cases, the spread was reduced when the two states were histogrammed separately, as with Oid-E898-O1. The thresholding method, as compared to the Gaussian fitting method, reduced the spread in some but not all looping probability distributions. We note that even in cases where the looping probability distributions for the no-promoter constructs were very broad, the E8 and TA histograms were still so similar that we could conclude there was no sequence dependence to looping at those lengths.
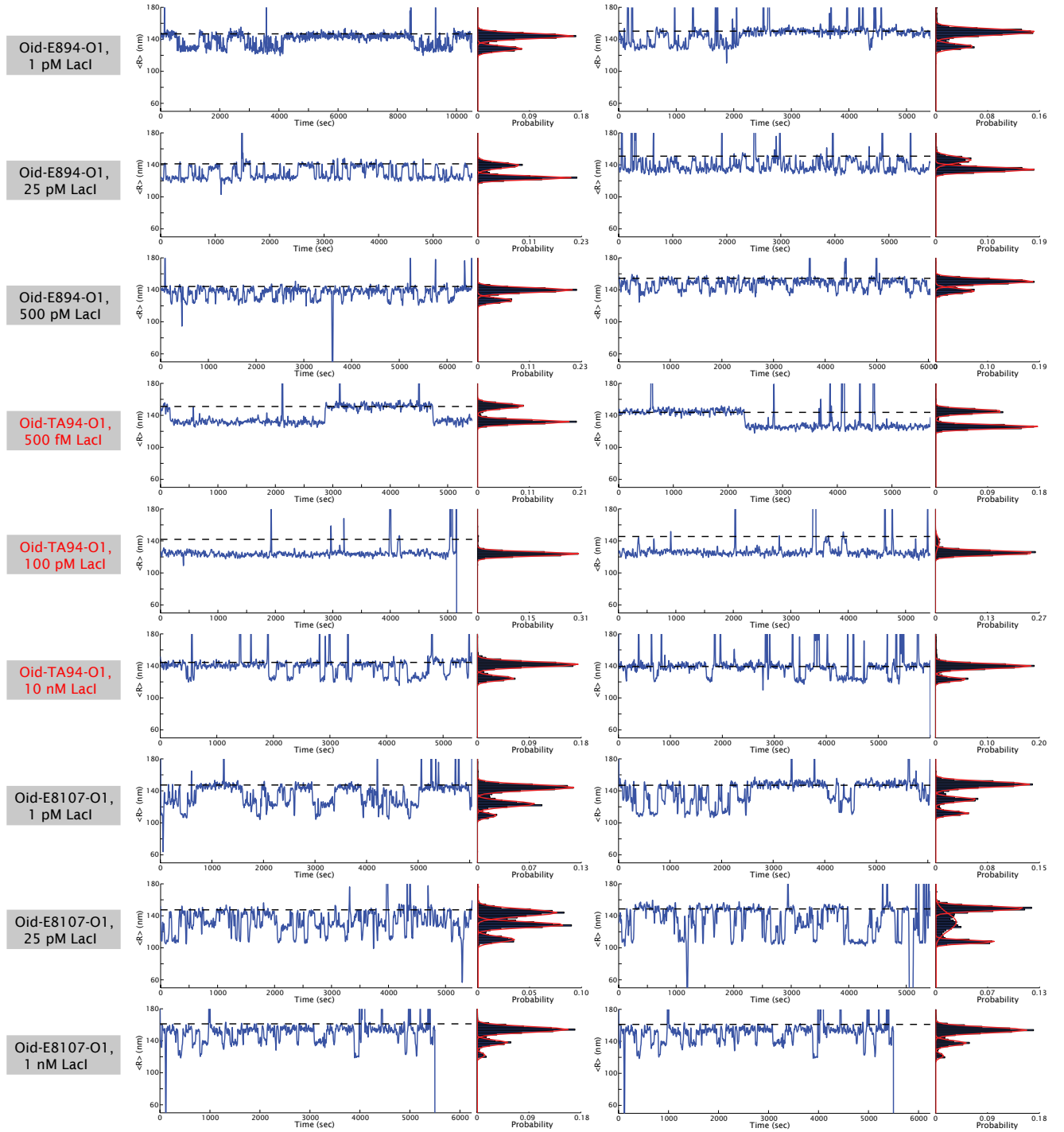
Figure S11: Representative examples of the average root-mean-squared motion ("<R>") of selected tethers as a function of time, for several of the DNA constructs shown in Fig. 2(D-F) in the main text, at several Lac repressor concentrations. To the right of each <R>-vs.-time trajectory is a histogram showing the probability of finding a given <R> over the whole trajectory. Red lines on this histogram show the results of a Gaussian fit, one way of determining the looping probability (see Sec. S4.1). The horizontal black dashed line in the plots of <R>-vs.-time indicate the average length of that particular tether in the absence of repressor. Excursions to RMS values less than about 80 nm are attributed to non-specific, transient adsorption of the bead to the surface, the DNA to the surface, or the DNA to the bead ("sticking" events); excursions to RMS values well above the horizontal black dashed lines are due to tracking errors (*e.g.* due to free beads in solution transiently entering the field of view). No trajectories for the O1-E894-O1 or O2-E894-O1 constructs are shown because they are essentially the same as those for Oid-E894-O1 (but see Fig. S12 for difficulties particular to the O2-E894-O1 construct).
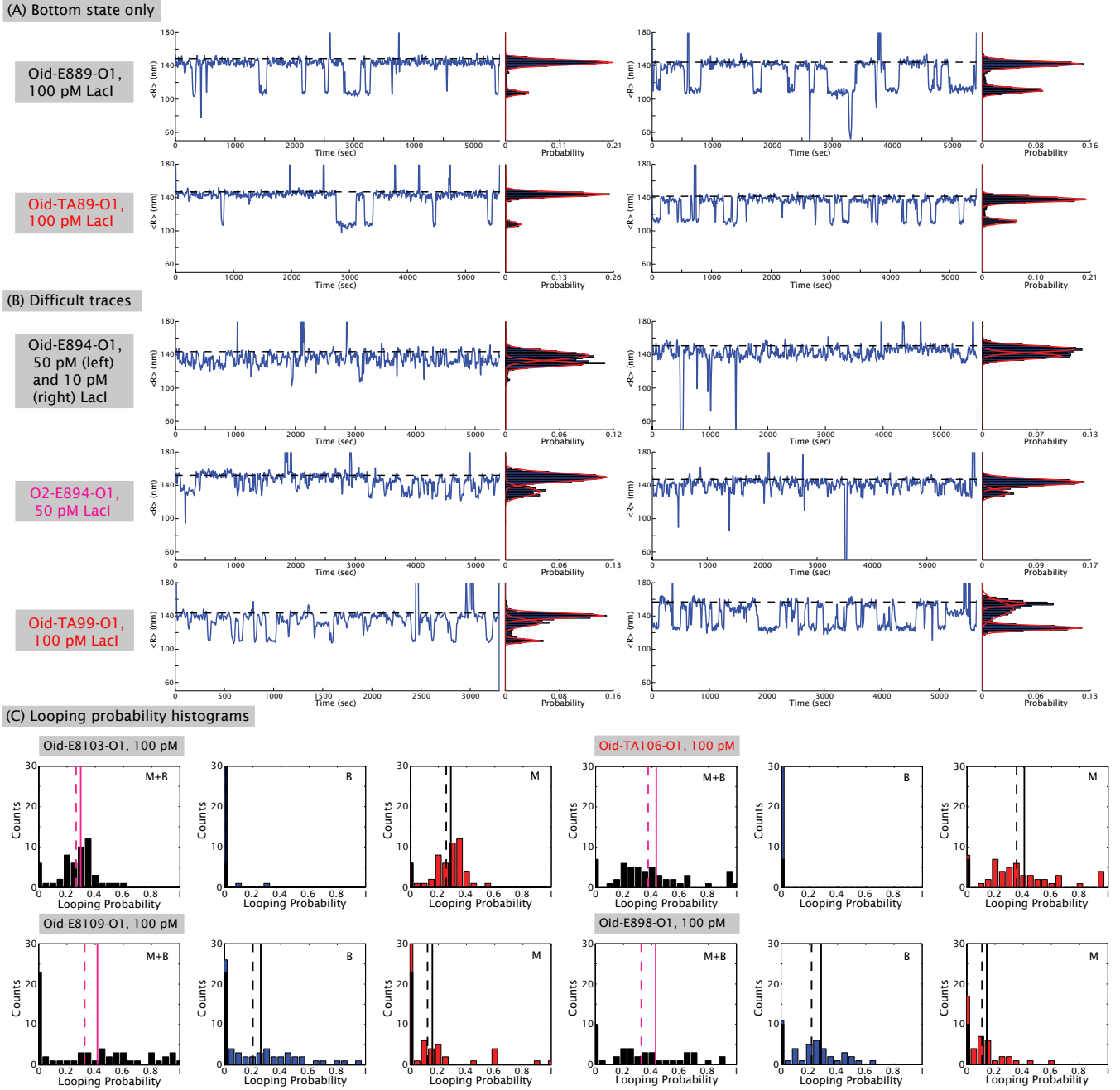
Figure S12: Representative examples of (A) DNA constructs with only the bottom state (to complement the middle-state-only and both-states examples in Fig. S11), (B) trajectories without the clear states shown in (A) here and in Fig. S11, and (C) distributions of looping probabilities. The plots in (A) and (B) are of the same kind as in Fig. S11 (see the caption of that figure for description). In (C), *unnormalized* looping probability histograms show the distributions of total ("M+B") looping probabilities, or the probabilities of the bottom ("B") or middle ("M") looped states, for four different DNA constructs at 100 pM repressor. The dashed line shows the mean looping probability with nonloopers included (see Sec. S4.3); the solid line shows the mean looping probability with nonloopers subtracted, which is the mean looping probability reported in the figures in the main text. In each "B" and "M" panel, the number of tethers that never loop at all are shown as a black bar in the zero bin; a blue or red bar above the black bar in the zero bin indicates the number of tethers that show no "B" state in the "B" panel, or no "M" state in the "M" panel.

31

# References

[1] Han, L., Garcia, H. G., S., B., Towles, K. B., Beausang, J. F., Nelson, P. C., and Phillips, R. (2009) Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS ONE,* **4**, e5621.

[2] Frank, D. E., Saecker, R. M., Bond, J. P., Capp, M. W., Tsodikov, O. V., Melcher, S. E., Levandoski, M. M., and Record, M. T. (1997) Thermodynamics of the interactions of *lac* repressor with variants of the symmetric *lac* operator: Effects of converting a consensus site to a non-specific site. *J Mol Biol,* **267**, 1305–1314.

[3] Zhan, H., Sun, Z., and Matthews, K. S. (2009) Functional impact of polar and acidic substitutions in the Lactose repressor hydrophobic monomer-monomer interface with a buried lysine.. *Biochemistry,* **48**(6), 1305–1314.

[4] Jacobson, H. and Stockmayer, W. H. (1950) Intramolecular reaction in polycondensations. I. The theory of linear systems. *J Chem Phys,* **18**, 1600–1606.

[5] Shimada, J. and Yamakawa, H. (1984) Ring-closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules,* **17**(4), 689–698.

[6] Edelman, L. M., Cheong, R., and Kahn, J. D. (2003) Fluorescence resonance energy transfer over 130 basepairs in hyperstable Lac repressor-DNA loops. *Biophys J,* **84**, 1131–1145.

[7] Wong, O. K., Guthold, M., Erie, D. A., and Gelles, J. (2008) Interconvertible Lac repressor-DNA loops revealed by single-molecule experiments. *PLoS Biol,* **6**, e232.

[8] Normanno, D., Vanzi, F., and Pavone, F. S. (2008) Single-molecule manipulation reveals supercoiling-dependent modulation of *lac* repressor-mediated DNA looping. *Nucleic Acids Res,* **36**, 2505–2513.

[9] Rutkauskas, D., Zhan, H., Matthews, K. S., Pavone, F. S., and Vanzi, F. (2009) Tetramer opening in LacI-mediated DNA looping. *Proc Natl Acad Sci USA,* **106**, 16627–16632.

[10] Mehta, R. A. and Kahn, J. D. (1999) Designed hyperstable Lac repressor·DNA loop topologies suggest alternative loop geometries. *J Mol Biol,* **294**, 67–77.

[11] Morgan, M. A., Okamoto, K., Kahn, J. D., and English, D. S. (2005) Single-molecule spectroscopic determination of Lac repressor-DNA loop conformation. *Biophys J,* **89**(4), 2588–2596.

[12] Towles, K. B., Beausang, J. F., Garcia, H. G., Phillips, R., and Nelson, P. C. (2009) First-principles calculation of DNA looping in tethered particle experiments. *Phys Biol,* **6**, 025001.

[13] Swigon, D., Coleman, B. D., and Olson, W. K. (2006) Modeling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation. *Proc Natl Acad Sci USA,* **103**, 9879–9884.

[14] Zhang, Y., McEwen, A. E., Crothers, D. M., Levene, S. D., and Fraser, P. (2006) Analysis of *in vivo* LacR-mediated gene repression based on the mechanics of DNA looping. *PLoS ONE,* **1**, e136.

[15] Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) Crystal structure of the Lactose operon repressor and its complexes with DNA and inducer. *Science,* **271**, 1247–1254.

[16] Barry, J. K. and Matthews, K. S. (1999) Thermodynamic analysis of unfolding and dissociation in Lactose repressor protein. *Biochemistry,* **38**, 6520–6528.

[17] Chen, J. and Matthews, K. S. (1994) Subunit dissociation affects DNA binding in a dimeric *Lac* repressor produced by C-terminal deletion.. *Biochemistry,* **33**, 8728–8735.

[18] Chen, J. and Matthews, K. S. (1992) Deletion of Lactose repressor carboxyl-terminal domain affects tetramer dissociation.. *J Biol Chem,* **267**(20), 13843–13850.

[19] Brenowitz, M., Mandal, N., Pickar, A., Jamison, E., and Adhya, S. (1991) DNA-binding properties of a Lac repressor mutant incapable of forming tetramers.. *J Biol Chem,* **266**(2), 1281–1288.

[20] Plischke, M. and Birger, B. (2006) Equilibrium statistical physics, World Scientific, Hackensack NJ 3rd edition.

[21] Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B., and Müller-Hill, B. (1994) Quality and position of the three *lac* operators of *E. coli* define efficiency of repression.. *EMBO J,* **13**(14), 3348–3355.

[22] Cloutier, T. E. and Widom, J. (2005) DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. *Proc Natl Acad Sci USA,* **102**, 3645–3650.

[23] Lowary, P. T. and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol,* **276**, 19–42.

[24] Widom, J. (2001) Role of DNA sequence in nucleosome stability and dynamics. *Quart Rev Biophys,* **34**, 1–56.

[25] Newman, M. E. J. and Barkema, G. T. (1999) Monte Carlo Methods in Statistical Physics, Oxford University Press, USA.

[26] Hsieh, W.-T., Whitson, P. A., Matthews, K. S., and Wells, R. D. (1987) Influence of sequence and distance between two operators on interaction with the *lac* repressor. *J Biol Chem,* **262**, 14583–14591.

[27] Whitson, P. A., Olson, J. S., and Matthews, K. S. (1986) Thermodynamic analysis of the Lactose repressor-operator DNA interaction. *Biochemistry,* **25**, 3852–3858.

[28] Whitson, P. A. and Matthews, K. S. (1986) Dissociation of the Lactose repressor-operator DNA complex: Effects of size and sequence context of operator-containing DNA. *Biochemistry,* **25**, 3845–3852.

[29] Segall, D. E., Nelson, P. C., and Phillips, R. (2006) Volume-exclusion effects in Tethered-Particle experiments: bead size matters.. *Phys Rev Lett,* **96**(8), 088306.

[30] Milstein, J. N., Chen, Y. F., and Meiners, J.-C. (2011) Bead size effects on protein-mediated DNA looping in tethered-particle motion experiments. *Biopolymers,* **95**, 144–150.