FLEXIBILITY AND CONTROL OF PROTEIN-DNA LOOPS

JASON D. KAHN, RAYMOND CHEONG[†], RUCHI A. MEHTA[†], LAURENCE M. EDELMAN[†], AND MICHAEL A. MORGAN[†]

Department of Chemistry and Biochemistry, University of Maryland
College Park, Maryland 20742-2021, United States of America
jdkahn@umd.edu

Protein-DNA loops are essential for efficient transcriptional repression and activation. The geometry and stability of the archetypal Lac repressor tetramer (LacI)-DNA loop were investigated using designed hyperstable loops containing *lac* operators bracketing a sequence-directed bend. Electrophoretic mobility shift assays, DNA cyclization, and bulk and single-molecule fluorescence resonance energy transfer (FRET) demonstrate that the DNA sequence controls whether the LacI-DNA loop forms a compact loop with positive writhe or an open loop with little writhe. Monte Carlo methods for simulation of DNA ring closure were extended to DNA loops, including treatment of variable protein hinge angles. The observed distribution of topoisomer products upon cyclization provides a strong constraint on possible models. The experiments and modeling imply that LacI-DNA can adopt a wide range of geometries but has a strong intrinsic preference for an open form. The flexibility of LacI helps explain *in vivo* observations that DNA looping is less sensitive to DNA length and shape than would be expected from the physical properties of DNA. While DNA cyclization suggests two pools of precursor loops for the 9C14 construct, single-molecule FRET demonstrates a single population. This discrepancy suggests that the LacI-DNA structure is strongly influenced by flanking DNA.

Keywords: Lac repressor, transcriptional regulation, DNA cyclization, DNA topology, single-molecule FRET, Monte Carlo simulation

1. Introduction

The regulation of gene expression often requires the integration of multiple inputs such as the availability of different nutrients, the activation of different signalling pathways from the cell membrane to the DNA, or the induction of stress responses. This integration occurs partially through protein-mediated DNA looping, whereby transcription factors bound at different sites can interact with each other or with common targets. Looping acts to enhance repression by increasing the local concentration of a DNA binding domain in the neighborhood of its binding site, via anchoring of the other end of the protein at a second site. More generally, looping can bring any two partners together, as in transcriptional activation by the *E. coli* NtrC protein or in the yeast two-hybrid system. Other proteins that affect DNA shape or flexibility can act at a distance in modulating gene expression by modulating DNA loop stability. DNA supercoiling and chromatin structure can have profound affects by affecting protein-DNA interaction directly and also by acting indirectly through their influence on looping and topology.

[†] Present Addresses: R.C., Dept. of Biomedical Engineering, Johns Hopkins Univ., Baltimore, MD 21218 USA; R.A.M., Lone Pine Capital, Greenwich, CT 06830 USA; L.M.E., Department of Emergency Medicine, Johns Hopkins Univ., Baltimore, MD 21287 USA; M.A.M., US Naval Research Lab, Washington, DC 20375, USA

Many studies of looping efficiency have assumed that DNA twist, shape, and stiffness control the local concentration of looping partners.[1-3] The effects of DNA length and shape on T4 ligase-mediated ring closure (cyclization) is modeled in the same way, as controlling the effective concentration of one properly-aligned and torsionally-phased end of the DNA in the neighborhood of the other.[4-6] This concentration is known as the *J* factor; for DNA molecules of 200-600 base pairs (bp) it ranges from ~100 pM to ~100 μM depending on the DNA shape and torsion. Although it is an oversimplification,[7] reasoning derived from elegant theory and experiment on cyclization has been applied to looping. For example, as seen for cyclization, looping is inefficient at short separations because loop formation requires bending a short DNA through a large angle, reaches a maximum as distance increases, and then decreases as the entropy of restraining the ends becomes more and more unfavorable.[8] Because the interacting partners presumably have a specific spatial relationship and bind a consistent face of the helix, superimposed on the overall length dependence there is torsional oscillation with a period equal to the helical repeat, which is ~11 in superhelical DNA because the contact area winds around the DNA.[2,9] DNA bending can assist in loop formation or break the loop depending on the bend direction relative to the direction of curvature needed to form the loop.[10,11]

In general, the overall trends observed for looping as a function of length, shape, and torsion often make qualitative sense. However, the effects are usually much weaker than one would predict from the in vitro physical chemistry of DNA. Looping efficiency in vivo can be used to derive values for the apparent persistence length *P* (a measure of bending stiffness) and torsional modulus *C*, and the derived values are usually much smaller (more flexible) than those determined in vitro. There are at least four possible, non-mutually exclusive explanations. First, the looping interactions can be multivalent, as for the NtrC-Eσ^{54} interaction.[12] NtrC is a hexamer, so in principle, six possible loops could be formed and the experiment would pick up the most stable. Second, non-specific DNA bending proteins such as HU, IHF, and H-NS in *E. coli* and histones and HMG proteins in eukaryotes induce random transient bends that result in enhanced apparent flexibility.[13,14] Third, the looping protein itself can be flexible or adopt different conformations, as observed for LacI and AraC. Fourth, it has been suggested that DNA may be much more flexible at short lengths than had previously been believed. This last idea is based on experiments that demonstrate surprisingly high *J* factors for cyclization of DNAs of < 100 bp, though the experiments are currently controversial.[15-17] Theoretical explanations of these results include transient denaturation bubbles or the idea that DNA bending has a double-well potential.[18,19] These ideas have in common that for large bending angles, once the large energy cost of disrupting stacking between adjacent base pairs has been incurred, the energy cost for further bending is small. There is evidence that such effects are sequence-dependent, being preferred for TATA box DNA.[20]

There are eukaryotic transcription activation systems where the expected dependencies of activation on DNA length and shape not observed at all.[14,21] This can be explained by taking the four reasons above to the limit of complete removal of the expected effect: interaction between large, conformationally flexible, multivalent

coactivators and large general transcription factors might be affected very little by DNA shape. Of course, activation could also be mediated through mechanisms other than looping (such as processive chromatin remodeling), or through very large DNA loops in eukaryotes,[22] such that their structures will be dominated by supercoiling or chromatin structure as opposed to local DNA structure. Also, we suggest that transcription activation is mechanistically less sensitive to DNA shape than repression: a 10-fold increase in the expression of a gene that is "on" 1% of the time could be due to only a 9% increase in the occupancy (from 1% to 10%), whereas 10-fold repression requires 90 % occupancy to block access by other factors.

One of the simplest looping interactions is the formation of DNA loops bridged by the tetrameric *E. coli* lactose repressor, the LacI protein. In the absence of inducer, LacI binds its primary operator sequence, at position +11 relative to the start site of transcription, and blocks transcription of the *lacZYA* operon. Secondary operators at -82 and +401 support DNA looping. Müller-Hill's group provided the clearest indication of their functional importance with deletion studies of the secondary operators, which reduced the efficiency of repression from 99.9+% to about 95%;[23] this appears to be a modest affect at first glance, but it reflects a 100-fold decrease in the repression ratio, the ratio of induced/repressed activity. Looping was confirmed as the cause through electron microscopy of loops and also from the dependence of repression efficiency on the distance between the primary and secondary operators.[8,24]

Crystallographic studies of the Lac repressor and the repressor-operator complex show that the repressor is a dimer of dimers.[25,26] The repressor forms a V shape, with the N-terminal headpieces that interact with DNA at the top, the dimeric core that binds inducer in the center, and a C-terminal 4-helix bundle at the vertex of the V. Deletion of the C terminus converts the repressor into a dimer that represses transcription much less effectively than the tetramer. The remainder of the tetramerization interface is made up of only a few contacts, although mutagenesis[27] as well as studies on the related Gal repressor suggests that they are in fact functionally relevant.[28] Steitz and coworkers suggested that LacI could change shape depending on loop size.[25] There are no high-resolution structures of the entire DNA loop, but the x-ray crystal structure of LacI bound to two separate operators was the basis for the well-known "Lewis model" of the LacI-DNA loop which shows the DNA "wrapping away" from the top of the V shape.[26] A DNA crossover in the tight loop is a positive writhe node, whereas LacI-DNA binding is stabilized by negative supercoiling.[29] The model also includes the CAP transcriptional activator, although on the wrong helical face of the DNA. Because of these issues, when we started work in the area we viewed the geometry of the loop as an open question.

Our focus has been to design hyperstable LacI-DNA loops with operator sequences bracketing sequence-directed bends, and then to use these molecules to probe loop structure and stability.[30-32] We made two constructs, denoted 9C14 and 11C12, that have different orientations of the lac operators with respect to sequence-directed curvature (Fig. 1). Initial mobility shift experiments showed that DNA bending confers hyperstability on the LacI DNA loop. The observed mobilities as well as DNA ring
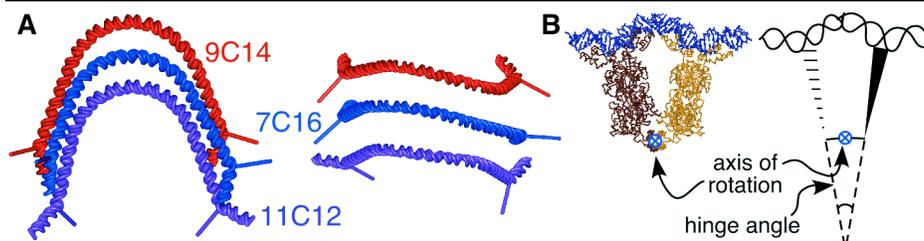
Fig. 1. Designed LacI-DNA looping constructs and the LacI hinging model. (*A*) Models of the equilibrium structures of the designed LacI looping constructs in the absence of LacI. The cylinders indicate the dyad axes of the *lac* operators, which are directed along the dyad axes of the LacI dimers shown at the right. When used, fluorophores are placed just adjacent to each operator on the internal side. (*B*) The x-ray co-crystal structure of LacI bound to two operator DNAs, from Ref. 26, and the hinge angle definition. The protein is a dimer of dimers, with each dimer binding one *lac* operator. The dimer on the right projects forward.

closure experiments suggested that operators directed away from the center of curvature (9C14) force the repressor to adopt a V-shape and give a tight positively supercoiled loop, whereas operators directed inward (11C12) give a more stable and open form requiring opening of a hinge at the repressor C-terminus.[31] These conclusions were verified by fluorescence resonance energy transfer (FRET), both in bulk[30] and on single 9C14 molecules.[32] Induction with IPTG changes the loop geometry[30] but does not release the protein entirely, in accord with *in vivo* studies on loops without DNA bending.[33]

This paper describes DNA cyclization studies and simulations on looped DNA, in the context of published experiments, all aimed at parsing the mutual influences of DNA shape and protein flexibility on the stability, geometry, and flexibility of the LacI-DNA loop. Previous experiments demonstrated different loop topologies, but the distributions were not interpreted quantitatively. We have performed additional quantitative cyclization experiments to measure *J* factors, we have characterized length variants of the 9C14 to differentiate twist and writhe effects, and we have applied the Monte Carlo simulation methods developed for DNA ring closure to the LacI-DNA loop.

2. Experimental Materials and Methods

DNA cyclization experiments were performed essentially as described previously. Body-labeled PCR products with BsaH I (GR|CGYC) or BssH II (G|CGCGC) ends were synthesized by PCR and restriction digestion from the 7C16, 9C14, 11C12, and Unbent templates.[31] BsaH I-ended variants of 9C14 that differed in length by –3 or +4 bp from the base sequence were synthesized using primers with internal insertions or deletions. Sequences of primers and cyclization substrates are given in the Appendix.

For ligation kinetics reactions, 1 nM DNA was incubated with 3-4 nM LacI for 30 minutes, followed by the addition of T4 DNA ligase (New England Biolabs). Aliquots from time courses were quenched and then analyzed on 6% polyacrylamide gels (75:1) containing 7.5 μg/ml chloroquine. Gels were quantitated and *J* factors determined essentially as described (graphs not shown).[20]

3. Simulation of the *J* factor for Cyclization of DNA Loops

In the Monte Carlo approach to modeling DNA cyclization, DNA chains are randomly generated using roll, tilt, and twist angles between base pairs that are drawn from Gaussian distributions with widths specified by the persistence length and torsional modulus. Each chain is formed by joining two half-chains from separate distributions to allow production of large numbers of chains (10^{10} chains from 2×10^5 half-chains are needed to adequately simulate $J \leq 1$ nM). The chains are then filtered for adherence to the boundary conditions for cyclization, which are co-linear helix axes and correct torsional phasing of ends.

Extension of this method to DNA looping is straightforward but computationally demanding; recent analytical[7] or rod mechanics[34-36] models are much more efficient, and atomic-level molecular dynamics simulations suggest that LacI flexibility is more complex than in our representation.[37] The Monte Carlo method has the advantages that consideration of sequence-dependent bending and flexibility is straightforward, and the entropy of cyclization is intrinsically considered. The joining of half-chains is done through a virtual bond that places the origins of the chains at positions of the DNA operators in a LacI-DNA complex represented according to the crystal structure.[26] Within each simulation, the protein geometry is considered to be fixed. To vary the LacI hinge angle in different simulations, the LacI dimers are rotated relative to each other about an axis through the C-terminal four-helix bundle, as shown in Fig. 1. Simulations of hinge angles from 30° to 180° in 15° increments were performed. Filtering for ring closure gives a model for the protein-DNA loop, referred to as the inner loop, with *J* factor given by J_{inner}.

For each protein geometry (i.e. hinge angle), there are four possible loop types that are considered individually, as shown in Fig. 2. There are two "parallel" loops, one a compact form with the looped DNA within the acute angle formed by LacI dimers and the other with the looped DNA on the outside.[38] Two antiparallel loops are distinguishable for a non-symmetric loop sequence. The relative free energy of loop formation is related to the *J* factor as follows:

$$\Delta G^0_{Loop\ 1} - \Delta G^0_{Loop\ 2} = -RT \ln\left(J_{inner,\ 1}/J_{inner,\ 2}\right). \tag{1}$$

Ring closure (cyclization) of the looped DNA is then modeled similarly to the inner loop, by anchoring half-chains at the operators but orienting them in the opposite direction. Closure gives a model of the outer loop which is then combined with the inner loop to make a complete LacI-DNA loop in a minicircle, with a *J* factor J_{outer} that would be the experimentally observed J factor for a pure population of the specified loop type and angle. This approach partitions the total free energy into the bending energy of the inner loop, the bending energy of the outer loop, and the bending energy of the protein without considering possible coupling terms. One might expect the cyclized loop to adopt a mixture of states specified by a partition function that depends on the total free energy of all the states, but the cyclization experiment is performed on pre-incubated loops, and under the conditions of the experiment equilibration among loop forms is very slow.
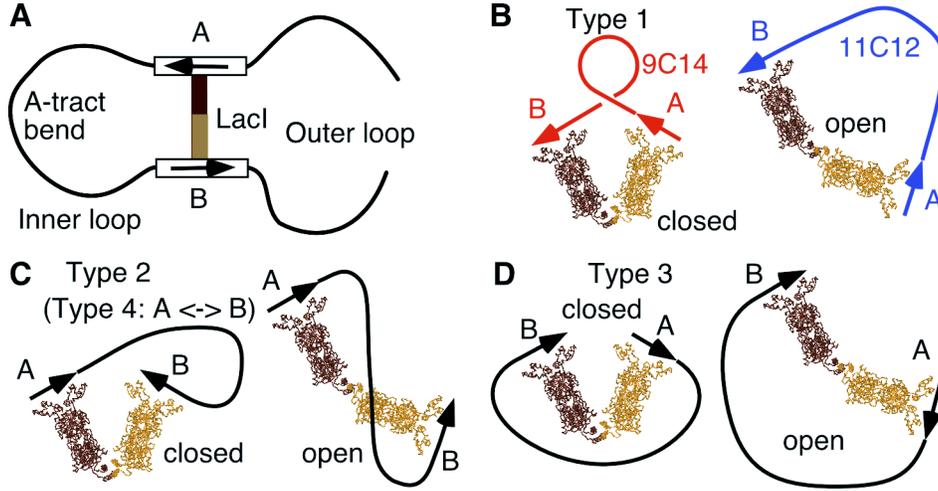
Fig. 2: The four loop topologies. For each loop type, the hinge angle can vary; closed (45°) and open (135°) forms are shown. The linear DNA contains two symmetric operators A and B that can bind LacI in either orientation. (*A*) The inner and outer loop approach to modeling ring closure of DNA loops. (*B*) The Type 1 parallel loop. (*C*) The Type 2 antiparallel loop. Type 4 is similar, with the operators switched. The overall molecules are not symmetric because the intervening sequence is directional, so Types 2 and 4 are not identical. (*D*) The Type 3 parallel loop.

Therefore, we assume that the loop type distribution is determined by the Boltzmann factor only for the inner loop, which is J_{inner}, the equilibrium constant for loop formation, multiplied by a factor from the deformation energy of the LacI-DNA complex, denoted P_{LacI}. The experimental J is the weighted average of J_{outer} over the fixed loop type and hinge angle θ distribution, as in Eq. (2). Other observables are computed analogously.

$$\langle J \rangle_{obs} = \frac{\sum_{type} \sum_{\theta} J_{inner}(type,\theta) P_{LacI}(\theta) J_{outer}(type,\theta)}{\sum_{type} \sum_{\theta} J_{inner}(type,\theta) P_{LacI}(\theta)}. \tag{2}$$

To extract topoisomer distributions, the final torsion angle probability distribution is fit to a Gaussian as described.[5] The corresponding Gaussians for length variants are obtained simply by shifting this curve. The total twist Tw is calculated from the sum of individual twist angles as in Eq. 3, and the writhe Wr is calculated from the helix axis coordinates for each base pair from Eq. 4:

$$Tw = \frac{1}{2\pi}\left[\sum_{i \in inner} \tau_i + \sum_{j \in outer} \tau_j\right] + \frac{2}{hr} \quad \text{and} \tag{3}$$

$$Wr = \frac{1}{2\pi}\sum_{k=1}^{N}\sum_{l=k+1}^{N} \frac{\left[(\mathbf{r}_k - \mathbf{r}_{k-1}) \times (\mathbf{r}_l - \mathbf{r}_{l-1})\right] \cdot \left[\mathbf{r}_k - \mathbf{r}_l\right]}{\left\|\mathbf{r}_k - \mathbf{r}_l\right\|^3}, \tag{4}$$

where τ_i is the twist angle for dinucleotide i, hr is the helical repeat, $\mathbf{r}_k$ is the position of the kth base pair, and N is the total number of base pairs. The linking number Lk, the sum of twist and writhe, is also described by a continuous distribution that is also fit to a Gaussian, and the relative probability of forming each topoisomer is calculated from the values of the Lk distribution at the integral values that must describe a physical closed circular DNA. Chains with crossover distance less than 25Å were removed from the final ensemble before analysis.

Sequence-dependent DNA bending was described by a scaled wedge angle model.[39,40] The DNA helical repeat (hr), torsional modulus (C), and persistence length (P) were varied between hr = 10.4-10.6 bp/turn, C = 1.5-3.0×10^{-19} erg·cm, and P = 463-550 Å (136-162 bp). Experimental J factors were in the 1-10 nM range, and to bring the simulated J factors down to about the same range (10-30 nM), the values were set to hr = 10.40 bp/turn, C = 2.0×10^{-19} erg·cm, and P = 550Å (162 bp). The helical repeat of the outer loop of the BsaH I molecules was estimated at hr = 10.60 bp/turn in order to improve the fit with 9C14 length variants.

Each set of flexibility parameters required 88 individual simulations corresponding to combinations of inner and outer loop, hinge angles, and loop types. Eight molecules were simulated, but the total number of simulations required was reduced because molecules with the same sticky ends (BssH II or BsaH I) have the same outer loop for any inner loop at a given hinge angle, inner loop structure (for 7C16, 9C14, or 11C12) is the same for any outer loop, and length variant results were derived from the parent 9C14. In total, 264 simulations per set were needed for all eight molecules for each set of flexibility parameters.

4. Results

The looping constructs 9C14 and 11C12 form hyperstable loops, which makes it possible to study a population that is entirely in a looped configuration; the wild-type loop is much less stable.[41] Electrophoretic mobility shift assays had shown that the 9C14-LacI and 11C12-LacI loops have markedly different mobilities, and DNA cyclization studies on extended constructs showed that 9C14 provided a positively supercoiled product whereas 11C12 did not.[31] Subsequent FRET experiments showed that energy transfer is very efficient in the 9C14 loop and weak in 11C12.[30] All of this provides solid evidence for the existence of at least two very different loop shapes. There was also evidence for a mixed population of 9C14 from DNA cyclization reactions: even though the $\Delta Lk = 0$ topoisomer appears much faster than the $\Delta Lk = +1$ topoisomer, the latter continues to accumulate after the former has saturated, suggesting separate reactant pools.

We turned to more quantitative DNA cyclization experiments accompanied by Monte Carlo simulation in order to test these ideas on loop variability and interconversion. Quantitative J factors were not be obtained in our earlier experiments, on molecules with very stable BssH II cohesive ends, because the ring closure reaction was too fast under the conditions used. We believe, however, that the distribution of product topoisomers is still informative. Less stable BsaH I ends were then used to decrease

ligation rates and allow determination of *J* factors for cyclization, and we also synthesized length variants of 9C14 to investigate whether positive supercoiling arises from changes in writhe or in twist.

Monte Carlo simulation methods that were previously used for analysis of DNA cyclization[5] were adapted to the problem of cyclization of DNA with embedded loops as described in Section 3. The results described here are based on flexibility parameters for DNA and for the LacI-DNA complex that give the best match to experiment, although since parameters are optimized one by one we cannot be sure that we have found the globally optimum set.

Some of the cyclization results are shown in Fig. 3 below. It is clear that Lac repressor binding alters the distribution of topoisomers, as observed previously, and it also changes the overall *J* factors. The +1 topoisomer is observed for the BsaH I - 9C14(-3) and 9C14(+4) molecules, but for all of the other molecules LacI binding induces a small negative ΔLk. The writhe component of the observed ΔLk is a combination of contributions from the protein-DNA (inner) loop and the outer loop closed by ligase, and the twist is similarly affected by the different twist changes in each lobe. This makes qualitative interpretation difficult, but the Monte Carlo simulations can help identify the putative species and processes contributing to the distribution of cyclized products.
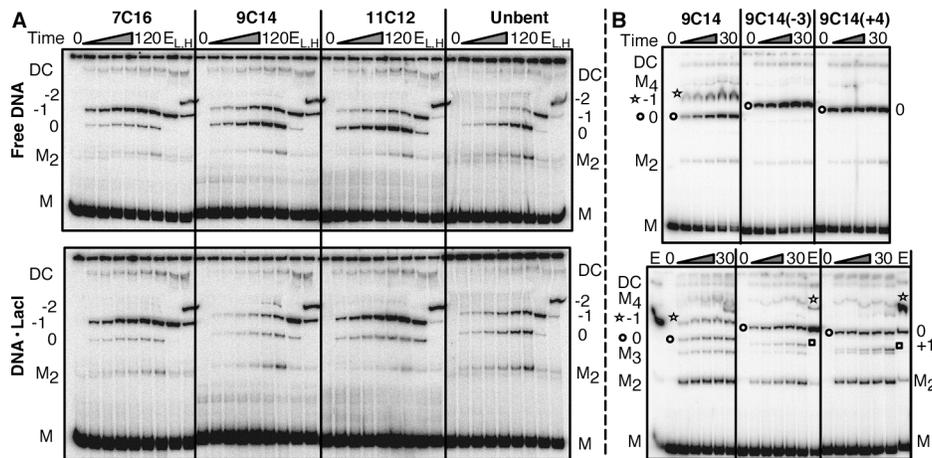

Fig. 3: Cyclization kinetics experiments. The labels DC, -2, -1, 0, M_2, and M denote dimer circle, $\Delta Lk = -2$ topoisomer, $\Delta Lk = -1$ topoisomer, relaxed topoisomer, bimolecular products, and monomer DNA, respectively. (*A*) LacI appears to reduce the overall extent of cyclization and to introduce slight unwinding, apparent as a shift from $\Delta Lk = 0$ to $\Delta Lk = -1$ topoisomers. Ligation kinetics experiments were performed using 1 nM body-labeled BsaH I-ended DNA molecules and 250 U/ml T4 DNA ligase, in the absence (*top panel*) or presence (*bottom panel*) of 4 nM LacI, at 21°C. Aliquots were quenched at 1, 2, 4, 10, 30, and 120 minutes. A control reaction without ligase (the 0 min time point) and two $\Delta Lk = -1$ topoisomer controls (lane E_L and E_H, from ligation in the presence of 0.1 and 0.3 μg/ml ethidium bromide, respectively) are also shown. (*B*) Ligation kinetics of BsaH I-ended 9C14 length variants shows LacI-induced formation of $\Delta Lk = +1$ topoisomers for some DNA lengths. Experiments were performed as in A except with 1.66 nM labeled DNA, and 0.5, 1, 2, 10 and 30 min time points. The $\Delta Lk = -1$ topoisomer control lane E was a 30 min ligation in the presence of 0.15 mg/ml ethidium bromide. The labels M_4 and M_3 denote linear tetramer and trimer. The $\Delta Lk = +1$ topoisomer, $\Delta Lk = -1$ topoisomer and relaxed (0) topoisomer are denoted with open square, star and circle, respectively.

The experiments in Fig. 3(B) show that much of the monomeric DNA is either non-ligatable or very slowly ligatable, presumably due to dephosphorylated or incorrectly restricted ends. Quantitative measurement of the J factor requires measurement of both the bimolecular ligation rate constant and the unimolecular cyclization rate constant, and this is difficult when there are too many uncharacterized species in the reaction. J tends to be underestimated in these cases because many molecules that cannot cyclize are still competent for bimolecular ligation. Topoisomers, however, can only be formed by cyclizable species, so the ratios between them should not be affected by the slow/dead fraction and are therefore much more accessible and reliable. Table 1 summarizes the quantitative results available from experiments like that of Fig. 3.

Table 1. Experimental and simulated *J* factors for different loop constructs.

	Experiment				Simulation			
	BsaH I; -LacI		BsaH I; +LacI		BsaH I; +LacI		BssH II; +LacI	
DNA	$\Delta Lk = 0$	$\Delta Lk = -1$	$\Delta Lk = 0$	$\Delta Lk = -1$	$P_{LacI} = 1$	$P_{LacI} = F_o$	$P_{LacI} = 1$	$P_{LacI} = F_o$
7C16	0.8	7.0	1.2	6.6	53.8	48.8	32.8	15.0
9C14	5.5	3.1	0.4	6.1	84.6	37.6	7.2	10.5
11C12	6.6	1.4	1.4	2.1	16.6	27.1	19.6	20.7
Unbent	1.0	0.2	0.8	1.0	27.3	51.8	14.6	22.3

All *J* factors are in nM. Experiments are accurate to within a factor of 3, with relative *J* factors being more reliable. The Simulation columns give the total *J* factor for all topoisomers; see Fig. 4 for distributions. The restriction enzyme ends are specified. Quantitative *J* factors could not be obtained for BssH II-ended molecules. P_{LacI} is the factor contributed by protein deformation, either a flat potential ($P_{LacI} = 1$) or one that favors LacI opening, $P_{LacI} = F_o = \exp(0.078 \times (\theta\text{-}28°))$, corresponding to $\Delta G° = -0.046$ kcal/mole per degree of LacI opening relative to the crystal structure hinge angle.

The results show that the simulations can capture the surprising observation that the *J* factors for these loops are generally similar to each other and also much smaller than those observed for cyclization of other bent molecules. However, the simulations to date do not agree well with the absolute magnitude of the observed *J* factors. The inclusion of the $P_{LacI} = F_o$ factor (as described in the legend to Table 1), improves the agreement with experiment, and its effect is much more obvious when we consider the predicted vs. experimental topoisomer distributions shown in Fig. 4 below. It is clear that rationalizing the experiment requires the LacI-DNA sandwich complex[42] (LacI bound to two DNA molecules, without a loop) to have a strong preference for a form much more open than the x-ray crystal structure. Much more complex and realistic forms for F_o could be used,[37] but we do not have enough data to discriminate among them.

The inclusion of the F_o weighting factor also helps rationalize the experimental results of Fig. 3(*B*), as shown in Table 2 below. The observed topoisomer distributions would be explained by a mixture of closed and open form loops, and the J_{outer} factor for the closed form is much smaller than that of the open form, as observed. The closed form induces $\Delta Lk \sim +0.6$ and the open form induces $\Delta Lk \sim -0.2$, in accord with our earlier models[31] assuming a small local untwisting ($\Delta Tw \sim -0.2$) from the LacI binding. The simulations do predict that the $\Delta Lk = +1$ topoisomer should be observed at low levels for 9C14, and we do not know why it was not seen in the experiments.
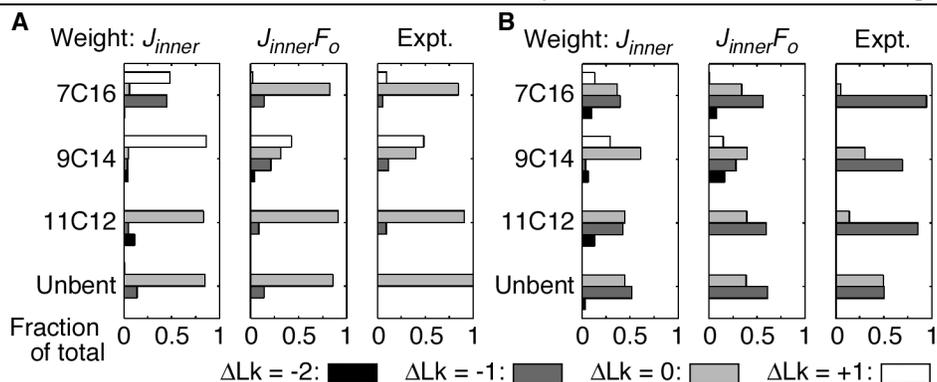
Fig. 4: Experimental vs. theoretical topoisomer distributions. The fraction of ΔLk = -2, -1, 0, and +1 topoisomers (*horizontal axes*) obtained upon ring-closure of BssH II-ended (*A*) and BsaH I-ended (*B*) molecules are shown. Simulation results comparing weighting functions $J_{inner}P_{LacI}$ with P_{LacI} = 1 (*left graph in each panel*) and $P_{LacI} = F_o$ (*middle graph*) are compared to experimental results (*right graph in each panel*). Data for BssH II-ended molecules is based on Fig. 6 of Ref. [31] and data for BsaH I-ended molecules is from Fig. 3(A) of this study. J_{inner} reflects the deformation free energy of the DNA loop and F_o reflects the deformation free energy of the LacI-DNA interaction and LacI conformation change. F_o is given in Table 1.

Fig. 5 below shows simulated Type 1 open form, Type 1 closed form, and Type 2 loops. The LacI-9C14 loop in Fig. 5(*B*) has a very small radius of curvature relative to the other two, and it would be very high in energy in the absence of intrinsic bending. The limaçon shape is due to DNA bending away from the LacI protein. Bending toward the protein would give a figure 8, and all the results would be essentially unchanged.

Table 2. 9C14 (BsaH I ends) length variant topoisomer distribution results.

Molecule (Est. twist)	Lk at ΔLk=0	Exptl. +LacI topoisomers	Sim. Type 1 closed form		Sim. Type 1 open form	
			Topoisomers	J_{outer}	Topoisomers	J_{outer}
9C14 (35.6 turns)	36	ΔLk = 0 ~ -1 Lk = 35, 36	64% Lk = 36 28% Lk = 37	0.2–0.8	56% Lk = 35 37% Lk = 36	12–60
9C14 (+4) (36.0 turns)	36	ΔLk = 0 > +1 Lk = 36, 37	88% Lk = 37	0.02–0.1	79% Lk = 36	21–80
9C14 (-3) (35.3 turns)	35	ΔLk = 0 > +1 Lk = 35, 36	92% Lk = 36	0.5–0.6	80% Lk = 35	19–58

The experimental observations are from Fig. 3(B). Simulated yields of topoisomers and the predicted J_{outer} factors (in nM) are for the major cyclized products (> 15%) of closed form (θ = 30°-75°) and open form (θ = 135°-180°). They are averages over θ based on $P_{LacI} = F_o$. The range given is for a range of flexibility parameters. Percentages do not add up to 100% due to minor topoisomers. The assignments of helical turns for each variant are a set of consistent estimates based on the topoisomer distributions.

In summary, the cyclization and simulation results confirm that phased DNA bending sequences can be used to both probe and control the geometry of LacI-DNA loops. We show that ring closure of a DNA containing a protein-DNA loop can be simulated using the same methods previously used for free DNA cyclization. Absolute J factors were modeled less successfully than topoisomer distributions, probably in part because the experimental data for the former are not as accurate. The P_{LacI} dependence
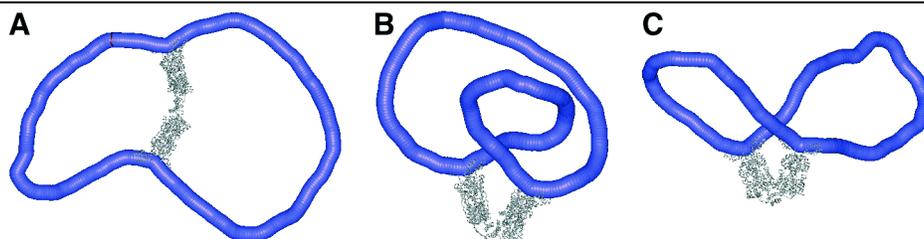
Fig. 5: Representative simulated LacI-DNA loops. The inner loop is the shorter segment. For the BsaH I molecules, total lengths are about 360 bp. (*A*) The open form Type 1 loop is 11C12, with a 135° hinge angle. (*B*) The closed form Type 1 wrapping away loop is 9C14, with a 45° hinge angle. The positive writhe is apparent. (*C*) The antiparallel Type 2 loop is 7C16, 30° hinge angle.

needed for the simulations to match experiment corresponds to the LacI-DNA sandwich complex having a strong intrinsic preference for an open form geometry. This might be due to electrostatic or steric repulsion between the DNA around the operators. The result rationalizes the ability of the 9C14 molecule to form two different loops: the closed form minimizes DNA bending and twisting free energy, whereas the open form minimizes LacI-DNA sandwich complex deformation free energy. The 11C12 molecule is more stable because the open form minimizes both components of the total free energy.

5. Discussion

Our initial approach to looping geometry was guided by the idea that the DNA that is most easily deformed to match the geometry imposed by a looping protein should form the most stable loop. We set out to study the low-resolution structure of the LacI-DNA loop by isolating stable complexes, through a binding selection-amplification protocol that used a semi-random DNA bending library synthesized essentially as described.[43] The library proved intractable, so we designed constructs with large intrinsic bends and operators oriented either outward (construct 9C14, Fig. 1) or inward (11C12) with respect to the center of curvature. To our surprise, hyperstable loops were observed for both constructs,[31] and there is no reason to believe that these are the most stable possible loops. The designed loops, however, are much more stable than natural loops, so it is also clear that natural loops have not evolved to maximal stability but rather to (presumably) optimal stability, perhaps in order to reduce fluctuations in gene expression.[44]

There is abundant evidence from topology and FRET[30] for at least two loop forms. All of our work is consistent with these being mainly the Type 1 closed form for 9C14 and the Type 1 open form for 11C12, as shown in Figure 5. For longer DNA the antiparallel loops are probably more stable, depending on length.[45] Loops formed from unbent DNA have electrophoretic mobilities and cyclization topologies similar to those of 11C12, so we believe that an open form loop is the best model for unbiased loops.

Our solution biochemistry and the simulations described above suggest that 9C14 can form two different shapes, and bulk FRET was interpreted as being consistent with this conclusion.[30] However, single-molecule FRET suggested that 9C14 adopts a single

closed-form conformation with extremely efficient energy transfer between operators.[32] While this result confirms the Type 1 closed form loop for 9C14, it does not agree with the idea of two forms. There are several possible resolutions. The addition of fluorophores to the DNA could alter the shape; we find this doubtful because the binding and electrophoretic properties of the labeled and unlabeled loops are similar. The loop could be deformed or released in the ligation experiments, although again there is no precedent for this. The cyclization experiment is, however, extremely sensitive to rare species if those species cyclize rapidly.[15,46] More interestingly, the origin of the discrepancy could lie in long-range effects of the loop tails. LacI requires about 40 bp of operator DNA to bind with full affinity, and this, along with extensive studies of the salt dependence of looping, has led Record and coworkers to propose that nonoperator DNA may wrap around the surface of the protein.[47] If wrapping occurs, it would predict different behavior for our cyclization constructs, with long tails outside the loop, and the fluorescent constructs, which have only enough DNA outside the loop to provide tight binding. These possibilities could be distinguished by experiments using fluorescent DNAs with long tails.

As a whole, this work shows the LacI protein is flexible, which we propose may explain why looping in vivo is less sensitive to shape than we would expect from the stiffness of DNA. In general, protein-DNA loops seem to have evolved to be flexible and dynamic, perhaps because some of them must be stable under a range of ionic conditions or with different nonspecific bending proteins bound to the loop DNA. Our results have not provided evidence in favor of sharp spontaneous bends, but they may also contribute.

Loop geometry can be controlled by manipulating DNA shape and sequence, and we can engineer them to be very stable. Exploration of a wider range of sequences might provide optimized sequences that could be locked into single conformations or else might be on the cusp between two forms, where the dynamics of interconversion would become interesting. The combination of new single-molecule methods, efficient modeling algorithms, and renewed interest in DNA physics in vivo[44,48] should make such molecules generally useful and interesting.

Acknowledgements

6. References

1.   G. R. Bellomy and M. T. Record, Jr., *Prog. Nucleic Acid Res. Mol. Biol.* **39**, 81 (1990).
2.   D. H. Lee and R. F. Schleif, *Proc. Natl. Acad. Sci. USA* **86**, 476 (1989).
3.   K. Rippe, P. H. von Hippel and J. Langowski, *Trends Biochem. Sci.* **20**, 500 (1995).
4.   H.-S. Koo, J. Drak, J. A. Rice and D. M. Crothers, *Biochemistry* **29**, 4227 (1990).

5.　J. D. Kahn and D. M. Crothers, *J. Mol. Biol.* **276**, 287 (1998).
6.　S. D. Levene and D. M. Crothers, *J. Mol. Biol.* **189**, 61 (1986).
7.　Y. Zhang, A. E. McEwen, D. M. Crothers and S. D. Levene, *Biophys. J.* **90**, 1903 (2006).
8.　J. Müller, S. Oehler and B. Müller-Hill, *J. Mol. Biol.* **257**, 21 (1996).
9.　G. R. Bellomy, M. C. Mossing and M. T. Record, Jr., *Biochemistry* **27**, 3900 (1988).
10.　M. Carmona and B. Magasanik, *J. Mol. Biol.* **261**, 348 (1996).
11.　R. B. Lobell and R. F. Schleif, *J. Mol. Biol.* **218**, 45 (1991).
12.　I. Rombel, A. North, I. Hwang, C. Wyman and S. Kustu, *Cold Spring Harbor Symp. Quant. Biol.* **63**, 157 (1998).
13.　Y. Hodges-Garcia, P. J. Hagerman and D. E. Pettijohn, *J. Biol. Chem.* **264**, 14621 (1989).
14.　E. D. Ross, P. R. Hardwidge and L. J. Maher, 3rd, *Mol. Cell. Biol.* **21**, 6598 (2001).
15.　T. E. Cloutier and J. Widom, *Mol. Cell* **14**, 355 (2004).
16.　T. E. Cloutier and J. Widom, *Proc. Natl. Acad. Sci. USA* **102**, 3645 (2005).
17.　Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia and A. Vologodskii, *Proc. Natl. Acad. Sci. USA* **102**, 5397 (2005).
18.　J. Yan and J. F. Marko, *Phys. Rev. Lett.* **93**, 108108 (2004).
19.　P. A. Wiggins, R. Phillips and P. C. Nelson, *Phys. Rev. E* **71**, 021909 (2005).
20.　N. A. Davis, S. S. Majee and J. D. Kahn, *J. Mol. Biol.* **291**, 249 (1999).
21.　E. D. Ross, A. M. Keating and L. J. Maher, 3rd, *J. Mol. Biol.* **297**, 321 (2000).
22.　B. Tolhuis, R. J. Palstra, E. Splinter, F. Grosveld and W. de Laat, *Mol. Cell* **10**, 1453 (2002).
23.　S. Oehler, E. R. Eismann, H. Krämer and B. Müller-Hill, *EMBO J.* **9**, 973 (1990).
24.　H. Krämer, M. Niemöller, M. Amouyal, B. Revet, B. von Wilcken-Bergmann and B. Müller-Hill, *EMBO J.* **6**, 1481 (1987).
25.　A. M. Friedman, T. O. Fischmann and T. A. Steitz, *Science* **268**, 1721 (1995).
26.　M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan and P. Lu, *Science* **271**, 1247 (1996).
27.　J. Chen, R. Surendran, J. C. Lee and K. S. Matthews, *Biochemistry* **33**, 1234 (1994).
28.　S. Semsey, M. Geanacopoulos, D. E. Lewis and S. Adhya, *EMBO J.* **21**, 4349 (2002).
29.　P. A. Whitson, W.-T. Hsieh, R. D. Wells and K. S. Matthews, *J. Biol. Chem.* **262**, 14592 (1987).
30.　L. M. Edelman, R. Cheong and J. D. Kahn, *Biophys. J.* **84**, 1131 (2003).
31.　R. A. Mehta and J. D. Kahn, *J. Mol. Biol.* **294**, 67 (1999).
32.　M. A. Morgan, K. Okamoto, J. D. Kahn and D. S. English, *Biophys. J.* **89**, 2588 (2005).
33.　N. A. Becker, J. D. Kahn and L. J. Maher, 3rd, *J. Mol. Biol.* **349**, 716 (2005).
34.　D. Swigon, B. D. Coleman and W. K. Olson, *Proc. Natl. Acad. Sci. USA* **103**, 9879 (2006).
35.　S. Goyal, N. C. Perkins and C. L. Lee, *J. Comp. Phys.* **209**, 371 (2005).
36.　A. Balaeff, L. Mahadevan and K. Schulten, *Phys. Rev. E* **73**, 031919 (2006).
37.　E. Villa, A. Balaeff and K. Schulten, *Proc. Natl. Acad. Sci. USA* **102**, 6783 (2005).
38.　S. Semsey, K. Virnik and S. Adhya, *Trends Biochem. Sci.* **30**, 334 (2005).
39.　T. E. Haran, J. D. Kahn and D. M. Crothers, *J. Mol. Biol.* **244**, 135 (1994).
40.　R. S. Manning, J. H. Maddocks and J. D. Kahn, *J. Chem. Phys.* **105**, 5626 (1996).
41.　M. Brenowitz, A. Pickar and E. Jamison, *Biochemistry* **30**, 5986 (1991).
42.　R. Fickert and B. Müller-Hill, *J. Mol. Biol.* **226**, 59 (1992).
43.　A. E. Lilja, J. R. Jenssen and J. D. Kahn, *J. Mol. Biol.* **342**, 467 (2004).
44.　J. M. Vilar and S. Leibler, *J. Mol. Biol.* **331**, 981 (2003).
45.　J. Yan, R. Kawamura and J. F. Marko, *Phys. Rev. E* **71**, 061905 (2005).

46. P. R. Hardwidge, J. D. Kahn and L. J. Maher, 3rd, *Biochemistry* **41**, 8277 (2002).

47. O. V. Tsodikov, R. M. Saecker, S. E. Melcher, M. M. Levandoski, D. E. Frank, M. W. Capp and M. T. Record, Jr., *J. Mol. Biol.* **294**, 639 (1999).

48. L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman and R. Phillips, *Curr. Opin. Genet. Dev.* **15**, 125 (2005).

Appendix: Kahn et al., Biophysical Reviews and Letters, 2006

Sequences of cyclization constructs for LacI loops.

Experiments by Ruchi Mehta, simulations by Raymond Cheong

```
BsaH I-ended 7C16
     CGCCAAAGCTGGGTACCGATATCTGCAGGTCAGTCTAGGTAATTGTGAGC
     GCTCACAATTAGATCTCAATTCACGGATCCGGTTTTTTGCCCGTTTTTTG
     CCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTT
     TGCGCTGACAACGCGTCCTAGAATCGAAGCTAGCTAATTGTGAGCGCTCA
     CAATTCGTTGTGGTAAAGCTTTGATATCAAGCTTATCGATACCGTCGACC
     TCGAGGGGGGGCCGCCACCGCGGTGGAGCTCCAATTCGCCCTATAGTGAG
     TCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGA
     AAACCCTGGCG

BsaH I-ended 9C14
     CGCCAAAGCTGGGTACCGATATCTGCAGGTCAGTCTAGGTAATTGTGAGC
     GCTCACAATTAGATCTCAATTCGTACGGATCCGGTTTTTTGCCCGTTTTT
     TGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTT
     TTTGCCCGTTTTTTGCGCTGAACGCGTCCTAGAATCGAAGCTAGCTAATT
     GTGAGCGCTCACAATTCGTTGTGGTAAAGCTTTGATATCAAGCTTATCGA
     TACCGTCGACCTCGAGGGGGGGCCGCCACCGCGGTGGAGCTCCAATTCGC
     CCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGT
     CGTGACTGGGAAAACCCTGGCG

BsaH I-ended 11C12
     CGCCAAAGCTGGGTACCGATATCTGCAGGTCAGTCTAGGTAATTGTGAGC
     GCTCACAATTAGATCTCAATTCCTGTACGGATCCGCAAAAAACGGGCAAA
     AAACGGCAAAAACGGGCAAAAAACGGCAAAAAACGGGCAAAAAACGGCAA
     AAAACGGGCAAAAAACCGCTACGCGTCCTAGAATCGAAGCTAGCTAATTG
     TGAGCGCTCACAATTCGTTGTGGTAAAGCTTTGATATCAAGCTTATCGAT
     ACCGTCGACCTCGAGGGGGGGCCGCCACCGCGGTGGAGCTCCAATTCGCC
     CTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTC
     GTGACTGGGAAAACCCTGGCG

BsaH I-ended Unbent
     CGCCAAAGCTGGGTACCGATATCTGCAGGTCAGTCTAGGTAATTGTGAGC
     GCTCACAATTAGATCTCAATTCCTGTACGGATCCACTGAATCCGGTGAGA
     ATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCA
     TTACGCTCGTCATCAAAATCACTACGCGTCCTAGAATCGAAGCTAGCTAA
     TTGTGAGCGCTCACAATTCGTTGTGGTAAAGCTTTGATATCAAGCTTATC
     GATACCGTCGACCTCGAGGGGGGGCCGCCACCGCGGTGGAGCTCCAATTC
     GCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAAC
     GTCGTGACTGGGAAAACCCTGGCG
```

```
Primers used for BsaH I-ended molecules:
BsaH I top Primer:
     5'-GACAGACTAGGCGCCAAAGCTGGGTACCGATATC-3'
BsaH I (-3) top Primer:
     5'-GATTGACTAGGCGCCGCTGGGTACCGATATCTGC-3'
BsaH I (+4) top Primer:
     5'-GACAGAATAGGCGCCAGTCAAAGCTGGGTACCGATATC-3'
M13/pUC (-47) Nco I primer (bottom primer)
     5'-GCTGCCATGGCGCCAGGGTTTTCCCAGTCACGAC-3'


BssH II-ended 7C16
     GCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGGTACCGATATCT
     GCAGGTCAGTCTAGGTAATTGTGAGCGCTCACAATTAGATCTCAATTCAC
     GGATCCGGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCGTTTTTTGCCCG
     TTTTTTGCCGTTTTTTGCCCGTTTTTTGCGCTGACAACGCGTCCTAGAAT
     CGAAGCTAGCTAATTGTGAGCGCTCACAATTCGTTGTGGTAAAGCTTTGA
     TATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCGCCACCGCGGT
     GGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGC


BssH II-ended 9C14
     GCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGGTACCGATATCT
     GCAGGTCAGTCTAGGTAATTGTGAGCGCTCACAATTAGATCTCAATTCGT
     ACGGATCCGGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGC
     CGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCGCTGAACG
     CGTCCTAGAATCGAAGCTAGCTAATTGTGAGCGCTCACAATTCGTTGTGG
     TAAAGCTTTGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCC
     GCCACCGCGGTGGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACGCGC
     GC


BssH II-ended 11C12
     GCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGGTACCGATATCT
     GCAGGTCAGTCTAGGTAATTGTGAGCGCTCACAATTAGATCTCAATTCCT
     GTACGGATCCGCAAAAAACGGGCAAAAAACGGCAAAAACGGGCAAAAAAC
     GGCAAAAAACGGGCAAAAAACGGCAAAAAACGGGCAAAAAACCGCTACGC
     GTCCTAGAATCGAAGCTAGCTAATTGTGAGCGCTCACAATTCGTTGTGGT
     AAAGCTTTGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCG
     CCACCGCGGTGGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACGCGCG
     C


BssH II-ended Unbent
     GCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGGTACCGATATCT
     GCAGGTCAGTCTAGGTAATTGTGAGCGCTCACAATTAGATCTCAATTCCT
     GTACGGATCCACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTT
     TCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTA
     CGCGTCCTAGAATCGAAGCTAGCTAATTGTGAGCGCTCACAATTCGTTGT
     GGTAAAGCTTTGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGG
     CCGCCACCGCGGTGGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACGC
     GCGC
```