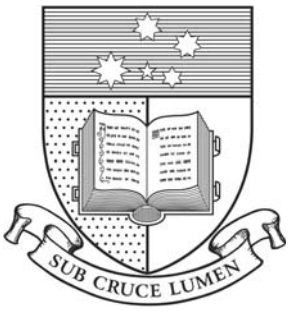


The University of Adelaide
School of Molecular & Biomedical Science
Discipline of Biochemistry

Testing the DNA loop domain model in *Escherichia coli*

A thesis submitted for the degree of Doctor of Philosophy

David Priest

Supervisors: Keith Shearwin & Ian Dodd

Thursday 4th September, 2014

Summary

The ability of DNA to form loops has been employed by evolution in almost every aspect of biology involving DNA, not least the regulation of gene transcription. The biophysical constraints on looping of the DNA polymer at short range (< 300 bp) have been extensively studied, however it is uncertain how the probability of DNA looping decays at longer range. The first part of this thesis presents a quantitative investigation of long range DNA looping both *in vivo* in *E. coli* and *in vitro*. DNA looping is more efficient *in vivo* than measured *in vitro* (by our collaborators) with the technique of Tethered Particle Motion (TPM), and we suggest that DNA supercoiling aids DNA looping *in vivo*. By measuring long-range looping *in vivo* using the two well-characterised looping proteins (the LacI and λCI repressors) and thermodynamic models of DNA looping, the decay in looping probability is quantified over the range 242–10000 bp. Furthermore this decay is shown to be a property of the DNA tether linking the loop, independent of the nature of the DNA looping protein(s).

Enhancers activate genes at long distance irrespective of position and orientation, so why don't enhancers activate the wrong genes? In other words, what mechanisms drive efficiency and specificity in enhancer-promoter looping? The loop domain model proposes that DNA loops formed by insulators pose a topological barrier that restricts the reach of enhancers to the vicinity of desired target promoter(s). Specifically, the model predicts that two DNA loops in an alternating arrangement should form somewhat mutually exclusively (i.e. they should interfere with one another's formation), whereas nested DNA loops are predicted to assist one another's formation, and side-by-side loops should form independently. In the second part of this thesis, the loop domain model is tested in *E. coli* by combining LacI and λCI-mediated DNA loops in these different orientations. Accordingly, we quantify DNA looping assistance and interference by fitting experimental data to a statistical-mechanical model, confirming the predictions of the loop domain model. Furthermore, TPM measurements of the same looping constructs support predictions that non-supercoiled DNA *in vitro* should facilitate DNA looping assistance, but not interference. In addition to confirming the loop domain model in *E. coli*, this thesis provides a strong experimental and theoretical foundation for further investigations of enhancer-promoter looping in prokaryotes and eukaryotes, and the relationship between chromatin architecture and gene expression in metazoans.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to David Priest and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

_____ _____

Signed Date

Acknowledgements

I would like to sincerely thank my supervisors Dr Keith Shearwin and Dr Ian Dodd for their years of generous mentorship and support. Two incredibly intelligent, skilled, passionate and kind individuals who demonstrate, every day, why it's worth pursuing this seemingly crazy, yet intensely challenging and rewarding career choice of molecular biological research and teaching. I thank them for giving me the opportunity to work on a project that was so well-designed from the outset. I take with me their passion for discovery and unique perspective on biological systems and look forward to future collaborations.

I would also like to thank the other members of the Shearwin lab for their friendship and willingness to lend a hand. Thanks Iain Murchland, Julian Pietsch, Cui Lun, Nan Hao, Alex Ahlgren-Berg and the many other members I've seen come and go during my 9 year stint!

Thanks to all those within the Biochemistry Discipline and School who work hard to provide such a good learning and teaching environment to which I would happily return in the future.

Thanks Mum and Dad. I would have never been able to make it this far without your unwavering support.

Contents

Chapter 1

Introduction

1.1 Part I. Investigating long-range DNA looping *in vivo*

2011 saw the 50th anniversary of the Operon Theory of gene regulation proposed in 1961 by François Jacob and Jacques Monod [Jacob and Monod, 1961; Yaniv, 2011; Beckwith, 2011]. In this paper, the authors suggested a new class of genes, called regulatory genes, that function to regulate the expression of structural genes such metabolic enzymes. Through an elegant discussion of various data, it was deduced that the genes for lactose metabolism in *E. coli* (encoded by the *lac* operon) were under coordinate repression through binding of a *trans*-acting repressor molecule to a *cis*-acting operator site, and that gene expression could be 'induced' through interaction of a chemical inducer with the repressor. Jacob and Monod were awarded the Nobel Prize in 1964 for their operon theory, which at the time, served to unite many disparate areas of molecular biology. The theory gained rapid acceptance, and informed a new phase of research into various systems, where similar mechanisms of negative, but also positive [Englesberg et al., 1965] control were identified.

Fast forward to 1974 and it was firmly established that repression of the *lac* operon was mediated through binding of the allosteric LacI repressor protein to a promoter-proximal operator sequence (now known as *lacO1*) [Gilbert and Müller-Hill, 1966; Bourgeois and Riggs, 1970; Gilbert and Maxam, 1973; Lewis, 2005]. In this year too, another *lac* operator sequence with lower affinity (later to be known as *lacO2*) was discovered within the *lacZ* coding se-

quence [Reznikoff et al., 1974]. The authors made the insightful prediction that this secondary operator may serve as a roadblock to RNA polymerase, but they also suggested it may "have resulted from an evolutionary accident" [Reznikoff et al., 1974]. Despite *in vitro* findings suggesting LacI tetramers could simultaneously bind two DNA fragments bearing *lac* operators [Culard and Maurizot, 1981], several more years elapsed before the first *in vivo* evidence for LacI-mediated DNA looping between two *lac* operators was published [Mossing and Record, 1986], where *O1* placed at three different positions upstream of a plasmid-based P_{lacI} reporter caused an $\sim$ 40 fold increase in repression. During this time, DNA looping was demonstrated for bacteriophage lambda CI [Hochschild and Ptashne, 1986; Griffith et al., 1986] and was also suggested to be important for other gene regulatory systems (such as the *ara*BAD operon [Dunn et al., 1984]), and for recombination (including bacteriophage integration), DNA replication and in the burgeoning field of eukaryotic enhancers (see below, Section 1.2) [Schleif, 1992; Matthews, 1992].

In the early 1980s, a separate body of DNA looping results emerged from studies on the *in vitro* cyclisation rate of DNA molecules catalysed by T4 DNA ligase [Shore et al., 1981; Shimada and Yamakawa, 1984]. A key observation was that the cyclisation rate of short DNA fragments oscillates with a period equivalent to the *in vitro* helical repeat of DNA ($\sim$ 10.5 bp). The persistence length is a basic mechanical property quantifying the stiffness of a polymer and is defined as the length over which the time-averaged orientation of the polymer becomes uncorrelated with the direction of its tangent [Peters and Maher, 2010]. Estimates of the persistence length for B-DNA *in vitro* are $\sim$ 150 bp. At distances below the persistence length, DNA behaves much like a flexible, elastic rod, which is resistant to twisting and bending. At distances above the persistence length, the contribution of bending and torsional rigidity diminishes, and the conformation of a DNA molecule is predicted by statistical polymer models, such as the freely jointed chain model [Rippe, 2001]. Therefore at shorter range, the probability of DNA looping is dominated by the inflexibility of the DNA molecule, whilst at longer range the free energy of forming a loop is dominated by the entropic cost of spatially constraining two DNA sites. The helical dependence of cyclisation probability arises from the requirement for an integral number of helical turns in a DNA circle; when the cohesive DNA ends are out of phase, the resistance

of the DNA to torsional strain results in a lower ring closure probability.

These *in vitro* studies laid the theoretical groundwork for quantitation of DNA looping *in vivo*. In a cyclisation experiment, ligation can occur between two ends of the same molecule (cyclisation) or between two separate molecules (intermolecular ligation). Cyclisation can be favoured since the DNA tether linking the two DNA ends increases the local concentration of one end in the vicinity of the other compared to the free ends of other DNA molecules. This concentration (termed J_{loop}) can be calculated from the cyclisation probability (see Shore and Baldwin [1983]; Rippe [2001] for review). For DNA looping *in vivo*, J_{loop} therefore quantifies the contribution by the DNA tether to increasing the effective concentration of a DNA looping protein above the cellular background concentration in the vicinity of each looping site.

Following Mossing and Record [1986], a series of *in vivo* P_{lac}–*lacZ* reporter studies [Oehler et al., 1990; Law et al., 1993; Oehler et al., 1994] investigated LacI-mediated DNA looping. Müller et al. [1996] found that successively changing the distance between *lac* operators gave alternating repression maxima and minima, therefore showing that DNA looping *in vivo* also had a helical dependence (with an *in vivo* helical repeat of $\sim$ 11.3 bp) (Fig 1.1, left). More recently, our understanding of short-range DNA looping *in vivo* has expanded [Becker et al., 2005b, 2007; Han et al., 2009; Johnson et al., 2012], however Müller et al. [1996] remained the last comprehensive study of longer-range DNA looping, and the extent of their operator spacing measurements was $\sim$ 1500 bp, leaving open the question of how steeply the strength of DNA looping decays over the 1500–10000 bp range.

Aim I: Investigate long-range DNA looping *in vivo* and *in vitro*

The first aim of the present study is therefore to investigate longer range DNA looping *in vivo*. At shorter spacings, the nature of the DNA looping protein – it's size, flexibility and the way it multimerises – will significantly affect the looping propensity, however at longer spacings these effects are expected to diminish and J_{loop} will be a property of the tether (the DNA and associated proteins) linking the two sites. We measure DNA looping *in vivo* (from 242–10000 bp) using two well-characterised DNA looping proteins, the lac repressor (LacI) and bacteriophage λ repressor (λCI), which bridge their operator sites via different molecular mechanisms. Whilst LacI binds the DNA as a preformed tetramer (leaving the other DNA binding head free to bind the other

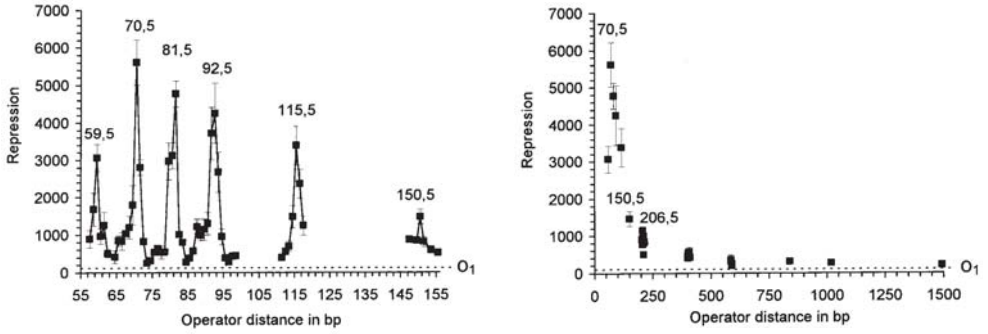


Figure 1.1: **Lac looping data of Müller et al. [1996].** LacZ units of chromosomally integrated *lac* looping reporters were measured in the absence and presence of LacI, the ratio of which yields the repression factor. **Left.** At shorter spacings, repression oscillates with a period of one helical turn. **Right.** At spacings greater than ~ 300 bp, the oscillations diminish, and repression decreases.

operator), λCI loops DNA through octamerisation of two separate tetramers bound cooperatively to each operator site [Dodd et al., 2004; Zurla et al., 2009]. Despite these differences, we find that J_{loop} vs distance for both looping proteins is similar, therefore providing strong evidence that the decay in J_{loop} we observe indeed represents a fundamental property of DNA in *E. coli*.

The persistence length for DNA *in vivo* – estimated to be as low as 20 bp, [Becker et al., 2005a] – is much shorter than for B-DNA *in vitro* (~ 150 bp). The increased flexibility of DNA in the cell is suggested to arise from (i) supercoiling of DNA *in vivo* and (ii) Nucleoid Associated Proteins (NAPs) (e.g. HU, IHF, H-NS and Fis), which bind both to specific genomic sites and to the bacterial nucleoid in a largely sequence-non-specific manner (reviewed in [Dillon and Dorman, 2010]). NAPs have been suggested to decrease the *in vivo* persistence length of DNA by introducing kinks and bends into the double helix. However it is unclear exactly how NAPs (and indeed nucleosomes) contribute to the increased flexibility of DNA *in vivo* and various biochemical and simulation methods are currently being employed to address this question [Peters and Maher, 2010].

Despite the large body of literature concerning short-range DNA looping *in vitro* [Dunlap et al., 2011; Johnson et al., 2012; Han et al., 2009], as is the case with the *in vivo* situation, little *in vitro* data exists over longer range [Hsieh et al., 1987; Zurla et al., 2009]. The second part of Aim I is therefore to systematically measure longer-range DNA looping *in vitro* using the technique of Tethered Particle Motion (TPM) and compare these results with the long-range

in vivo data. These experiments were carried out by our collaborators Dr David Dunlap and Dr Sandip Kumar at Emory University (Atlanta, USA). As predicted, J_{loop} was found to be lower *in vitro* than *in vivo* (i.e. weaker DNA looping) thereby confirming the influence of context on DNA looping at long range. Determinants of increased long-range DNA looping *in vivo* (such as DNA supercoiling and NAPs) will be the subject of future studies.

1.2 Part II. Testing the loop domain model in *E. coli*

During the 1980s 'action at a distance' (long-range gene regulation) was observed in Eukaryotes, with the discovery of stretches of DNA called enhancer elements that were capable of increasing gene expression when positioned from ~ 100 bp up to many kilobases upstream or downstream the promoter [Banerji et al., 1981; Müller et al., 1988; Maniatis et al., 1987]. Studies of the enhancers of small DNA tumour viruses (such as SV40 virus) provided early insights, such as the observation that enhancers are modular in structure, consisting of combinations of DNA motifs bound by various transcription factors.

The discovery of enhancers began to help explain earlier evidence suggesting that the surrounding chromatin environment could influence gene expression. Position effect variegation (PEV) in the fly *Drosophila melanogaster* arises when chromosomal translocations juxtapose normally active (euchromatic) genomic regions with silent (heterochromatic) regions [Lewis, 1950]. Variable spreading of heterochromatin into the euchromatic region causes stochastic silencing of genes near the breakpoint, leading to a variegated expression pattern in *D. melanogaster* embryos and flies. Also, translocations into non-heterochromatic regions can yield non-variegated position effects (chromosomal position effect (CPE)), which are now known to result from adventitious enhancer-promoter (EP) communication across the translocation breakpoint ('enhancer hijacking').

The widespread presence of enhancers in *D. melanogaster* was shown using the 'enhancer trap' technique, where a *lacZ* reporter gene fused to a P-element transposon generated tissue- and stage-specific expression patterns (i.e. CPE) depending on the genomic site of insertion [Bellen et al., 1989]. This observation that a naïve transgene can respond to the surrounding cis-regulatory environment of the genome suggests that rather than being promoter-specific ele-

ments, enhancers can activate multiple promoter types in a general manner [Geyer and Clark, 2002]. Work on the immunoglobulin heavy chain (Igh) and β-globin loci showed enhancers to be a common feature of eukaryotic developmental loci, and that genes within these loci responded to multiple enhancer sequences located far from promoters (See Müller et al. [1988]; Maniatis et al. [1987]; Noordermeer and Duboule [2013] and refs therein). Recent advancements in genomics technology (such as Chromatin Immunoprecipitation followed by next-generation DNA sequencing (ChIP-seq)), have allowed the prediction of enhancer elements based upon characteristic chromatin signatures (such as H3K4me1 and H3K27ac) and transcription factor binding (such as the presence of CBP/p300) [Zentner et al., 2011]. Furthermore, we can now directly screen large numbers of DNA sequences for enhancer activity [Arnold et al., 2013; Shlyueva et al., 2014]. Enhancers are now estimated to outnumber promoters with 50–100,000 enhancers predicted in mice and humans [ENCODE Project Consortium et al., 2012; Thurman et al., 2012; Andersson et al., 2014].

Genes are often activated by multiple enhancers located many kilobases upstream and downstream from the promoter [Marinić et al., 2013] and even within introns of unrelated genes [Lettice et al., 2002]. Enhancers recruit transcription factors and the transcriptional machinery and have been shown to directly loop to target promoters [Carter et al., 2002; Tolhuis et al., 2002], thereby increasing the local concentration of these factors at promoters. The observation that enhancers act over such long genomic distances raises the question: what drives *efficient* and *specific* long-range EP contacts? At the β-globin locus – where expression of the β-globin subunit of heamoglobin is controlled by erythroid-specific transcription factors such as GATA-1, EKLF and Ldb1 – EP specificity appears to be provided by interactions between these specific factors bound at enhancer and promoter. In the absence of GATA-1 or Ldb1, DNA looping between the Locus Control Region (LCR, the strong globin enhancer) and the βmaj gene is abolished, resulting in low βmaj expression. GATA-1 binds the DNA at both the LCR and βmaj promoter, whilst Ldb1 is a co-factor that cannot bind DNA itself. Interestingly however, in the absence of GATA-1, whilst Ldb1 binding is lost at βmaj, it remains associated with the LCR. Using a GATA-1-deficient erythroblast cell line, Ann Dean and colleagues [Deng et al., 2012], were able to restore LCR-βmaj contact (and partially restore gene expression) by using an engi-

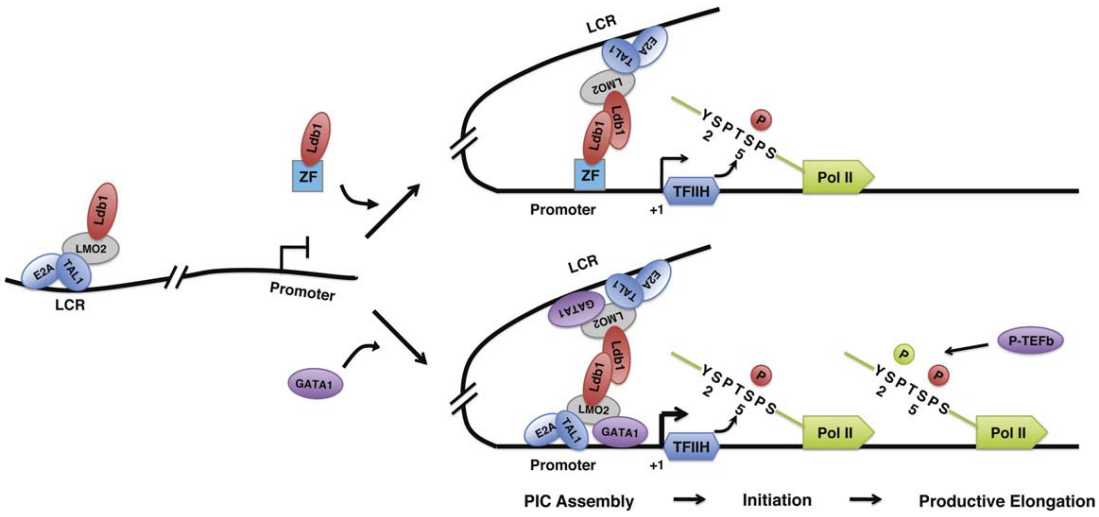


Figure 1.2: Figure 7 from Deng et al. [2012] showing a hypothetical model of the β-globin locus where transcription factor Ldb1 was suggested to bridge contacts between the LCR and globin promoters (bottom). In the absence of GATA-1 (and DNA looping), targeting Ldb1 to the promoter with a zinc finger fusion protein (top) restored LCR-promoter contact.

neered zinc-finger DNA binding fusion protein to target Ldb1 to the βmaj promoter (Fig 1.2). It was suggested that Ldb1 therefore bridges interactions between the LCR and globin promoters thus showing that transcription factors themselves can be a driving force in EP interactions.

Compatibility between enhancers and promoters is often determined by the assortment of transcription factors bound at each site, however given the well-documented promiscuity of many enhancers, there must exist mechanisms to insulate promoters from off-target EP contacts, which could otherwise lead to transcriptional noise and gene mis-regulation. Genetic elements capable of insulation (termed 'enhancer blockers') were first discovered in *D. melanogaster* through careful analysis of mutants that resulted from the presence of a ~ 7.5 kb retrotransposon, called *gypsy*, at various genomic loci. *Gypsy* was discovered as a class of phenotypes that could be suppressed by mutations at the Suppressor of Hairy-wing (*Su(Hw)*) locus, which encodes a protein that binds 12 degenerate sites within *gypsy*. One such allele, y^2, is caused by insertion of *gypsy* 700 bp upstream of the *yellow* gene, which is required for pigmentation of cuticle structures in the developing and adult fly [Geyer et al., 1988]. Expression of *yellow* in different cuticle parts is governed by a series of enhancers ~ ± 2 kb from the promoter, and remarkably the *gypsy*

insertion of y^2 only causes the *yellow* phenotype (lack of *yellow* expression) in tissues whose enhancers are *distal* to the insertion site, that is, not between *gypsy* and the promoter [Geyer and Corces, 1992]. Therefore, Su(Hw)-bound *gypsy*, when interposed between a promoter and enhancer can block their communication.

At the same time the enhancer-blocking (EB) activity of *gypsy* was being demonstrated, a number of other elements capable of insulation were discovered [Gyurkovics et al., 1990; Kellum and Schedl, 1991; Holdridge and Dorsett, 1991; Geyer and Corces, 1992; Cai and Levine, 1995; Bell and Felsenfeld, 1999]. In one such study, Kellum and Schedl [1991] found that flanking an enhancer trap construct (consisting of the *white* gene driven by a minimal promoter) with the scs and scs' elements (from the 87A heat shock locus) insulated the reporter from CPE, whilst a construct bearing scs only at its 5' side was still responsive to CPE. This apparent non-autonomy of scs/scs' (they only function as pairs) seemed at first to contradict *gypsy* enhancer-blocking experiments, where only one element yielded insulation, however further studies showed that *gypsy* too could function as a pair [Holdridge and Dorsett, 1991] and that its EB activity was enhanced when a second *gypsy* element was placed downstream of the *white* reporter gene [Roseman et al., 1993], suggesting that both enhancer blocking and insulation from position effects required insulator 'pairing'. These studies showed that different insulators are capable of enhancer blocking irrespective of the nature of the enhancer or target gene, suggesting that insulation is a generic phenomenon.

The mechanisms by which insulators block EP communication has been the subject of considerable debate [Corces, 1995; Bell and Felsenfeld, 1999; Kuhn and Geyer, 2003; Capelson and Corces, 2004; Phillips and Corces, 2009; Bulger and Groudine, 2011; Krivega and Dean, 2012; Pennacchio et al., 2013; Phillips-Cremins and Corces, 2013; Noordermeer and Duboule, 2013; Chetverina et al., 2014]. Two prominent models of enhancer blocking are the decoy model and the loop domain model. In the decoy model, insulators are proposed to compete with promoters for contacts with enhancers, thereby sequestering enhancer signals away from promoters [Geyer and Clark, 2002] (Fig 1.3A). One obvious objection to the decoy model is that if insulators serve as enhancer sinks, then they should elicit enhancer blocking when placed *upstream* of an enhancer, however this is not observed. A range of other data (see below) is unexplainable by

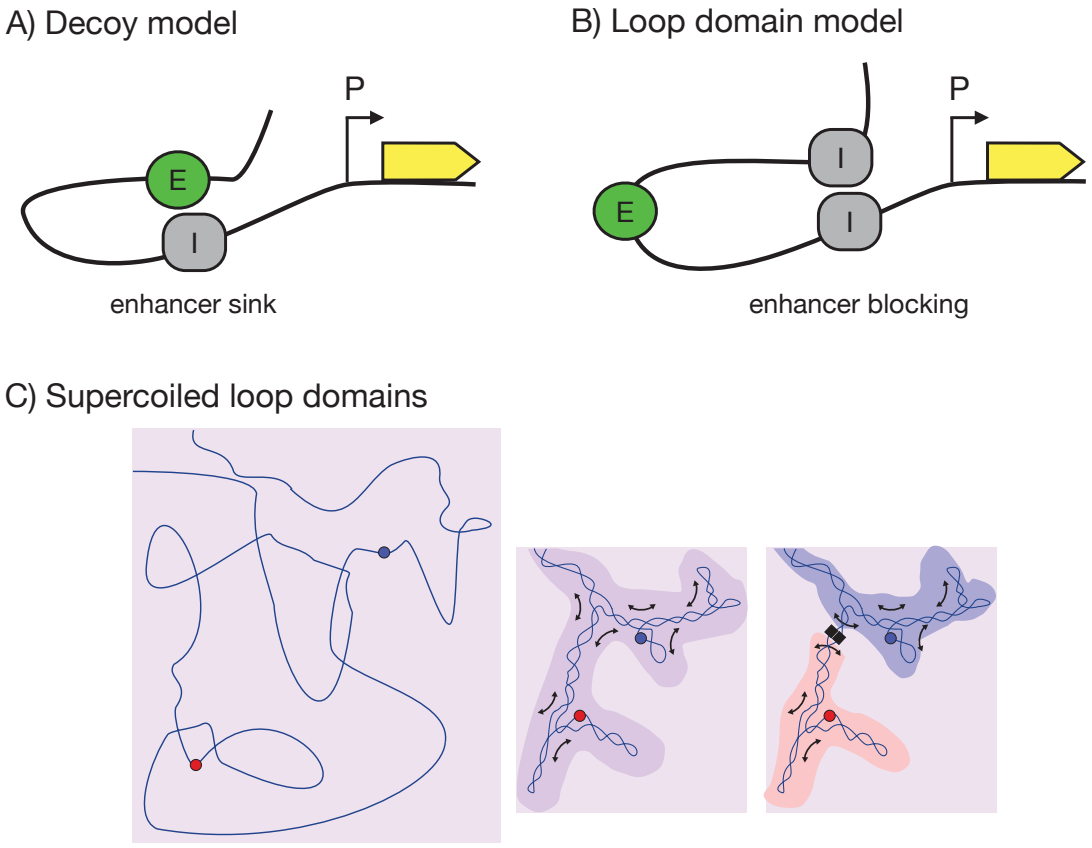


Figure 1.3: **Models of insulator function.** **A.** Decoy model. Insulators loop directly to enhancers and promoters, neutralising them and preventing enhancer-promoter communication. **B.** Loop domain model. Insulators loop to one another and sequester enhancers and promoters in separate topological loop domains. **C.** Supercoiling and loop domains. Left. In a non-supercoiled polymer, two sites find one another via collisions in 3D. Middle. Supercoiling reduces the search space for the two sites to the darker purple area. Right. A DNA loop (black boxes) is a barrier to supercoiling and the sites must again search the whole space (light purple) to find one another. Figure 1.3C by Ian Dodd, see Chapter 3.

decoy mechanisms yet consistent with the predictions of the loop domain model.

In the loop domain model [Corces, 1995], DNA looping between insulators divides the genome into a series of DNA loop 'domains' (Fig 1.3B). The key proposition of this model is that the DNA loop *itself* forms a topological barrier, which reduces the contact probability between sites in separate loops. The loop domain model predicts that one DNA loop (say between a pair of insulators) can *interfere* with the formation of an alternative (e.g. EP) loop (Fig 1.4, left). Therefore, enhancer-blocking insulation can be achieved by placing an enhancer and promoter in separate DNA loops (Fig 1.3B). Although DNA looping is driven by specific insulator-insulator contacts, the actual mechanism of enhancer blocking (DNA looping interference) is passive, and therefore relies on some biophysical property. DNA supercoiling is suggested to be a prominent force because sliding of a supercoiled DNA strand along itself would have the effect of reducing the search space for two sites on the DNA (Fig 1.3C). The base of a DNA loop should pose a barrier to the propagation of supercoiling because a bound DNA looping protein prevents the DNA strands from twisting relative to one another. As a result, two sites located in separate loops cannot encounter one another via supercoiling-aided sliding; their interaction is limited to less-frequent collisions in three dimensional space.

Under the loop domain model, segregation of enhancers and promoters into separate DNA loops provides evolution with the means to partition the regulatory landscape of the genome. A key requirement therefore is the formation of specific insulator-mediated DNA loops (insulator pairing). Early studies hinted insulator pairing to be a possible mechanism of enhancer blocking [Roseman et al., 1993; Kellum and Schedl, 1991], and clustering of insulators was suggested by observations that insulator proteins coalesce into discrete subnuclear foci termed 'insulator bodies' [Byrd and Corces, 2003; Capelson and Corces, 2004] (however see [Schoborg et al., 2013]). Compelling evidence for insulator pairing came from two studies published in 2001 [Muravyova et al., 2001; Cai and Shen, 2001]. Consistent with previous results, the authors demonstrated enhancer blocking activity of a single Su(Hw) site placed between an enhancer and reporter transgene, however, remarkably, interposition of a second Su(Hw) site resulted in loss of EB activity. Whilst DNA looping between a Su(Hw) site interposed between the enhancer and promoter and another site upstream of the enhancer (or downstream of the reporter), would

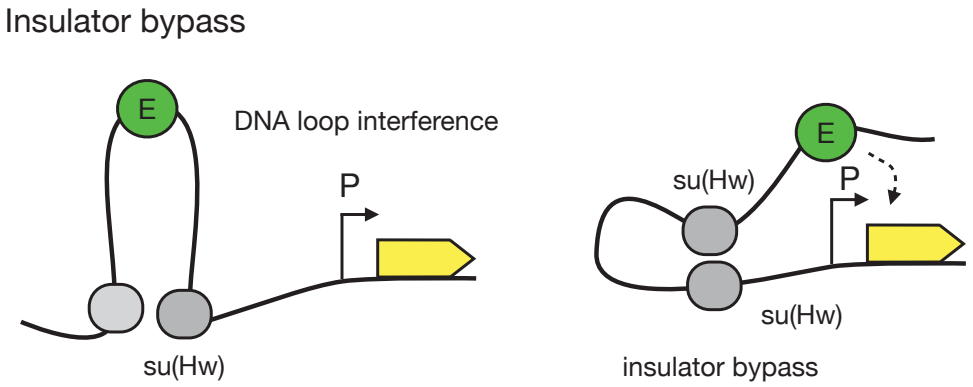


Figure 1.4: **Insulator bypass** Left. A single Su(Hw) site between enhancer and promoter can loop to another exterior insulator leading to enhancer blocking. Right. Two interposed Su(Hw) sites will loop to one another, returning the enhancer and promoter to the same loop domain.

place enhancer and promoter in separate DNA loops, looping between two interposed Su(Hw) sites returns the enhancer and promoter to the same loop domain (Fig 1.4). This phenomenon of 'insulator bypass' was also shown for other insulators [Melnikova et al., 2004], and is reversed with the addition of a third Su(Hw) insulator site [Kuhn et al., 2003]. More recently, direct evidence of DNA looping between insulators (and indeed between enhancers and promoters [Tolhuis et al., 2002]) came from the Chromosome Conformation Capture (3C) technique [Dekker et al., 2002], which detects the spatial proximity of selected chromosomal loci in the nucleus, allowing an estimate of their relative contact frequency [Blanton et al., 2003; Comet et al., 2011; Kyrchanova et al., 2013].

Recent experimental support in *D. melanogaster* for the loop domain model comes from studies employing more complex transgenic constructs [Savitskaya et al., 2006; Gohl et al., 2011; Kyrchanova et al., 2013]. A criticism of earlier enhancer blocking studies is that the spacer sequences between insulators, enhancers and promoters were small (<1 kb), and if insulators work by forming DNA loops, then the resulting small DNA loops may sterically hinder access of transcription factors and RNA polymerase. Notably however, Kyrchanova et al. [2013] still observed enhancer blocking of the *white* gene, when its enhancer was placed at the centre of a ~ 4.3 kb DNA sequence flanked by *gypsy* insulators, which was itself another 2 kb upstream of the *white* promoter, thereby showing that enhancer blocking could be elicited by rather large DNA

loops. Also, Paul Schedl and coworkers [Gohl et al., 2011] tested an elegant series of constructs harbouring multiple enhancers, insulators and promoters, all of which were consistent with the loop domain model.

Investigations of enhancer blocking are not limited to *D. melanogaster*. Two studies using supercoiled plasmid templates *in vivo* [Ameres et al., 2005] and *in vitro* [Bondarenko et al., 2003] showed that sequestering an enhancer in a controllable DNA loop reduced its ability to activate a reporter gene outside the loop. Ameres et al. [2005] transiently transfected HeLa cells with plasmids bearing an SV40-luciferase reporter under control of an SV40 enhancer flanked on either side with a series of *tet* operators. DNA looping between the *tet* operators was driven by a co-transfected construct encoding a fusion protein of the tet DNA binding domain and a dimerisation domain. When the fusion protein was absent, or when DNA binding was abolished by treatment with doxycycline, the SV40 enhancer drove robust luciferase expression. However in the presence of the fusion protein, enhancer blocking was observed, due presumably to sequestration of the SV40 enhancer in a separate DNA loop. Two weaknesses of this study however were that the size of the DNA loop around the enhancer was relatively small (344 bp), and that plasmid, not chromosomally-integrated reporters were used, which may have yielded unexpected effects due to interactions between two separate plasmids. Bondarenko et al. [2003] performed a similar study *in vitro*, separating the bacterial NtrC enhancer from its *glnAp2* target promoter with a LacI-driven DNA loop and measuring enhancer blocking using *in vitro* transcription assays. In this study, enclosing the promoter within a 3.7 kb lac loop yielded robust enhancer blocking, and it was suggested that the lac loop reduced EP communication by presenting a barrier to the propagation of supercoiling (as in Fig 1.3C). Although this study was quantitative and used well-defined components, it was nonetheless performed on DNA *in vitro*, which is known to behave differently to DNA inside the cell (see above, Section 1.1).

Presently, the only confirmed insulator protein in vertebrates is CCCTC-binding factor (CTCF), a zinc finger DNA binding protein (for review see [Phillips and Corces, 2009; Holwerda and de Laat, 2013; Ong and Corces, 2014]). CTCF ChIP-seq peaks often overlap with those of the Cohesin complex, which was originally identified as important for sister chromatid cohesion, but has now been implicated in establishing/maintaining long-range chromatin looping [Wendt

et al., 2008]. Perhaps the strongest evidence in mammalian cells for enhancer blocking comes from studies with the β-globin locus [Hou et al., 2008; Noordermeer and Duboule, 2013]. The human β-globin locus is flanked at its 5' end (upstream of the LCR) and 3' end (downstream of the β-globin gene) by DNAseI hypersensitive sites HS5 and 3'HS1 respectively, both of which harbour CTCF sites, and which loop to one another in early erythroid cells prior to globin gene activation (Fig 1.5A). During differentiation, activation of β-globin expression is accompanied by looping of its promoter to the LCR. Using mice transgenic for the human β-globin locus, Ann Dean and colleagues [Hou et al., 2008] introduced an ectopic copy of HS5 between the LCR and globin genes and found (using 3C) that it contacted the natural HS5 element upstream of the LCR. Coincident with the appearance of this ectopic contact they observed a decrease in contacts between the LCR and β-globin promoter. This result strongly supports a model whereby DNA looping between the ectopic and natural HS5 sites sequesters the LCR in an insulating loop, preventing it from looping to and activating the globin genes (Fig 1.5A). Interestingly, this effect was dependent upon the CTCF site within the ectopic HS5, since deletion of the site (or siRNA depletion of CTCF) restored LCR-globin contacts and globin gene expression [Hou et al., 2008].

The above studies provide compelling evidence that separating nearby enhancers and promoters into separate DNA loops can inhibit their contact, however these studies all used artificial transgenic constructs, and the degree to which this phenomenon has been employed by evolution in the natural genomic context is unclear. At the *bithorax* (BX-C) homeotic gene complex in *D. melanogaster*, there is some suggestion that developmentally regulated DNA looping between insulators may serve to restrict enhancer-promoter contacts in a tissue-specific manner, however this has not been directly shown (for review see [Maeda and Karch, 2011; Chetverina et al., 2014]).

One locus where alternate loop formation is suspected to inhibit EP contacts is the parentally imprinted *Igf2/H19* locus (for review see [Phillips and Corces, 2009]). This locus encodes two genes, insulin-like growth factor 2 (Igf2), which is exclusively expressed from the paternal allele, and the long non-coding RNA *H19*, which is expressed only from the maternal allele (Fig 1.5B). The paternal allele is methylated in the male germline, and methylation of the *H19* promoter

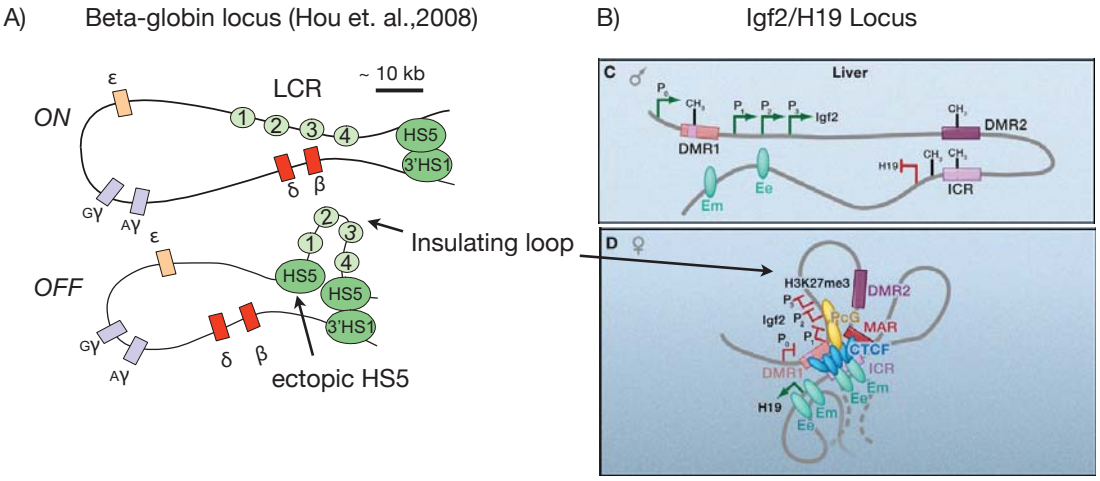


Figure 1.5: **Enhancer blocking in vertebrates A.** β-globin locus. Top. In the natural locus looping between HS5 and 3'HS1 facilitates contact between the LCR HS sites (HS1-4) and globin promoters. Bottom. Loop formation between HS5 and an ectopically introduced HS5 inhibits LCR-promoter contact by placing them in separate loops. Image is 6D from Hou et al. [2008]. **B.** *Igf2/H19* locus. Top. Paternal allele. ICR and DMRs are methylated preventing CTCF binding. *H19* promoter is methylated and inactive. Enhancers access and activate *Igf2* promoter. Bottom. Maternal allele. CTCF binds to unmethylated ICR and DMRs, isolating *Igf2* in separate loop from enhancers. *H19* promoter is unmethylated and activated by downstream enhancers. Image is Fig 1 from Phillips and Corces [2009].

silences its expression. On the paternal allele, the enhancers downstream of *H19* loop back to and activate *Igf2*. On the unmethylated maternal allele, the *H19* promoter is unmethylated and active, however *Igf2* is sequestered within a CTCF/Cohesin-mediated DNA loop between the Imprinting Control Region (ICR), and Differentially Methylated Regions (DMRs) upstream. This loop doesn't form on the paternal allele since methylation of the ICR and DMRs prevents CTCF binding. On the maternal allele, silencing of *Igf2* involves polycomb repression, however it was also suggested a loop domain mechanism comes into play whereby the enhancers cannot access *Igf2* since they are in an alternate DNA loop [Murrell et al., 2004] (Fig 1.5B).

Chromosomes in interphase nuclei do not intermingle completely, rather they occupy distinct chromosome territories that are largely invariant throughout the cell cycle [Cremer and Cremer, 2001, 2010]. Chromosomes have a rough inside-outside organisation and seem to show preferred sub-nuclear positions, however the relative locations of territories from one cell to the next is largely random [Nagano et al., 2013]. The existence of long-range gene regulation shows us that

genome structure is important for genome function, however a central question in the field is: does genomic function *care* about genomic structure? Specifically, how much of the observed chromatin architecture is a rather irrelevant consequence of 'functional' phenomena such as long range enhancer-promoter looping and transcription? Conversely, to what degree does the establishment of chromatin architecture – through for example enhancer blocking by insulator pairing – have a causative influence on genomic output? Drawing a line between these two ideas may turn out to be a matter of semantics, however the latter (function follows form) seems to be supported by studies of enhancer blocking and may rest on the validity of the loop domain model. As will be discussed below, findings from recently-developed genome-wide technologies have seen a revolution in our understanding of how genomic structure relates to its functional output.

In a Chromosome Conformation Capture (3C) experiment, cells are fixed with formaldehyde followed by restriction enzyme digestion of DNA. Subsequent re-ligation permits junctions to form between cross-linked DNA fragments that were in close proximity in the nucleus, yet may be distant in the linear genome [Dekker et al., 2002]. Suspected DNA loops are then interrogated semi-quantitatively using qPCR with primers that amplify across junctions. 3C allows detailed dissection of chromatin architecture, however because it uses pairs of PCR primers to detect pairwise interactions between preselected locations, its throughput is limited to hypothesis-driven investigations of small genomic regions. Recently however, the advent of next-generation DNA sequencing has facilitated unbiased interrogation of 3C libraries to detect contact frequencies genome-wide in an 'one-to-all' (4C) or 'all-to-all' manner (5C, Hi-C and ChIA-PET) [Lieberman-Aiden et al., 2009; Dixon et al., 2012; Shen et al., 2012; Sexton et al., 2012; Nora et al., 2012; Sanyal et al., 2012; Hou et al., 2012]. The remarkable finding from these studies is that metazoan genomes are subdivided into discrete self-associating regions, termed Topologically Associating Domains (TADs), with a median size of $\sim$ 880 kb in mammals [Dixon et al., 2012] (61 kb in *D. melanogaster* [Hou et al., 2012]). These domains are clearly observable on interaction heat maps as distinctive triangles (Fig 1.6). Intriguingly, TADs are established in embryonic stem (ES) cells and their boundaries are often invariant throughout differentiation and even between mouse and human [Dixon et al., 2012]. On the sub-TAD level

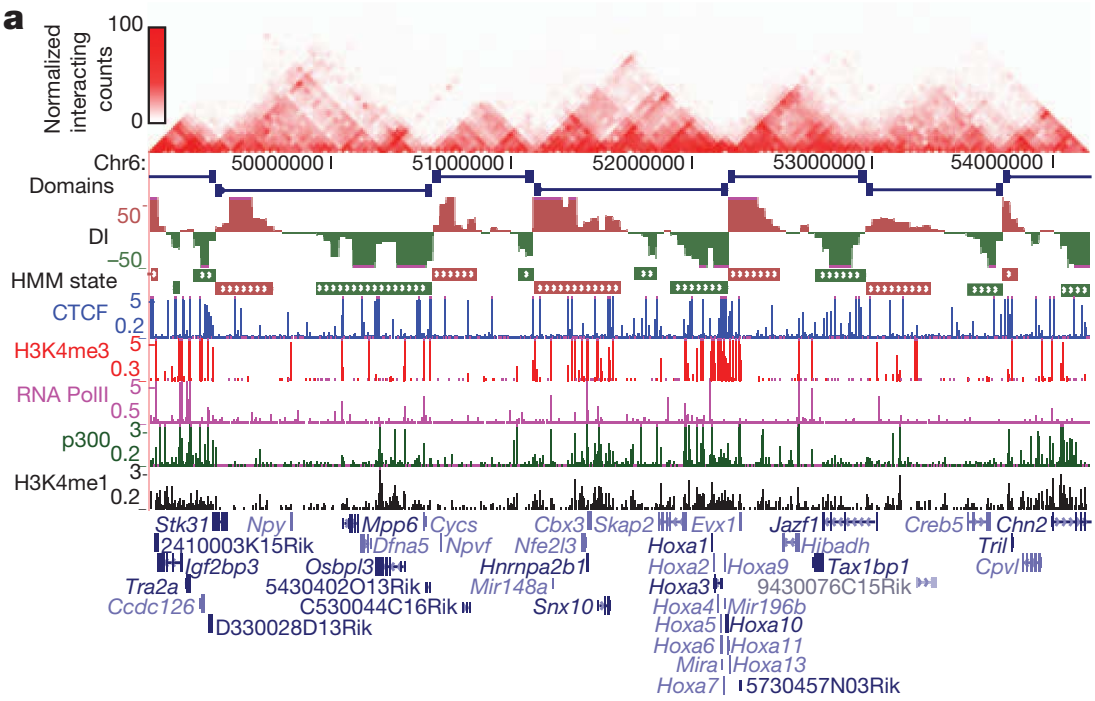


Figure 1.6: **Topologically-associating domains (TADs).** Shown is Figure 1A from the Hi-C study of Dixon et al. [2012]. The chromatin interaction matrix is tilted on its side to show interaction frequencies along a genomic region. TADs clearly stand out as triangles on the heat map, i.e. regions that have more inter-domain than intra-domain contacts. The bottom of the figure shows genes and ChIP-seq data sets.

however, extensive cell-type specific contacts are observed, which could reflect the establishment of distinct enhancer-promoter and insulator-insulator interactions in different cell lineages. The plaid pattern in Hi-C interaction matrices suggests chromosomes preferentially segregate into two broad zones (termed A and B compartments) [Lieberman-Aiden et al., 2009], however cell-to-cell reproducibility in genome conformation doesn't seem to emerge until the $\sim$ 1 Mb length scales of TADs, suggesting TADs to be a fundamental building block of the genome.

The existence of long-range gene regulation in metazoans combined with the observation that enhancers often activate sets of co-regulated genes (e.g. at the *Hox* clusters [Andrey et al., 2013]) should impose an evolutionary constraint on the rearrangement of genes by chromosomal translocations and inversions. Indeed, blocks of microsynteny, containing two or more genes and associated enhancer elements (termed Genomic Regulatory Blocks (GRBs)) have been conserved across diverse metazoan lineages [Kikuta et al., 2007; Irimia et al., 2012, 2013], thus

showing that cis-regulatory constraints can conserve ancient gene associations. In mammals, enhancer-trap studies have shown that broad genomic regions (~ 1 Mb) can drive similar tissue-specific expression patterns of a transgenic reporter construct [Ruf et al., 2011], which suggests that the genome is divided into distinct cis-regulatory landscapes over which enhancers have broad sway to define tissue-specific expression profiles. Early attempts to assign enhancers to their target promoters often resorted to a nearest-neighbour approach despite observations that enhancers do not always regulate their closest gene [Lettice et al., 2002; Zhang et al., 2013]. More recently, by comparing the ChIP-seq signatures of promoters and enhancers and transcriptomes across 19 diverse tissues and cell types, Bing Ren and colleagues [Shen et al., 2012] were able to accurately deduce EP pairs based on their co-activity in cell type subsets (and see Kieffer-Kwon et al. [2013]; Jin et al. [2013]; Andersson et al. [2014] for more recent developments). This study also found clusters of co-regulated genes, termed Enhancer Promoter Units (EPUs), which covered ~ 50% of the mouse genome, and notably, aligned quite well with mouse TAD borders. Furthermore, a recent enhancer trap study from François Spitz and co-workers showed that cis-regulatory landscapes align well with TADs [Symmons et al., 2014]. In this study, multiple reporter insertions within one TAD usually followed a similar tissue-specific expression profile in mid-gestation mouse embryos, whereas insertions across TAD borders often displayed differing expression. In another study, analysis of an ~ 4 Mb region encompassing the X inactivation centre (XIC) identified TADs covering the region and also revealed correlated expression patterns for genes in the same TAD [Nora et al., 2012]. These results suggest that the borders of topological domains define the extent of cis-regulatory landscapes, with TAD borders limiting the influence of enhancers to the domain in which they reside; or in other words, TAD borders are insulators.

Differentiation is accompanied by activation of cell-type specific enhancer repertoires, which go on to form specific contacts with other enhancers, insulators and promoters [Kieffer-Kwon et al., 2013; Nord et al., 2013; Stergachis et al., 2013]. However, an important question in the field concerns the *dynamics* of EP looping. For example, given some stimulus-responsive enhancer, is this enhancer already in contact with its target promoter prior to stimulation, or are they unconnected and diffusing about some nuclear volume? If they are unconnected, does

binding to the enhancer of a stimulus-dependent transcription factor drive EP contact? Even if EP contacts are not pre-formed, data suggests that the ability of any genomic locus to diffuse throughout the nucleus is restricted due to it being tethered within its chromosome territory. This was elegantly demonstrated by Noordermeer et al. [2011], who showed that an ectopically introduced β-globin LCR was only able to activate the endogenous *β-globin* gene (located on a different chromosome) in a few 'jackpot cells', in which the chromosomes harbouring the ectopic LCR and *β-globin* gene were adequately aligned. If EP contacts undergo significant changes upon cell stimulation, then 3C profiles would be expected to be drastically different in treated vs un-treated cells. However, Bing Ren and colleagues recently showed that while TNF-α treatment of IMR90 cells resulted in activation of NF-κB-responsive enhancers and their target genes, the treatment did not significantly change EP contacts [Jin et al., 2013]. Even when stimulus-dependent changes in local genome structure are observed (e.g. [Wood et al., 2011]), they may be more a fine-tuning or refinement of an existing pre-structure. Therefore, it appears that passage along a cellular lineage necessitates a reconfiguration of genome conformation, yet in any one interphase nucleus EP contacts may be largely pre-formed (poised), and indeed that these pre-structures actually shape a cell's genetic response to stimuli, highlighting genome structure as a crucial determinant of cellular identity (e.g. see [Fanucchi et al., 2013]). Nevertheless, some studies suggest that long-range movement of DNA within the nucleus in response to stimuli does occur [Li et al., 2013], and furthermore, whether or not genes migrate to 'transcription factories' upon activation is a point of contention in the field [Papantonis and Cook, 2013; Zhao et al., 2013; Cisse et al., 2013].

The defining feature of TADs is a reduced frequency of interaction between genomic sequences either side of a domain boundary (compared to sequences within the same domain), however the mechanisms that define boundaries remain unclear [Phillips-Cremins and Corces, 2013]. TADs often correlate with blocks of epigenetic associations (such as lamina associated domains (LADs) and the heterochromatin mark H3K9me3) [Dixon et al., 2012; Sexton et al., 2012], however TAD boundaries were not disrupted on the X-chromosome in ES cells lacking the enzymes responsible for deposition of H3K27me3 (*Eed*$^{-/-}$) or H3K9me2 (*G9a*$^{-/-}$), suggesting epigenetic blocks are a consequence, not a cause of TAD boundaries [Nora et al., 2012]. No-

tably, a 58 kb deletion over a TAD boundary at the X-inactivation centre (XIC) led to partial merging of the two adjacent domains [Nora et al., 2012], suggesting specific DNA sequence elements reside at boundaries.

Since Hi-C identifies boundaries on a genome-wide scale, it is possible to identify the types of factors statistically associated with those boundaries, and for both mammals and *D. melanogaster* two defining features seem to be an enrichment for insulators and active genes. In *D. melanogaster* a number of insulator binding proteins (IBPs) are known (e.g. BEAF-32, dCTCF, CP190 and Su(Hw)), and their binding sites are significantly associated with transcription start sites (TSSs). IBP sites are enriched in clusters at domain boundaries and thus so too are clusters of expressed TSSs [Sexton et al., 2012; Hou et al., 2012]. *D. melanogaster* insulators loop to one another and these loops may bring enhancers close to target promoters whilst blocking their access to off-target promoters via looping interference. Co-operative looping between clusters of IBP sites at domain boundaries may facilitate strong/frequent boundary-boundary interactions. However, since TSSs are also clustered at TAD borders it is likely that borders would engage in considerable interactions with cis-regulatory sequences in each adjacent domain. This integration of EP contacts at a TAD boundary was recently elegantly shown at the *HoxD* homeotic cluster in mammals, which strikingly resides at the border between two TADs [Andrey et al., 2013]. Therefore since borders may engage both in contacts with other borders as well as the domain interior, a key question arises: when moving through a TAD border, what mechanisms determine the sudden shift in the polarity of contact preference between sequences in the upstream domain, to sequences in the downstream domain? Precise genetic dissection of TAD boundaries should provide answers to these questions.

In mammals, whilst the bulk of CTCF sites (85%) reside within TADs, CTCF ChIP-seq peaks are nonetheless strongly enriched at ~70% of TAD borders [Dixon et al., 2012]. Mammalian TAD borders are also enriched in housekeeping genes (such as tRNA genes, which have a broad expression profile), suggesting that transcription may have a role in defining TAD boundaries. Elongating RNA polymerase generates positive and negative supercoiling ahead and behind of itself respectively, and increased concentration of TSSs at TAD boundaries suggests they may be sources of supercoiling, and raises the question of whether supercoiling (and other features

of highly expressed chromatin such as loose packaging (DNAseI hypersensitivity [Thurman et al., 2012])), may be a driving factor behind boundary formation. Notably, a recent study that identified Hi-C domains in *Caulobacter crescentus*, found that ectopic insertion of a highly expressed gene resulted in the formation of a strong nascent domain boundary [Le et al., 2013]. Also, a recent study measuring levels of supercoiling across human chromosome 11 discovered ~ 600 'supercoiling domains' on that chromosome and that ~ 30% of TAD boundaries lay within 20 kb of supercoiling domain boundaries [Naughton et al., 2013]. The interplay between supercoiling and TAD boundary definition will be an exciting area of future research and will help elucidate the degree to which supercoiling may help define insulated loop domains.

Further questions about TADs concern domain hierarchy and chromatin dynamics. *Drosophila* has large TADs (~ 600 kb), small TADs (~ 6 kb) and all sizes in between [Hou et al., 2012], and whilst some large TADs appear to have a largely homogeneous internal contact profile, other large TADs appear to consist of clusters of ever-smaller domains. The apparent lack of smaller TADs in mammals may or may not simply be a technical matter [Dixon et al., 2012]. Assuming existence of large numbers of specific insulator-insulator and enhancer-promoter loops, it would make sense for larger domains to be made up of smaller domains and it will be important to keep in mind this possible 'fractal' nature of genomic domains. To provide clearer answers, we require more higher resolution [Jin et al., 2013] and more comprehensive single-cell [Nagano et al., 2013] Hi-C studies.

The degree to which the loop domain model comes into play across three orders of magnitude of chromatin length (1–1000 kb) is a key question. Do ~ 1 Mb sized TADs not interact because they are separately-supercoiled loop domains? Furthermore, on the sub-TAD length scales of 1-100 kb, how often are EP contacts channeled through insulator loop formation? Indeed, in any particular cell, what are the dynamics of genome structure [Miyanari et al., 2013]? How often does an enhancer need to be in contact with its target promoter to cause productive transcription [Liu et al., 2013]? How reproducible is chromatin structure from cell to cell? To what degree can genes migrate around the nucleus? Indeed, upon activation, do genes 'loop out' of their chromosome territory into a transcriptionally permissive inter-chromatin compartment? [Sutherland and Bickmore, 2009]. Does elongating RNA polymerase track along chromatin or is it immo-

bilised in so-called 'transcription factories' [Papantonis and Cook, 2013]? Answers to these questions will require advancements in many areas such as the field of live single-cell, single molecule microscopy.

Aim II: Test the loop domain model *in vivo*

Given the large body of literature and discussion suggesting that insulators may function through a loop domain mechanism, it is surprising how little attention has been focused on quantitatively testing the biophysical underpinnings of this model. The loop domain model predicts that forming an insulating loop around an enhancer (or promoter) can inhibit formation of the alternate enhancer-promoter loop. Therefore at its most basic level, the model predicts that alternating DNA loops will *interfere* with one another's formation. Whilst DNA looping interference has been shown *in vitro* [Bondarenko et al., 2003], it has yet to be shown in any organism *in vivo*. Different research groups have attempted to test loop domain mechanisms in metazoans through genetic dissection of model developmental enhancer-promoter systems (such as the β-globin locus [Hou et al., 2008]), or introduction of transgenic reporters, e.g. [Gohl et al., 2011]. This approach of making targeted changes to DNA and their associated transcription factors has uncovered mechanistic details, however these studies struggle to provide quantitative insights into the biophysical basis of DNA loop insulation.

Two approaches might be taken to study the loop domain model in mammalian cells either through the use of transgenic reporters or by manipulation of DNA looping at developmental gene loci. The second approach is addressed in a research proposal at the end of this thesis (Section 5.3). For the first approach, genetic reporters could be constructed based on well-characterised components, e.g. the SV40 enhancer and luciferase reporter gene, and operator sites for DNA looping proteins could be placed around the enhancer and reporter gene in much the same way Su(Hw) sites were used in the early enhancer-blocking studies [Kuhn et al., 2003]. However there are three important technical considerations for this approach. Firstly, it would require development of DNA looping proteins that are orthogonal to the host, since using host components (such as insulator binding proteins and their cognate DNA binding sites) could be subject to off target interactions with other parts of the genome that may confuse the results. λCI is one such a candidate DNA looping protein, however modifications to its protein sequence may

be required e.g. addition of a nuclear localisation sequence. Secondly, the constructs would need to be integrated into the genome at defined sites, rather than maintained on episomal vectors because plasmid DNA would not recapitulate all of the features of chromatin. Current tools for creating transgenic cell lines (such as the PiggyBac system) are subject to issues including multiple integrant copy number and instability, and next generation techniques for integrating test constructs at single copy into genomic 'safe-harbour' locations (involving CRISPR/Cas and TALEN genome editing tools) would need optimisation [Sander and Joung, 2014]. Finally, a rigorous test of DNA looping interference requires quantitation of looping probabilities, and therefore requires well-characterised, quantitative promoter/reporter systems. Achieving this with currently-available luciferase reporters and mammalian promoters would require further optimisation.

The bacterial nucleoid differs greatly in both size and structure from metazoan chromatin, however *E. coli* nevertheless provides an ideal, tractable model system to test the basic phenomenon of DNA looping interference, which underpins the loop domain model. Indeed, the relative simplicity of the *E. coli* genome (lack of nucleosomes and the associated epigenetic code) should be viewed as a benefit for the first quantitative *in vivo* test of the loop domain model. *E. coli* has long been used as a model organism for discovering and testing basic biological phenomena and as such, is replete with well-characterised molecular biological tools, including a system for easy, single copy, targeted genomic integration [Haldimann and Wanner, 2001; St-Pierre et al., 2013] (see Chapter 4). Since *E. coli* lacks long-distance gene regulation *per se*, we take the genetic reporter approach. Therefore in order to quantify loop interference, we require two orthogonal, quantifiable *in vivo* DNA looping proteins. Conveniently, the manuscript for Aim I provides a detailed analysis of two such DNA looping systems, the Lac and λCI repressors (Chapter 2). The manuscript in Chapter 3 addresses Aim II by leveraging these tools to provide the first quantitative test of the loop domain model *in vivo*.

We find that loop interference is a real phenomenon and quantitate the degree to which DNA loops interfere with one another. Furthermore, by using a nested configuration of DNA loops, we quantitate the degree to which one loop can *assist* another's formation, which is reminiscent of the way insulator pairing has been suggested to assist EP looping [Ong and Corces, 2014].

We also use Tethered Particle Motion (TPM) to show that (as expected) loop interference is less strong *in vitro* and we speculate that this is due to the lack of supercoiling in the TPM setup. This study shows that, in principle, looping between insulators can drive EP contact specificity, and it provides a quantitative framework for making similar measurements in Eukaryotes with their larger DNA length scales and different DNA packaging into chromatin.

Chapter 2

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

David G. Priest[1], Lun Cui[1], Sandip Kumar[2], David Dunlap[2], Ian B. Dodd[1*], Keith E. Shearwin[1]

[1]School of Molecular and Biomedical Science (Biochemistry), University of Adelaide, Adelaide SA 5005, Australia.

[2]Department of Cell Biology, Emory University, Atlanta GA 30322, USA.

Proc Natl Acad Sci U S A. 2014 Jan 7;111(1):349-54. doi: 10.1073/pnas.1317817111.

Statement of Authorship

Title of Paper	Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and Lambda repressors
Publication Status	◉ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	David G. Priest, Lun Cui, Sandip Kumar, David D. Dunlap, Ian B. Dodd, and Keith E. Shearwin. Proc Natl Acad Sci U S A, 111(1): 349–54, Jan 2014. doi: 10.1073/pnas.1317817111.

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	David Priest		
Contribution to the Paper	Designed Lac looping constructs. Made and assayed Lac looping constructs. Made constructs for TPM experiments. Assisted collaboration correspondence. Helped prepare figures for paper. Helped write paper with Ian Dodd.		
Signature		Date	

Name of Co-Author	Lun Cui		
Contribution to the Paper	Designed, made and assayed lambda looping constructs, analysed data, assisted in writing supplementary material in manuscript.		
Signature		Date	

Name of Co-Author	Sandip Kumar		
Contribution to the Paper	Designed and performed TPM experiments, analysed data and assisted in writing the manuscript.		
Signature		Date	

Name of Co-Author	David Dunlap		
Contribution to the Paper	Designed experiments, analysed data and assisted in writing the manuscript		
Signature		Date	

Statement of Authorship

Title of Paper	Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and Lambda repressors
Publication Status	◉ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	David G. Priest, Lun Cui, Sandip Kumar, David D. Dunlap, Ian B. Dodd, and Keith E. Shearwin. Proc Natl Acad Sci U S A, 111(1): 349–54, Jan 2014. doi: 10.1073/pnas.1317817111.

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)			
Contribution to the Paper			
Signature		Date	

Name of Co-Author	Ian B. Dodd		
Contribution to the Paper	Supervised project, designed experiments, analysed data, performed modelling, wrote correspondence, wrote bulk of manuscript in collaboration with David Priest		
Signature		Date	

Name of Co-Author	Keith E. Shearwin		
Contribution to the Paper	Supervised project, designed experiments, analysed data and assisted in writing the manuscript.		
Signature		Date	

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

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

David G. Priest[a], Lun Cui[a], Sandip Kumar[b], David D. Dunlap[b], Ian B. Dodd[a,1], and Keith E. Shearwin[a]

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

Edited by Mark Ptashne, Memorial Sloan Kettering Cancer Center, New York, NY, and approved November 18, 2013 (received for review September 20, 2013)

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

j factor | TPM | promoter-enhancer

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

The effect of DNA tethering can be quantified by the factor j_{LOOP} (M), the effective molar concentration of one site on the DNA relative to the other, or as the free energy of the DNA looping reaction ΔG_{LOOP} (Fig. 1) (5–9). These parameters are interconvertible: $\Delta G_{LOOP} = -RT\ln j_{LOOP}$ (kcal/mol; the reference j is 1 M). The formation of a naked DNA loop is in itself an energetically unfavorable reaction (ΔG_{LOOP} is positive) under physiological conditions due to the enthalpic cost of DNA bending and twisting (particularly important for short DNA segments) and the entropic cost of restricted configurational freedom of the DNA (the major limitation for long DNA loops). Thus, protein-mediated DNA looping reactions are driven by thermodynamic linkage to favorable protein–protein and protein–DNA interactions (Fig. 1). Nevertheless, the DNA tether can help assemble DNA–protein complexes because the effective concentration of the DNA-bound protein at the distant site, j_{LOOP}, can be greater than the concentration of available free protein (Fig. 1). Despite its critical role in DNA looping interactions, there are few measurements of j_{LOOP} in vivo, and these are mostly restricted to short site separations.

Many in vivo techniques used to detect DNA looping interactions can quantitate relative contact efficiencies but do not permit measurement of j_{LOOP}. Chromosome conformation capture (3C) and related methods have revealed a complex network of in vivo interactions, many over megabase pair DNA distances, between DNA sites in the genomes of organisms ranging from bacteria to humans (4, 10, 11). Assays using DNA recombinases have shown, as expected from DNA polymer theory, that specific contact efficiencies decrease with increasing DNA separation between the sites over the 1- to 100-kb range (12, 13). None of these methods reveal the fraction of the time that the sites are in contact.

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

Significance

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

Author contributions: D.G.P., L.C., S.K., D.D.D., I.B.D., and K.E.S. designed research; D.G.P., L.C., S.K., D.D.D., and I.B.D. performed research; D.D.D. and K.E.S. contributed new reagents/analytic tools; D.G.P., L.C., S.K., D.D.D., I.B.D., and K.E.S. analyzed data; and D.G.P., S.K., D.D.D., I.B.D., and K.E.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: ian.dodd@adelaide.edu.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1317817111/-/DCSupplemental.

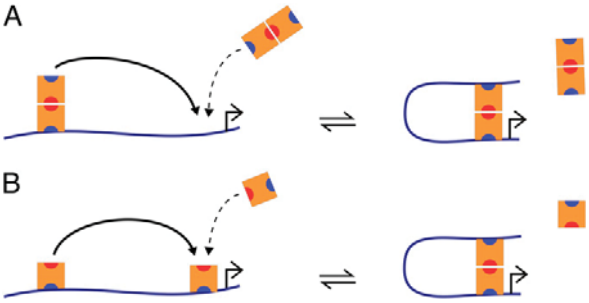


Fig. 1. The tethering effect of DNA. The assembly of a protein–DNA complex at a promoter (or other site) can be assisted if part or all of the complex is bound at a separate site on the same DNA molecule. The formation of the complex can occur more efficiently because the DNA tether can cause the effective concentration of the distally bound proteins at the target site [j_{LOOP}] (solid arrows) to be greater than their concentration in solution (dashed arrows). The formation of the complex may repress the promoter, as studied here, or may activate the promoter (e.g., when the distal site is an enhancer). The complex may be a multimer of the same protein, as shown, or may comprise different protein subunits. (*A*) Case where the protein–protein interactions (red) are very strong and loop formation is driven by protein–DNA interactions (blue), as for the Lac repressor. (*B*) Case where the protein–DNA interactions are strong and loop formation is driven by protein–protein interactions, as for the λ CI repressor.

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

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

In theory, j_{LOOP} for DNA loops formed by LacI or λ CI should be the same at large separations, where protein conformation effects are minimal. To resolve this apparent discrepancy and to extend the estimates of j_{LOOP} for long DNA separations in vivo, we performed systematic repression-based j_{LOOP} assays over the 300-bp to 10-kb range using the LacI and λ CI repressors in *E. coli*. To test whether the in vivo enhancement of looping seen at short separations holds at long separations, we measured j_{LOOP} in vitro with TPM using the Lac repressor for 600- to 3,200-bp separations with the same DNA constructs. The results for the two different repressors were similar and show that long-range DNA looping in vivo is considerably more efficient than for naked DNA in vitro, indicating long distance suppressing factors inside cells. Strategies for maximizing looping efficiency are discussed.

Results

System to Estimate j_{LOOP} Using Lac Repression in Vivo. We made chromosomally integrated *lacZ* reporters driven by the catabolite activator protein (CAP)-independent *PlacUV5* promoter controlled by a proximal *lac* operator (O_P) at the normal *O1* position (centered at +11) and an upstream distal *lac* operator (O_D; Fig. 2*A*). To characterize the system, we used three natural operators (*O1*, *O2*, and *O3*) at O_P, and three operators (*Oid*, a high affinity operator; *O1*; and a mutant operator, O^-) at O_D, with a 300-bp spacing (center to center) between O_D and O_P. Expression of *lacZ* was measured in the absence of Lac repressor (*L0*) or at two different LacI concentrations (*L1* and *L2*), expressed from $PlacI^+$.*lacI* or $PlacI^q$.*lacI* constructs integrated at a separate chromosomal site (Fig. 2*A*). The results showed, as expected, that the presence of *O1* or *Oid* at O_D increased repression substantially, depending on the operator combination and the LacI concentration (Fig. 2*B*).

By fitting the data to a statistical-mechanical model of LacI repression (Fig. 2*C*), we were able to estimate a number of in vivo parameters (Fig. 2*D*). The model is similar to Han et al. (9), with the addition of RNA polymerase (RNAP) binding to *PlacUV5* in competition with LacI (24). The cellular concentration of LacI tetramers supplied from the WT *lacI* promoter has been estimated at 18 nM (25) (11 tetramers in a 1-fL volume) and was used for *L1*. The model does not permit determination of absolute binding constants, and *K1* was arbitrarily fixed to 1 nM. The relative dissociation constants obtained for the four operators are similar to those found by Garcia and Phillips (25) (0.22, 1, 4.5, and 333 for *Kid*, *K1*, *K2*, and *K3*, respectively). Western blotting and $PlacI^+$ and $PlacI^q$ activities (Fig. S4) supported the estimate of the ratio of the high and low Lac repressor concentrations of *L2*/*L1* = 10.9. Other reporter measurements (*SI Materials and Methods*) supported the estimate of 18.1 background LacZ units from the *PlacUV5.lacZ* reporter (Fig. 2*D*).

The fitting gives an estimate of $j_{LOOP}/L1$ that is independent of the fixed values chosen for *K1* or *L1*. Assuming *L1* = 18 nM gives j_{LOOP} = 1,400 nM (for this 300-bp spacing), similar to estimates of ~1 μM for short *lac* operator spacings in vivo (6, 8, 19).

Effect of DNA Spacing on LacI Looping in Vivo. To measure j_{LOOP} over longer distances, we used *Oid-O2* reporters with operator spacings ranging from 242 to 5,600 bp. The sequence of spacer DNA can affect the efficiency of short DNA loops in vitro (26), so we made two series of spacers: series 1 (300–5,600 bp) and series 2 (242–3,200 bp). We also inserted *Oid* (or O^-) in the chromosome 500 kb away from *O2* (*SI Materials and Methods*).

The upstream operator aided repression most strongly at the shortest spacing (242 bp). Although its effect weakened with increasing spacer length, there remained a significant effect of the upstream operator at 5,600 bp at the *L1* concentration (Fig. 3*A*). The operator placed 500 kb away had no effect on repression.

The reporter data were fitted using the model and the parameters from Fig. 2*D* but allowing $j_{LOOP}/L1$ to vary for each operator spacing (*SI Materials and Methods*). These $j_{LOOP}/L1$ values (Fig. 3*B* and Fig. S5) show that the effective concentration of the operators relative to each other ranges from 180-fold the *L1* concentration of LacI in solution at 242 bp to 1.8-fold *L1* at 5,600 bp. Thus, even at a 5,600-bp separation, a Lac repressor at the distal *lac* operator is seen by the promoter almost two times more frequently than all of the other Lac repressors in the cell combined. In contrast, the data for the 500-kb spacing was fitted with a $j_{LOOP}/L1$ value of 0.09 (Fig. 3*C*); a Lac repressor bound 500 kb away contacts the proximal operator no more frequently than any of the other ~11 Lac repressors in the cell.

The $j_{LOOP}/L1$ values allow calculation of the fraction of looping for *Oid* and *O2* at each separation and [LacI] (Fig. 3*C*). LacI looping is sensitive to concentration; at low concentrations, neither operator is occupied, whereas at high concentrations, looping is obstructed by the formation of doubly bound species (species 6; Fig. 2*C*). Maximal looping decreases with increasing spacing, but even at separations of 5,600 bp, *Oid-O2* are looped at least 30% of the time over a 100-fold range of [LacI] (0.2–20 nM; Fig. 3*C*).

Effect of Separation on Long-Range Looping by λ Repressor in Vivo. We used λ CI to obtain an independent measurement of the relationship between j_{LOOP} and DNA separation in vivo. DNA looping by λ CI can be detected by repression of the λ *PRM*

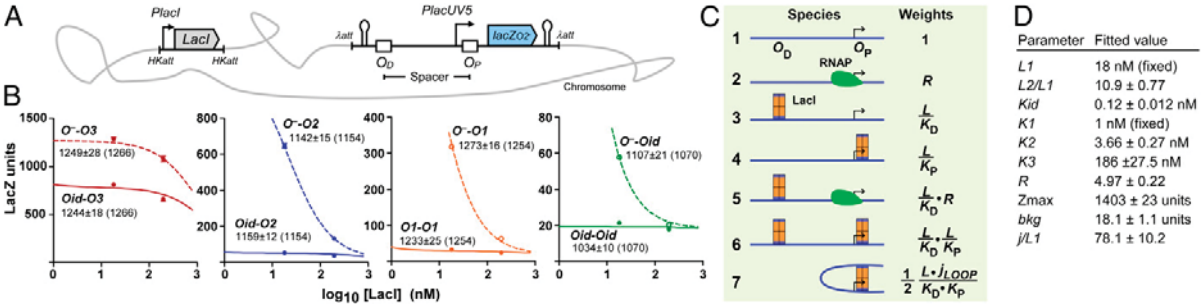


Fig. 2. Using regulation by LacI looping to measure j_{LOOP} in vivo. (*A*) Chromosomal PlacUV5.lacZ reporters with a *lac* operator at the promoter (O_P) and 300 bp upstream (O_D), with LacI supplied from a chromosomal *lacI* gene with placI$^+$ (low [LacI] = *L1*) or placI^q (high [LacI] = *L2*). (*SI Materials and Methods* and Figs. S1–S3). (*B*) Data and model fits for the four O_D–O_P combinations tested, assuming *L1* = 18 nM. The numbers on the plots are data (model) in the absence of LacI. Data errors are 95% confidence limits; $n \geq 8$. (*C*) Statistical mechanical model of regulation by LacI tetramers. *L* is the concentration of LacI tetramers, *K*s are dissociation constants, and *R* is a dimensionless constant describing RNAP occupation of the promoter. (*D*) Parameters obtained from data fitting. *Kid*, *K1*, *K2*, and *K3* are dissociation constants for *Oid*, *O1*, *O2*, and *O3*, respectively. *Zmax* is the LacZ activity obtained if the promoter were to be fully occupied by RNAP (average shown; *SI Materials and Methods*). *bkg* is the LacZ units obtained if the promoter is completely repressed. Errors are SDs from 100 data fittings (*SI Materials and Methods*).

promoter (7, 23, 27). At low concentrations, CI dimers form tetramers on *OL1.OL2* and *OR1.OR2*, activating *PRM*, and these tetramers can interact to form an octamer-bridged DNA loop (Fig. 4*A*; free energy ΔG_{oct}). The loop allows a dimer bound at *OL3* to help a dimer bind to the weak *OR3*, forming a *trans*-tetramer and repressing *PRM* (Fig. 4*A*; dimer-dimer interaction free energy ΔG_{tet}). Thus, repression at moderate CI concentrations requires the distal *OL* site. With a statistical-mechanical model for CI regulation (7, 23), the free energy for the DNA looping reaction in vivo, ΔG_{oct}, can be extracted from a comparison of *PRM* activity ± *OL* at high CI concentrations.

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

We made a series of *OR.PRM.lacZ* reporters with *OL* (or an *OL*$^-$ sequence) at distances from 150 to 10,000 bp upstream of *OR* (Fig. 4*A*, Figs. S1 and S2, and *SI Materials and Methods*). A chromosomally integrated *cI* gene (or vector only) provided 3.3 WT lysogenic units (WLUs) of CI (or no CI) (23). *OL*-dependent repression of *PRM* was strongest for the 250-bp spacing and became weaker with increasing distance (Fig. 4*B* and Fig. S6). However, even at 10,000 bp, the presence of *OL* aided repression of *PRM*. In contrast, *OL* 500 kb away on the chromosome gave no enhancement of *PRM* repression (Fig. S7).

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

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

These j_{LOOP} values produce efficient looping in the CI system, with *OL-OR* looped over 50% of the time at the 10,000-bp separation at 3.3 WLUs CI (Fig. 4*D*). Unlike the Lac system, the looped fraction increases with increasing CI concentrations (Fig. 4*D*) because the CI multimer in the looped complex, the octamer, does not form readily in solution, and loop-blocking species are only likely to form at considerably higher CI concentration (29).

Long-Range Looping by lac Repressor in Vitro. Previous studies indicate that LacI looping with short DNA tethers (<500 bp) is less efficient in vitro than in vivo. To test whether this difference holds for longer DNA tethers, we examined LacI looping in single molecules by TPM (9, 30–32), where looping of a DNA molecule attaching a bead to a surface can be detected as a restriction in the Brownian motion of the bead (Fig. 5*A*). The fragments were the same as the 600-, 900-, 1,200-, and 3,200-bp spacing in vivo constructs, but with *O1* at O_P (Fig. 5*A*, Fig. S9, and *SI Materials and Methods*).

Looped and unlooped states were followed over time over a range of LacI concentrations. Excursion values from the motion records of all selected beads in one experimental condition were assembled in a histogram from which average looping probabilities were determined (Fig. 5*B*). As expected, the probability of looping goes through a maximum with respect to [LacI] (Fig. 5*C*). At LacI concentrations at which *Oid* is fully occupied, the decrease in looping with increasing [LacI] allows estimation of j_{LOOP} independently of the affinity of *O1* (Fig. 5*D*), because looping or unlooping is a simple binding competition between bound or free LacI (9).

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

Discussion

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

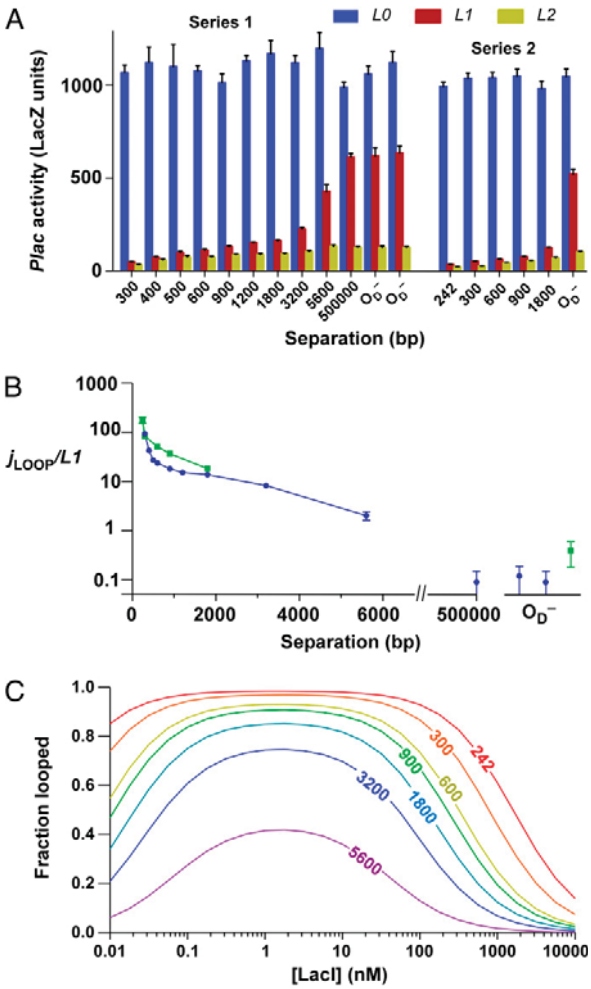


Fig. 3. Effect of DNA separation on LacI looping in vivo. (*A*) Activity of Oid-spacer-PlacUV5.O2.lacZ reporters with no (*L0*), low (*L1*), or high (*L2*) LacI levels. Errors are 95% confidence limits ($n \geq 4$). (*B*) Fitted $j_{LOOP}/L1$ values vs. spacing (*SI Materials and Methods* and Fig. S5) for the series 1 (blue) and series 2 (green) spacers. The effect of the different spacer sequences was small, with a maximal difference of approximately twofold (for the 600- and 900-bp spacers). (*C*) Model predictions of the effect of DNA separation (j_{LOOP} calculated assuming $L1 = 18$ nM) and [LacI] on the efficiency of looping. Average j_{LOOP} values from series 1 and 2 were used for 300, 600, 900, and 1,800 bp.

and λ CI were in remarkable agreement, especially considering the differences in looping mechanism, with formation of a LacI loop primarily a DNA–protein interaction (Figs. 1*A* and 2*C*) and formation of the CI loop dominated by a protein–protein interaction (Figs. 1*B* and 4*A*). The different modeling approaches and the distinct additional parameters needed to extract j_{LOOP} (*L1* for LacI and ΔG_{PTN} for CI) make the estimates quite independent of each other. In addition, as discussed, our estimates for shorter spacings agree well with previous in vivo measurements (6, 19, 20).

At separations <300 bp, j_{LOOP} was >1 μM and decayed with separation roughly as a power law with exponent –1.2 (Fig. S10). In the 5- to 10-kb separation range, j_{LOOP} was still ~20–40 nM, substantially higher than the 1.7 nM concentration of a single molecule within a cell volume of 1 fL (1 μm^3). Thus, as long as the concentration of a protein in the cell is reasonably low, even a distant operator can strongly increase its effective local concentration. As a result, even at these distances, the fraction of time that the DNA sites spend looped can be considerable. For *Oid-O2* at 5,600-bp separation, the fraction looped could be almost 40% (Fig. 3*C*). For λ *OL-OR* at 10-kb separation, the fraction looped can be at least 50% at CI concentrations >3 WLU (Fig. 4*D*).

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

In Vivo Factors Increase j_{LOOP} at Long Separations. The in vitro j_{LOOP} values obtained for LacI looping by TPM were ~5- to 45-fold lower and appeared to decay faster with separation than the in vivo values (Fig. 5*E*), fitting a power law with an exponent of –1.5 (Fig. S10).

Our TPM estimates are roughly comparable with published in vitro values. Our j_{LOOP} = 75 nM for the 600-bp LacI loop is higher than the ~10- and ~30-nM values obtained for LacI by TPM for ~100- and ~300-bp loops, respectively, by Han et al. (9), but this is consistent with the expected increase in j_{LOOP} over these separations due to relief of enthalpy costs (5). The TPM results of Johnson et al. (26) indicate even poorer looping with short spacers, with j_{LOOP} values of 0.3–4 nM for different ~100-bp LacI loops. Our 900-bp estimate of j_{LOOP} = 54 nM is similar to the j_{LOOP} = 37 nM obtained for an 870-bp loop formed by Cre recombinase (15). At the 3,200-bp spacing, our in vitro j_{LOOP} fell to ~6 nM, which is comparable to 18 nM at 3,044 bp for Cre (15) and ~10 nM for DNA cyclization at 4,000 bp (14). TPM analysis of CI looping at 2,300 bp gave j_{LOOP} = 24 nM (17), somewhat higher than our TPM values (Fig. 5*E*); however, this may be an overestimate because a low value of ΔG_{tet} was used.

Thus, we are confident that the in vivo–in vitro difference is real and applies over two orders of magnitude of the separation between sites (100–10,000 bp).

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

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

Creating Efficient Long-Range Looping. Given a fixed j_{LOOP} between any two DNA sites, how can DNA looping be maximized?

The LacI and λ CI proteins represent extremes of a biochemical continuum (Fig. 1). In the case of LacI, the protein–protein

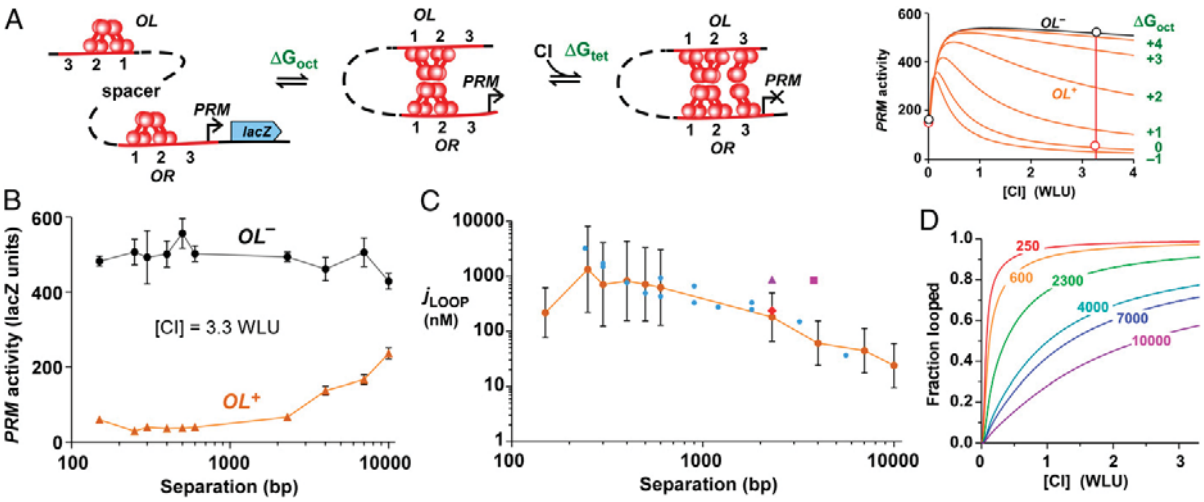


Fig. 4. Effect of DNA separation on λ CI looping in vivo. (*A*) Approach for measuring j_{LOOP} with λ CI. Structure of the *OL*-spacer-*OR.PRM.lacZ* reporters. CI activates *PRM* at low concentrations and represses it at high concentrations. Repression is dependent on the presence of *OL* and the free energy of loop formation between *OL* and *OR*, ΔG_{oct}. (*B*) Decreased repression of *PRM* with increasing *OL*-*OR* separation. *OR.PRM.lacZ* reporters ± *OL* placed various distances upstream were assayed at 3.3 WT lysogenic units (WLU) of CI and in the absence of CI (*SI Materials and Methods* and Fig. S7). Errors are 95% confidence limits; $n = 4$. (*C*) Fitted j_{LOOP} values; errors are SDs (*SI Materials and Methods* and Fig. S7). Also shown are previous j_{LOOP} estimates for in vivo CI looping (square, ref. 7; triangle, ref. 23; red diamond, ref. 20); as well as the j_{LOOP} values for LacI (Fig. 3*C*), assuming $L1 = 18$ nM (blue circles). (*D*) Predicted fraction of looping for *OL* and *OR* at different separations up to 3.3 WLU, using j_{LOOP} estimates from *C*.

interactions that connect the two DNA-binding ends of the complex are strong, and the DNA-looping multimer, the tetramer, forms at exceedingly low concentrations (37). For CI, the assembly of dimers into tetramers, and tetramers into the loop-forming octamer is relatively weak so that the DNA looping complex does not form readily in solution (29).

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

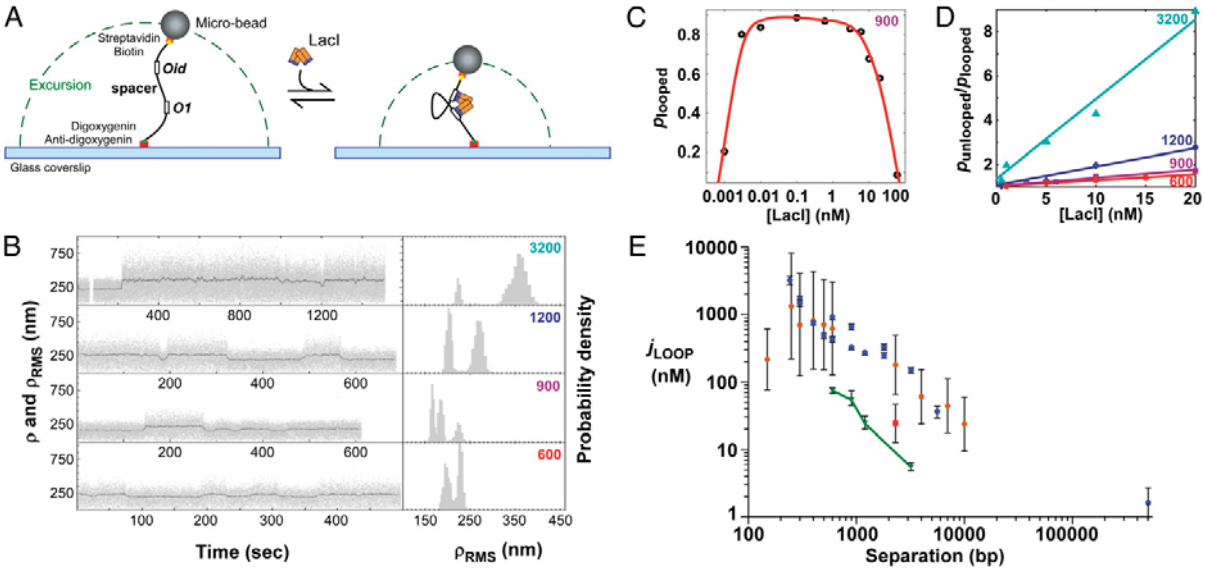


Fig. 5. Measurement of j_{LOOP} by TPM. (*A*) The TPM setup. (*B*) Examples of excursion vs. time traces for single beads, showing transitions between looped and unlooped states (*Left*) and mean excursion probability distributions (*Right*) for tethers with lac operators separated by 600–3,200 bp at 10 nM LacI. Different Lac loop conformations (two peaks) were distinguishable for the 900-bp loop. (*C*) The probability of looping vs. LacI concentration for the 900-bp separation. (*D*) j_{LOOP} for each DNA was determined from the average fraction of time the DNA is looped (p_{looped}) at each [LacI], using the [LacI] dependence of loop blockage at [LacI] 0.5–20 nM (*Oid* was filled at 0.1 nM), according to the equation $p_{unlooped}/p_{looped} = 2K1/j_{LOOP} + 2[LacI]/j_{LOOP}$ (9) (*K1* was fixed at 1 nM). The slope of each $p_{unlooped}/p_{looped}$ vs. [LacI] plot equals $2/j_{LOOP}$. (*E*) Comparison of the TPM j_{LOOP} values with the in vivo values (LacI, blue; CI, brown). The red square shows a previous TPM estimate for λ CI (17).

For CI, efficient looping requires the free energy of the protein–protein interaction, ΔG_{PTN}, to compensate for the unfavorable ΔG_{LOOP}. The interaction of two CI tetramers is estimated to provide –9.1 kcal/mol to drive the looping reaction (28, 29). However, this octamerization reaction alone would fail to give more than 50% looping for DNA separations above ~1,000 bp, where $\Delta G_{LOOP} > +9.1$ kcal/mol ($j_{LOOP} < 384$ nM). One way to improve CI looping would be to strengthen ΔG_{PTN} by increasing the strength of the interactions between protein subunits. However, such changes would tend to increase the formation of larger complexes in solution, which would block looping.

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

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

Materials and Methods

Reporter and Expression Constructs. The *lacZ* reporters and LacI and CI expression constructs were chromosomally integrated using the OSIP system and its precursors (23, 39) into MG1655 *rph*$^+$ Δ*lacIZYA* (*SI Materials and Methods*). Cells were grown at 37 °C in minimal medium (LacI looping strains) or in rich medium (CI looping strains) and assayed by a modified LacZ microtiter plate method (7) (*SI Materials and Methods*).

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

TPM. TPM experiments were conducted as previously described (30–32) with some modifications (*SI Materials and Methods*). Drift correction was improved by subtracting the motion of one to five stuck beads in the same field of view. Stringent selection of motion records based on symmetry and unlooped excursion amplitude was used. The mean square excursion of the bead was used to determine tether length by reference to a calibration curve with tethers of known length.

ACKNOWLEDGMENTS. We thank Laura Finzi, Iain Murchland, Julian Pietsch, and other members of the Shearwin, Dunlap, and Finzi laboratories, as well as Kim Sneppen, for discussions. We also acknowledge Suleyman Ocuncuoglu and Weng Lee Lim in the Emory Physics Department for enhancements to bead tracking software and Kathleen Matthews for the Lac repressor. Support was from Human Frontiers Scientific Program Grant RGP0051, a University of Adelaide PhD scholarship (to D.G.P.), the China Scholarship Council (L.C.), Australian Research Council Grants DP110100824 and DP11010470, the WH Elliott Fellowship in Biochemistry and National Health and Medical Research Council (Australia) Grant APP1025549 (to I.B.D.), National Institutes of Health Grant RGM084070A (to D.D.D.), and the Center for Pediatric Nanomedicine in the Department of Biomedical Engineering, Georgia Institute of Technology and Children's Healthcare of Atlanta (D.D.D.).

1. Ptashne M (1986) Gene regulation by proteins acting nearby and at a distance. *Nature* 322(6081):697–701.
2. Matthews KS (1992) DNA looping. *Microbiol Rev* 56(1):123–136.
3. Bulger M, Groudine M (2010) Enhancers: The abundance and function of regulatory sequences beyond promoters. *Dev Biol* 339(2):250–257.
4. Gibcus JH, Dekker J (2013) The hierarchy of the 3D genome. *Mol Cell* 49(5):773–782.
5. Rippe K (2001) Making contacts on a nucleic acid polymer. *Trends Biochem Sci* 26(12):733–740.
6. Vilar JM, Leibler S (2003) DNA looping and physical constraints on transcription regulation. *J Mol Biol* 331(5):981–989.
7. Dodd IB, et al. (2004) Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev* 18(3):344–354.
8. Zhang Y, McEwen AE, Crothers DM, Levene SD (2006) Analysis of in-vivo LacR-mediated gene repression based on the mechanics of DNA looping. *PLoS ONE* 1:e136.
9. Han L, et al. (2009) Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS ONE* 4(5):e5621.
10. Qian Z, Dimitriadis EK, Edgar R, Eswaramoorthy P, Adhya S (2012) Galactose repressor mediated intersegmental chromosomal connections in Escherichia coli. *Proc Natl Acad Sci USA* 109(28):11336–11341.
11. Cagliero C, Grand RS, Jones MB, Jin DJ, O'Sullivan JM (2013) Genome conformation capture reveals that the Escherichia coli chromosome is organized by replication and transcription. *Nucleic Acids Res* 41(12):6058–6071.
12. Ringrose L, Chabanis S, Angrand PO, Woodroofe C, Stewart AF (1999) Quantitative comparison of DNA looping in vitro and in vivo: Chromatin increases effective DNA flexibility at short distances. *EMBO J* 18(23):6630–6641.
13. Stein RA, Deng S, Higgins NP (2005) Measuring chromosome dynamics on different time scales using resolvases with varying half-lives. *Mol Microbiol* 56(4):1049–1061.
14. Shore D, Langowski J, Baldwin RL (1981) DNA flexibility studied by covalent closure of short fragments into circles. *Proc Natl Acad Sci USA* 78(8):4833–4837.
15. Shoura MJ, et al. (2012) Measurements of DNA-loop formation via Cre-mediated recombination. *Nucleic Acids Res* 40(15):7452–7464.
16. Hsieh WT, Whitson PA, Matthews KS, Wells RD (1987) Influence of sequence and distance between two operators on interaction with the lac repressor. *J Biol Chem* 262(30):14583–14591.
17. Zurla C, et al. (2009) Direct demonstration and quantification of long-range DNA looping by the lambda bacteriophage repressor. *Nucleic Acids Res* 37(9):2789–2795.
18. Mossing MC, Record MT, Jr. (1986) Upstream operators enhance repression of the lac promoter. *Science* 233(4766):889–892.
19. Vilar JM, Saiz L (2005) DNA looping in gene regulation: From the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev* 15(2):136–144.
20. Hensel Z, Weng X, Lagda AC, Xiao J (2013) Transcription-factor-mediated DNA looping probed by high-resolution, single-molecule imaging in live E. coli cells. *PLoS Biol* 11(6):e1001591.
21. Müller J, Oehler S, Müller-Hill B (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J Mol Biol* 257(1):21–29.
22. Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, 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.
23. Cui L, Murchland I, Shearwin KE, Dodd IB (2013) Enhancer-like long-range transcriptional activation by λ CI-mediated DNA looping. *Proc Natl Acad Sci USA* 110(8):2922–2927.
24. Sanchez A, Osborne ML, Friedman LJ, Kondev J, Gelles J (2011) Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. *EMBO J* 30(19):3940–3946.
25. Garcia HG, Phillips R (2011) Quantitative dissection of the simple repression input-output function. *Proc Natl Acad Sci USA* 108(29):12173–12178.
26. Johnson S, Lindén M, Phillips R (2012) Sequence dependence of transcription factor-mediated DNA looping. *Nucleic Acids Res* 40(16):7728–7738.
27. Dodd IB, Perkins AJ, Tsemitsidis D, Egan JB (2001) Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. *Genes Dev* 15(22):3013–3022.
28. Rusinova E, Ross JB, Laue TM, Sowers LC, Senear DF (1997) Linkage between operator binding and dimer to octamer self-assembly of bacteriophage lambda cI repressor. *Biochemistry* 36(42):12994–13003.
29. Senear DF, et al. (1993) The primary self-assembly reaction of bacteriophage lambda cI repressor dimers is to octamer. *Biochemistry* 32(24):6179–6189.
30. Finzi L, Dunlap D (2003) Single-molecule studies of DNA architectural changes induced by regulatory proteins. *Methods Enzymol* 370:369–378.
31. Nelson PC, et al. (2006) Tethered particle motion as a diagnostic of DNA tether length. *J Phys Chem B* 110(34):17260–17267.
32. Dunlap D, Zurla C, Manzo C, Finzi L (2011) in *Methods in Molecular Biology*, eds Peterman EJG, Wuite GJL (Humana Press, New York), Vol 783, pp 295–313.
33. Becker NA, Kahn JD, Maher LJ, 3rd (2007) Effects of nucleoid proteins on DNA repression loop formation in Escherichia coli. *Nucleic Acids Res* 35(12):3988–4000.
34. Czapla L, Swigon D, Olson WK (2008) Effects of the nucleoid protein HU on the structure, flexibility, and ring-closure properties of DNA deduced from Monte Carlo simulations. *J Mol Biol* 382(2):353–370.
35. Dillon SC, Dorman CJ (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* 8(3):185–195.
36. Olson WK, Grosner MA, Czapla L, Swigon D (2013) Structural insights into the role of architectural proteins in DNA looping deduced from computer simulations. *Biochem Soc Trans* 41(2):559–564.
37. Barry JK, Matthews KS (1999) Thermodynamic analysis of unfolding and dissociation in lactose repressor protein. *Biochemistry* 38(20):6520–6528.
38. Dröge P, Müller-Hill B (2001) High local protein concentrations at promoters: Strategies in prokaryotic and eukaryotic cells. *Bioessays* 23(2):179–183.
39. St-Pierre F, et al. (2013) One-step cloning and chromosomal integration of DNA. *ACS Synth Biol* 2(9):537–541.

Supporting Information

Priest et al. 10.1073/pnas.1317817111

SI Materials and Methods

DNA Constructions. The parent strain for all reporter assays was E4643, which was constructed from BW30270 (CGSC7925) MG1655 *rph*$^+$ by precise deletion of *lacIZYA* (EcoCyc MG1655: 360527–366797) by recombineering (1). EC100D *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80d*lacZ*ΔM15 Δ*lac*X74 *recA1endA1 araD139* Δ(*ara,leu*)7697 *galU galK* λ$^-$ *rpsL nupG pir*$^+$ (DHFR; Epicentre) was used for propagation of R6γK *ori* (*pir*-dependent) plasmids.

DNA constructions used commercial DNA synthesis (GenScript; gBlocks from Integrated DNA Technologies; primers from Geneworks), restriction enzyme–based cloning, and isothermal Gibson assembly (2). DNA sequences of manipulated regions were confirmed, except for some of the larger spacers, the sizes of which were confirmed by PCR.

The reporter and LacI and CI expression constructs (Figs. S1–S3) were made using a plasmid integration system developed from the CRIM plasmids (3), the pZ plasmids (4), and an *O2*$^-$ *lacZ* reporter gene (5) preceded by an RNaseIII cleavage site (6). As in the CRIM system, phage integrase proteins were used to integrate the plasmids into phage attachment sites in the bacterial chromosome. Integration was at λ*attB* (EcoCyc (7) MG1655 sequence position: 806551), ϕHK022 *attB* (position 1055419), ϕ80 *attB* (position 1308595), or ϕ186 *attB* site (position: 2783828) using original or modified CRIM integrase plasmids. PCR was used to screen for correct single-copy integrants. All sequences are available on request.

Reporter Constructs. To measure DNA loop-dependent repression, a modular DNA looping *lacZ* reporter "chassis" was designed to be amenable to both restriction enzyme–based cloning and Gibson isothermal assembly (Fig. S1). The various chassis structures were inserted into pIT-HF-CL.lacZ and were integrated into the λ*attB* site. LacZ expression was from *PlacUV5.O* or *OR.PRM* modules, with various *lac* or λ CI operator modules (Fig. S2) located elsewhere on the chassis. The chassis is invertible; when module 1 points toward *lacZ*, LacI looping is measured by repression of *PlacUV5*, whereas in the inverted orientation, module 4 points toward *lacZ* and λ CI looping can be measured by repression of *PRM*.

Spacer DNA between the operator/promoter modules (Fig. S1) was made up of sequences from within the *Escherichia coli* genes *ftsK* (position 932456–936438), *rne* (114410-1143589), and *valS* (4479008-4481858) to minimize the likelihood of incorporation of cryptic promoters.

Control tests showed that *PLacUV5.O*$^-$ reporters were not responsive to LacI either in the absence of any *lac* operators in the looping chassis or with *Oid* situated 300 bp upstream.

Background LacZ units that are refractory to LacI repression were estimated by the measuring the activity of the Series 1 O1.300. O1 reporter in the *L2* strain in the presence of multiple copies of the *lacI* gene on the pUHA1 plasmid (4). The value of 20.6 units is close to the 18.1 units estimated by modeling (Fig. 2*D*).

Close to twofold differences in the effect of the upstream *Oid* operator and j_{LOOP} values were seen between series 1 and series 2 reporters with 600- and 900-bp operator separations (Fig. 3 *B* and *C*), presumably resulting from differences in the interoperator sequence. As shown in Fig. S1*A*, the main differences between the two series are the spacer (*valS*, series 1; *ftsK*, series 2) and terminators (λtI, series 1; 186 Tw, series 2). Series 1 reporters showed less repression, and it is possible that the valS or λtI sequences contain an element that is refractory to DNA looping (e.g., intrinsically bent DNA or a binding site for a DNA-bending protein). An alternative explanation is that the *valS* spacer in series 1 could contain a cryptic promoter facing toward *lacZ* that is strong enough to drive polymerase through the λtI terminator and the LacI roadblock at O2 and contribute to the *PLacUV5* measurement.

Reporter with *lac Oid* at 500-kb Spacing. pOSIP-KP (a highly modified CRIM plasmid) (8) carrying *Oid* or *O*$^-$ was integrated into the *attφ80* site in E4643 carrying the series 1 *O*$^-$0.300.*PlacUV5.O2.lacZ* reporter (Fig. S1*B*) at *attλ*. In these integrants, the FLP-excisable integration and Kan^R module was removed through transformation of a FLP recombinase-expressing plasmid (8).

LacI Expression Constructs. DNA fragments carrying *lacI* and its natural *PlacI* promoter or the ∼10-fold stronger *PlacI*^{q} promoter mutant were obtained from the pUHA1 or pDM1.1 plasmids, respectively (4), by digestion with SalI. The fragments extended from the –78 position of *PlacI* (366837) to 18 bp downstream of the *lacI* stop codon (365634; includes *O3*) and were inserted into the SalI site of pIT3-SH (Fig. S3). Single-copy integrants of pIT3-SH, pIT3-SH-lacI$^+$, or pIT3-SH-lacI^q at the φHK022 att site were used to provide LacI at the *L0*, *L1*, and *L2* concentrations, respectively.

CI Expression Constructs. The λ CI was expressed from pIT3-TO-λcI-OL3-4 (9) integrated at 186 *attB* and containing a *PRM.cI.OL* module that produces 3.3 ± 0.53 WT lysogenic units (WLUs) of CI (9). The control strain expressing no CI contained the empty vector pIT3-TO at 186 *attB*. The copies of *OL* and *OR* present on the expression construct do not influence the reporter at λ*attB* by DNA looping because they are located 1.98 Mbp away.

LacI Westerns. Cultures of the LacI-expressing strains (*L0*, *L1*, and *L2*) were grown to OD_{600} = 0.4 in L broth (1% Bacto-tryptone, 1% NaCl, and 0.5% yeast extract, pH 7.0) and resuspended at 1/40 culture volume in B-Per lysis buffer (Pierce) containing 0.25 U/μL Benzonase (Novagen) and incubated at 4 °C for 15 min to lyse the cells and digest nucleic acids; 4× NuPAGE SDS loading buffer (Invitrogen) was added, and tubes were heated to 70 °C for 10 min. Equal volumes of SDS samples were run on 4–12% (wt/vol) Bis Tris gels with Mops running buffer (Novex) and transferred to 0.2-μm PVDF membranes using an iBlot transfer apparatus (Invitrogen). Membranes were blocked in 5% (wt/vol) BSA and then incubated with a 1/200 dilution of polyclonal rabbit anti-LacI antibody (Rockland) preadsorbed against an extract of the parental strain (E4643), followed by a 1/4,000 dilution of Cy5-labeled goat-anti-rabbit 2° antibody (GE Healthcare). Membranes were dried and imaged on the Cy5 channel of a GE Typhoon imager (Amersham Biosciences). Western images were analyzed using ImageJ software, and the intensity of *L1* LacI bands was estimated using the dilution series of *L2* into *L0* extracts as a standard (Fig. S4*A*).

***lacI* Translational Fusions.** The plasmid pRS414 (10) is for making translational fusions to *lacZ*; however, the *lacZ* gene is *O2*$^+$. Thus, the BamHI/Bsu36I fragment from pRS414 was cloned into pIT3-CL-lacZtrim (Fig. S4*B*) to give pIT3-CL-lacZtrimfuse, where *lacZ* is *O2*$^-$. A 227 bp PCR fragment from pUHA1, encompassing the promoter region (up to position –119), the ribosome binding site, and the first 15 codons of *lacI* (*PlacI.lacI*) and pDM1.1 (*PlacI*^{q}*.lacI*), was cloned into the Acc65I/BamHI site of pIT3-CL-LacZtrimfuse, and the plasmid was integrated into λ*att* site in E4643.

LacZ Assays. LacZ assays for the CI looping reporters were carried out using the microtiter plate method (11), with the modification that cultures grown in L broth at 37 °C to late log phase were added to a combined lysis-assay buffer, with each well of a microtiter plate containing: 50 μL culture + LB (usually 20 μL culture + 30 μL LB), 150 μL TZ8 (100 mM Tris·HCl, pH 8.0, 1 mM $MgSO_4$, 10 mM KCl), 40 μL *o*-nitrophenyl-β-D-galactoside (ONPG; 4 mg/mL in TZ8), 1.9 μL 2-mercapoethanol, and 0.95 μL polymyxin B (20 mg/mL; Sigma). Assays were repeated at least four times.

The Lac looping reporters were assayed as above, with the modification that cultures were grown in M9 minimal medium [1× M9 salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.01 mM $(NH_4)_2Fe(SO_4)_2{\cdot}6\ H_2O$, and 0.4% glycerol] (10× M9 salts = 67.8 g of NaH_2PO_4, 30.0 g of KH_2PO_4, and 5 g NaCl/L H_2O). We found that robust growth of our strains in minimal medium required an iron source (12).

Modeling of Looping in Vivo. ***Estimation of lac system parameters from the 300-bp spacing reporter data.*** Following the results of Sanchez et al. (13) that indicate mutually exclusive binding of LacI at *O1* and RNAP at the promoter, we allowed only two RNA polymerase (RNAP)-bound species: one with O_D empty and one with O_D occupied. Different forms of RNAP–promoter complexes were not distinguished and, because [RNAP] was not varied, RNAP occupation was described by a dimensionless constant, *R*.

The reporter data for the eight different (O_D-300-O_P) operator combinations (Fig. 2*B*; O^--*O3*, *Oid-O3*, O^--*O2*, *Oid-O2*, O^--*Oid*, *Oid-Oid*, O^--*O1*, *Oid-O1*) were globally fitted using the model of Fig. 2*C* (the O^- data were fitted to the model with species 1, 2, and 4 only). Promoter activity at each [LacI] (tetramer) was taken to be proportional to the statistical weight of species with RNAP bound at the promoter and was calculated as $bkg + Zmax \times (R + R \times L/K_D)/Z$, where *L* is [LacI], K_D is the LacI dissociation constant at the distal operator, and *Z* is the partition sum. The maximal activity (*Zmax*) is the activity of the promoter in LacZ units if it were fully occupied by RNAP; the basal activity of the promoter (in the absence of LacI) is thus $bkg + Zmax \times R/(1 + R)$. Different *Zmax* values were allowed for each *Plac.O_P* combination, because the sequence differences at the proximal operator could conceivably affect promoter activity; however, the four fitted values for *Zmax* (1,263 ± 18, 1,485 ± 29, 1,364 ± 18, and 1,500 ± 26 for *Oid*, *O1*, *O2*, and *O3*, respectively) were within 10% of their mean, indicating little effect of the sequence changes on the promoter. The *L1* and *K1* parameters were held fixed at 18 and 1 nM, respectively.

The fitting algorithm minimized $\Sigma[(observed - expected)^2/expected]$, comparing the model-generated expected promoter LacZ values with the 24 observed LacZ data points of Fig. 2*B*, giving equal weight to all points. In a Monte Carlo algorithm, parameter values were iteratively varied by successive random steps, with each modified parameter set accepted if the fit improved. Fitting convergence was tested by repeating the fitting process with different randomly chosen initial parameter values and was found to be highly reliable.

The fitted values for the dissociation constants (*K1*, *K2*, *K3*, *Kid*) and *R* are dependent on the fixed values chosen for *K1* and *L1*; thus, the model can only resolve relative in vivo binding strengths for the different operators. The fitted value for *j*/*L1* is independent of the fixed value chosen for *K1*, *L1*, or *R*.

To gain an appreciation of the error in the estimates of the LacI system parameters (Fig. 2*D*), we repeated the fitting process with "jiggled" data; that is, each of the 24 data points used in the fitting (Fig. 2*B*) was changed randomly based on the *t*-distribution according to its measured SEM and number of observations (*n*). By making many sets of jiggled data and doing parameter fitting to each of these, we sampled how variation in the data and variation in the parameter fitting procedure combine to affect the parameter estimates. The parameter values in Fig. 2*D* are the means and SDs from 100 good-scoring fits.

Estimating j/L1 from the variable spacing lac reporter data. Each of the reporter data sets (series 1 or 2) from the *Oid-O2* constructs with different DNA spacers (Fig. S5 and Fig. 3*A*) were globally fitted by fixing the parameter estimates for *R*, *L2*/*L1*, *Kid*, *K2*, and *bkg* to their values in Fig. 2*D* and by allowing different *j*/*L1* values for each construct. A single *Zmax2* (maximal activity of *PlacUV5.O2*) was also fitted for each series (which for series 1 included pooled results from two O^- constructs; Fig. S5).

As before, we performed the same data jiggling procedure, with the data means randomly shifted based on their SEM and *n*. We also incorporated the uncertainty in the fixed parameter estimates by randomly jiggling their values based on their SDs (Fig. 2*D*) and the normal distribution. Thus, each fitting run used slightly different parameter values and data means. The *j*/*L1* estimates for each spacing (Fig. S5) were obtained from 90 fitting runs that produced good-scoring fits.

Estimating j_{LOOP} from the variable spacing λ CI reporter data. Four data points, OL^+ or OL^- and [CI] = 0 or 3.3 WLU, were collected for each of the 10 spacings (150–10,000 bp; Fig. S6*A*). We expect that spacing will only affect the OL^+ [CI] = 3.3 values. However, we noticed that the OL^- [CI] = 0, OL^- [CI] = 3.3, and OL^+ [CI] = 0 values for the shorter spacings (150–500 bp) were systematically lower than those for the longer spacings (600–10,000 bp; Fig. S6*A*), due most likely to day-to-day variations in the LacZ assay. For the fitting, we therefore normalized the data by dividing the four data points (and SEMs) for each spacing by a factor (0.88–1.17, depending on the spacing) to compensate for this systematic error. These normalized data are shown in Fig. S6*C* and were used for Fig. 4*B*.

The data from the 10 different spacings were fitted globally, using the means of the OL^- [CI] = 0, OL^- [CI] = 3.3, and OL^+ [CI] = 0 values and each individual OL^+ [CI] = 3.3 value. All parameters were fixed except for *max_prm* (the maximal activity of *PRM*), which was applied to all spacings, and ΔG_{oct}, which allowed individual values for each spacing. The parameter values (Fig. S6*B*) were essentially as used in our previous study (9), except for ΔG_{tet}, as discussed in the main text. The change to ΔG_{tet} resulted in slight changes to certain parameters to optimize the fit to our previous data set (9). The list of species in the λ model is given in Fig. S6*D*; note that because the constructs did not contain the UP element near *OL*, and *PR* was inactive, species involving these elements were ignored. The fitting procedure was as described previously (9) and is essentially the same as for the LacI system modeling. To gain an idea of the error in the ΔG_{oct} estimates, both the fixed parameter values and the data means were jiggled, as for the LacI system.

The obtained ΔG_{oct} values for each spacing were converted to j_{LOOP} as follows:

$$j_{LOOP}(\mathrm{M}) = \exp[-(\Delta G_{oct} + 9.1)/RT]$$

where RT = 0.616 kcal/mol.

Programs were written in Fortran 77 and run in a Cygwin environment on a laptop computer and are available on request.

Tethered Particle Motion Methods. ***DNA preparation.*** DNAs were prepared by Taq PCR with biotin- and digoxygenin-labeled oligonucleotides (Invitrogen, Life Technologies Corporation or Integrated DNA Technologies), using plasmid templates with the series 1 looping chassis (Fig. S1; except with *O1* at O_P) cloned into pUC57. All amplicons were purified by using silica membrane–based purification kits (Qiagen), and the lengths were checked by gel electrophoresis. The details of the plasmids and primers are available on request.

Tethered particle motion experiments. Tethered particle motion (TPM) experiments were conducted as previously described

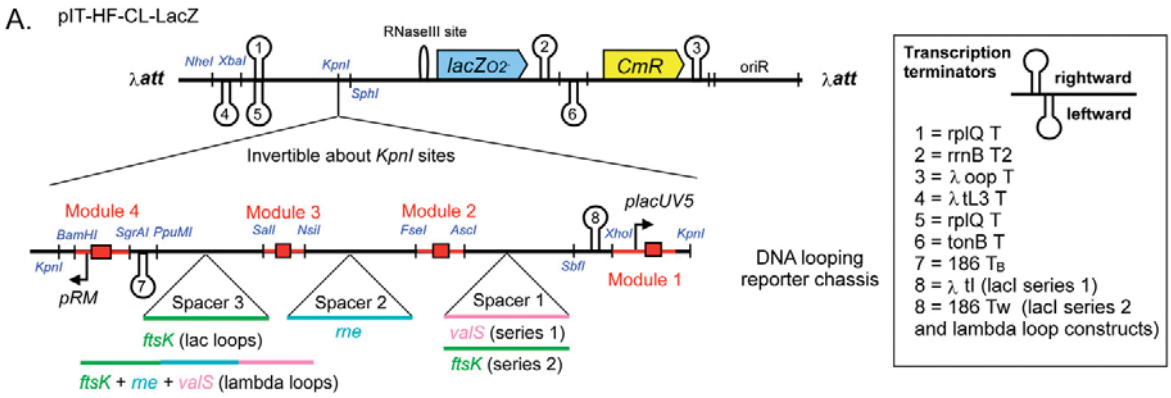
(14–17) with some modifications (18). Chambers were assembled between a coverslip and a microscope slide using a parafilm spacer. To attach tethers to the glass surface, 50–100 pM of DNA labeled with digoxigenin on one end and biotin on the other end (Fig. 5*A*) was incubated in binding buffer (10 mM Tris·HCl, pH 7.4, and 200 mM KCl) for 15 min. After gentle flushing with three volumes of wash buffer (10 mM Tris·HCl, pH 7.4, 200 mM KCl, and 0.5m g/mL α-casein), a solution of 0.5–1 nM streptavidin-coated polystyrene beads (160-nm radius; Spherotech) in binding buffer was introduced to incubate for 10 min. The chamber was gently flushed with three volumes of wash buffer, two of λ buffer (10 mM Tris·HCl, pH 7.4, 200 mM KCl, 5% DMSO, 0.1 mM EDTA, 0.2 mM DTT, and 0.1 mg/mL α-casein), and then with 2 volumes of LacI in λ buffer to make sure that the chamber had the desired protein and buffer composition. The chamber was incubated for 30 min on the microscope, and then 3–5 immobile beads and 3–20 mobile beads were tracked. Further details about the instrumentation and the analysis are available (18).

Data preprocessing: drift calculations and symmetry selection. Only immobile beads that were tracked for the entire recording were used for drift calculations. A 4-s moving average (center of mass) for selected immobile beads in a video frame was calculated and subtracted from each bead in the same field of view to remove the drift. Immobile beads exhibit very low amplitude motion, and a very short time averaging window (4 s as opposed to 40 s for mobile beads) can determine the anchor point of the bead very accurately. Such short time averaging increases the cutoff frequency of drift motion that can be subtracted while minimizing attenuation of the motion of tethered beads. The drift-corrected recordings of tethered beads were then analyzed and selected for symmetry before further analysis.

Analysis. The mean square excursion of the bead was calculated using the formula $\rho_{8s}^2 = [(x - x_{8s})^2 + (y - y_{8s})^2]_{8s}$. A long time window for averaging, 8 s, was necessary to improve the resolution of looped and unlooped states in the case of 600-bp loops. In plots of ρ and $\sqrt{\rho_{8s}^2}$ vs. time for individual beads (Fig. 5*B*), steps in the amplitude signal loop formation and breakdown and the histogram of the mean excursion shos the time spent in each state for DNA tethers that form 600-, 900-, 1,200-, or 3,200-bp loops. A calibration curve (Fig. S9*A*) was used to verify the magnitude of the looped and unlooped excursion for each recording and select only the recordings with expected excursions. All selected recordings in the same experiment condition were concatenated, and an overall histogram was plotted. A threshold was determined, and time points with excursion smaller than the threshold were taken to be looped, whereas time points with a higher excursion were considered unlooped. The looping probability for different LacI concentrations was determined by dividing the time spent in a looped state by the total observation time. The linear fits used to determine j_{LOOP} from the looping probabilities at each [LacI] (Fig. 5*C*) are shown in Fig. S9*B*.

1. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA* 97(12):6640–6645.
2. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343–345.
3. Haldimann A, Wanner BL (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol* 183(21):6384–6393.
4. Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* 25(6):1203–1210.
5. Müller J, Oehler S, Müller-Hill B (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J Mol Biol* 257(1):21–29.
6. Linn T, St Pierre R (1990) Improved vector system for constructing transcriptional fusions that ensures independent translation of *lacZ*. *J Bacteriol* 172(2):1077–1084.
7. Keseler IM, et al. (2011) EcoCyc: A comprehensive database of Escherichia coli biology. *Nucleic Acids Res* 39(Database issue):D583–D590.
8. St-Pierre F, et al. (2013) One-step cloning and chromosomal integration of DNA. *ACS Synth Biol* 2(9):537–541.
9. Cui L, Murchland I, Shearwin KE, Dodd IB (2013) Enhancer-like long-range transcriptional activation by λ CI-mediated DNA looping. *Proc Natl Acad Sci USA* 110(8):2922–2927.
10. Simons RW, Houman F, Kleckner N (1987) Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene* 53(1):85–96.
11. Dodd IB, Perkins AJ, Tsemitsidis D, Egan JB (2001) Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. *Genes Dev* 15(22):3013–3022.
12. Paliy O, Gunasekera TS (2007) Growth of E. coli BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Appl Microbiol Biotechnol* 73(5): 1169–1172.
13. Sanchez A, Osborne ML, Friedman LJ, Kondev J, Gelles J (2011) Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. *EMBO J* 30(19):3940–3946.
14. Finzi L, Dunlap D (2003) Single-molecule studies of DNA architectural changes induced by regulatory proteins. *Methods Enzymol* 370:369–378.
15. Nelson PC, et al. (2006) Tethered particle motion as a diagnostic of DNA tether length. *J Phys Chem B* 110(34):17260–17267.
16. Dunlap D, Zurla C, Manzo C, Finzi L (2011) in *Methods in Molecular Biology*, eds Peterman EJG, Wuite GJL (Humana Press, New York, NY USA), Vol 783, pp 295–313.
17. Gao N, Shearwin K, Mack J, Finzi L, Dunlap D (2013) Purification of bacteriophage lambda repressor. *Protein Expr Purif* 91(1):30–36.
18. Kumar S, et al. (2013) Enhanced tethered particle motion analysis reveals viscous effects. *Biophys J*, in press.

A.



B.

Lac Looping Constructs (series 1)

Spacing (bp)	Module 4 (pRM-)	Spacer 3 (bp)	Module 3	Spacer 2 (bp)	Module 2	Spacer 1	Module 1
O_D-	O-	300	O-	300	O-	300	*pLacO2* (or *O3*, *O1*, *Oid*)
300	O-	300	O-	300	*Oid* (or *O1*)	300	*pLacO2* (or *O3*, *O1*, *Oid*)
400	O-	300	O-	300	*Oid*	400	*pLacO2*
500	O-	300	O-	300	*Oid*	500	*pLacO2*
600	O-	300	*Oid*	300	O-	300	*pLacO2*
900	*Oid*	300	O-	300	O-	300	*pLacO2*
1200	O-	300	*Oid*	600	O-	600	*pLacO2*
1800	O-	300	*Oid*	900	O-	900	*pLacO2*
3200	O-	300	*Oid*	1400	O-	1800	*pLacO2*
O_D-	O-	300	O-	1400	O-	1800	*pLacO2*
5600	*Oid*	Long spacer (as per lambda spacer 3)					*pLacO2*

Lac Looping Constructs (series 2)

Spacing (bp)	Module 4 (pRM-)	Spacer 3 (bp)	Module 3	Spacer 2 (bp)	Module 2	Spacer 1	Module 1
O_D-	O-	222, no *ftsK*	O-	142, no *rne*	O-	600	*pLacO2*
242	O-	222	O-	142	*Oid*	242	*pLacO2*
300	O-	222	O-	142	*Oid*	300	*pLacO2*
600	O-	222	O-	142	*Oid*	600	*pLacO2*
900	O-	222	O-	142	*Oid*	900	*pLacO2*
1800	O-	222	O-	142	*Oid*	1800	*pLacO2*

Lambda Looping Constructs

Spacing (bp)	Module 4	Spacer 3 (bp)	Module 3	Spacer 2 (bp)	Module 2	Spacer 1 (bp)	Module 1 (pLac-)
150	*oR+(pR-)123*	150	*oL+(pL-)321*	142, no rne	O-	242	O-
250	*oR+(pR-)123*	250	*oL+(pL-)321*	142	O-	242	O-
300	*oR+(pR-)123*	300	*oL+(pL-)321*	142	O-	242	O-
400	*oR+(pR-)123*	400	*oL+(pL-)321*	142	O-	242	O-
500	*oR+(pR-)123*	500	*oL+(pL-)321*	142	O-	242	O-
600	*oR+(pR-)123*	600	*oL+(pL-)321*	142	O-	242	O-
2300	*oR+(pR-)123*	2300	*oL+(pL-)321*	142	O-	242	O-
4000	*oR+(pR-)123*	4000	*oL+(pL-)321*	142	O-	242	O-
7000	*oR+(pR-)123*	7000	*oL+(pL-)321*	142	O-	242	O-
10000	*oR+(pR-)123*	10000	*oL+(pL-)321*	142	O-	242	O-

Fig. S1. Reporter constructs. (*A*) Reporter vector and chassis. The basic structure of the integrated reporter vector and inserted looping chassis is shown, together with relevant restriction sites and transcription terminators. Modules contained operators (red boxes) for either LacI or CI or a standard operator minus sequence (O^-). (*B*) Detailed maps of the reporter constructs. The blue shading indicates the looped region. Spacer lengths are measured between the centers of the operators within the modules (*OL2* and *OR2* for λ).

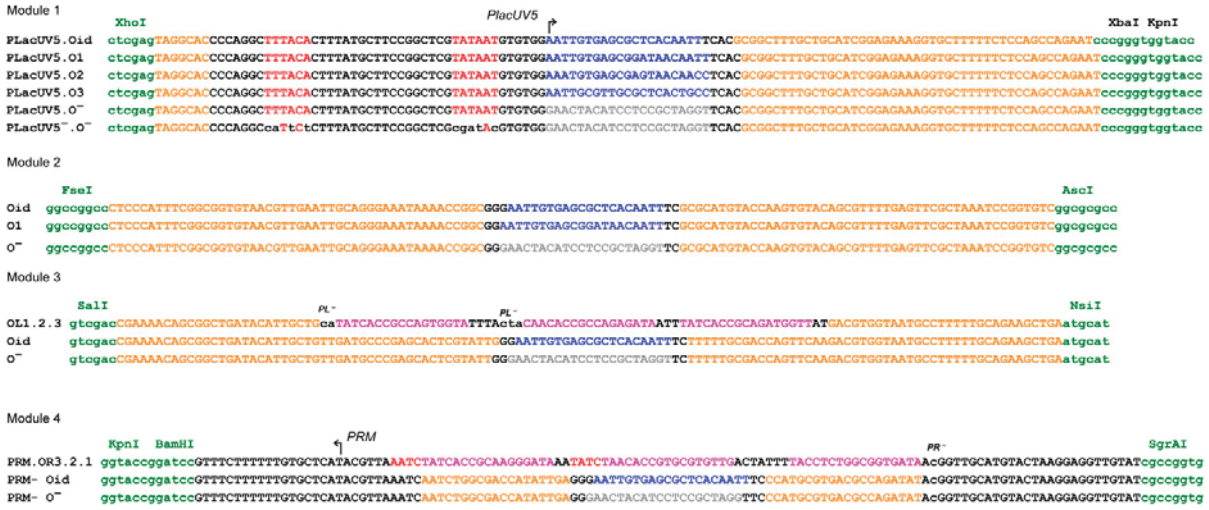


Fig. S2. Sequences of reporter modules. Each 120- or 121-bp module is listed in the orientation shown in Fig. S1 and flanked by chassis restriction sites (lowercase green). Non-*lac* or λ sequences are in brown. *Lac* operators are shown in blue (*lacO*⁻ in gray) and CI operators in purple. Promoter sequences (–10 and –35) for *PlacUV5* and *PRM* are in red; lowercase bases are mutations used to inactivate promoters.

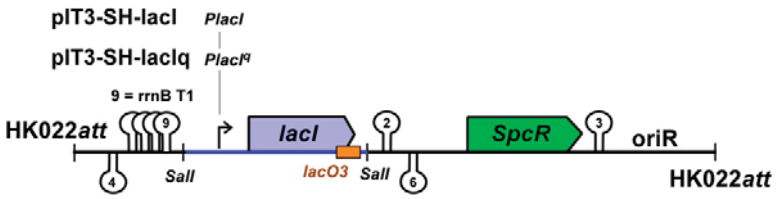


Fig. S3. LacI expression constructs. Terminators are as in Fig. S1. SpcR, spectinomycin resistance.

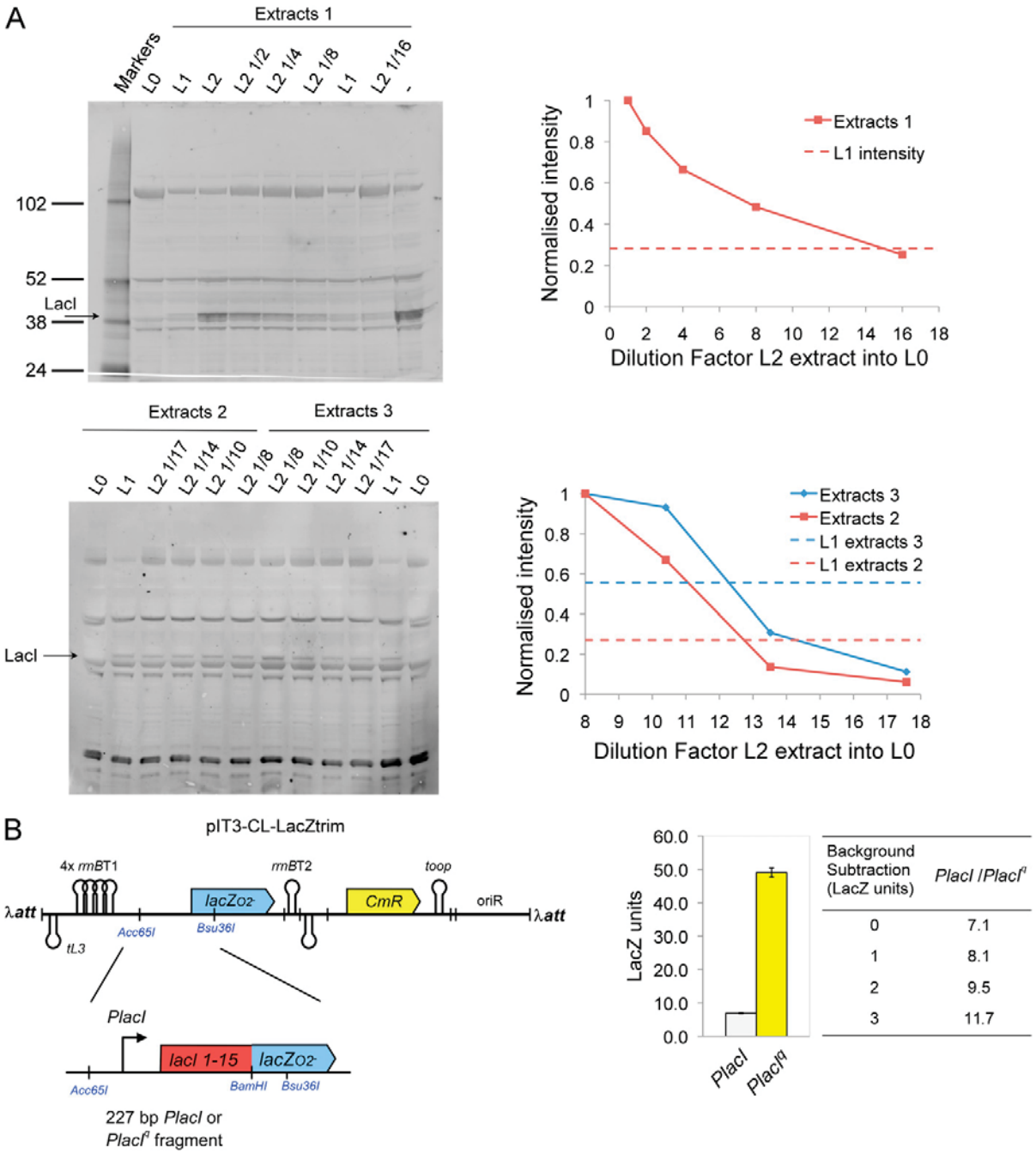


Fig. S4. Experimental estimates of *L2*/*L1* for the high- and low-expressing LacI strains. (*A*) Extracts of cells with no LacI (*L0*), low LacI (*L1*), or high LacI (*L2*) levels were analyzed by Western blotting using LacI antibody (*SI Materials and Methods*). *L2* extracts were diluted in *L0* extract as indicated. The *L2*/*L1* ratio was estimated as 15.5 (upper blot) and 12.5 (lower blot; 12.25 and 12.75), giving an average over three extracts: *L2*/*L1* = 13.5. (*B*) LacZ assay results are shown for the *PlacI* (*L1*) and *PlacI*^{q} (*L2*) translational fusion reporters (*SI Materials and Methods*) ±95% confidence limits (n = 15, 16). The ratio of LacI expression from the *PlacI* and *PlacI*^{q} promoters is dependent on an unknown level of background LacZ units that could come from extraneous promoters. Ratios *PlacI*^{q}/*PlacI*, calculated after subtracting 0–3 units of background are shown and should approximate *L2*/*L1*.

A

Series	Spacing	Average LacZ units			Standard deviation			No. values			Fitted	SD
		L0	*L1*	*L2*	*L0*	*L1*	*L2*	*L0*	*L1*	*L2*	j_{LOOP}/LI	
1	O–	1059.9	619.6	127.5	69.82	69.16	13.39	12	12	12	0.12	0.07
1	O–	1122.9	636.9	127.3	104.56	65.88	8.94	15	16	16	0.09	0.06
1	300	1067.7	49.5	35.1	64.15	2.55	2.53	12	12	12	94.38	9.11
1	400	1123.1	74.8	59.4	51.92	2.39	3.49	4	4	4	42.52	3.78
1	500	1100.0	99.7	74.7	74.89	4.36	5.96	4	4	4	27.30	2.63
1	600	1077.7	112.5	75.7	44.33	10.09	6.17	12	12	12	23.92	2.29
1	900	1013.5	132.3	88.7	76.30	7.81	5.26	12	12	12	18.28	1.51
1	1200	1131.1	153.3	88.5	35.38	4.28	8.80	8	8	8	15.26	1.35
1	1800	1169.8	163.3	93.8	128.66	11.51	6.76	15	16	16	13.80	1.31
1	3200	1121.3	227.3	106.1	72.54	16.84	10.72	16	16	16	8.26	0.80
1	5600	1198.1	428.7	132.7	54.44	22.43	6.19	4	4	4	2.02	0.40
1	500000	1006.5	632.2	146.1	32.2	22.4	5.2	8	8	8	0.09	0.06
2	O–	1046.9	522.1	103.9	65.11	38.04	4.68	12	12	12	0.39	0.21
2	242	992.9	36.1	23.5	34.13	1.89	1.12	11	12	12	178.81	24.83
2	300	1037.3	53.0	27.9	46.10	1.72	1.15	12	12	12	83.09	8.74
2	600	1040.4	63.9	45.7	45.10	3.29	1.41	12	12	12	51.51	4.89
2	900	1049.0	77.9	56.0	71.57	3.08	4.01	16	16	16	36.66	3.63
2	1800	982.5	126.1	70.6	47.43	5.28	5.36	8	15	7	18.37	1.83

B

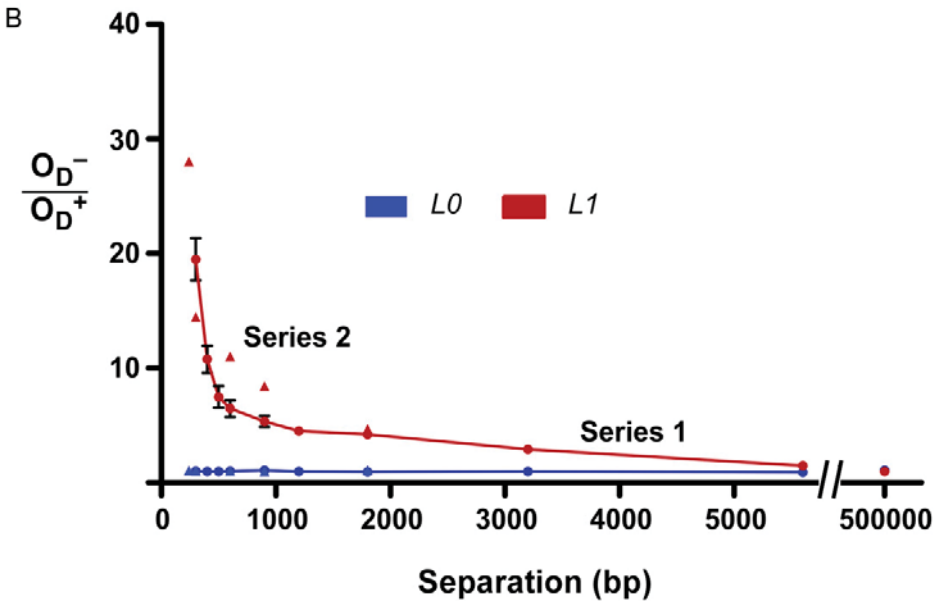
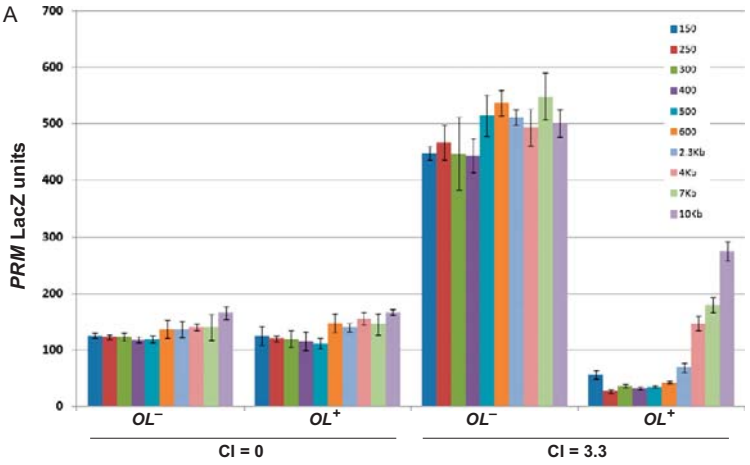


Fig. S5. LacI looping vs. separation. (*A*) LacI reporter data and derived $j_{LOOP}/L1$ values for different *Oid-O2* operator spacings used in Fig. 3. (*B*) The O_D effect: O_D^-/O_D^+ = (O_D^- LacZ units – *bkg*)/(O_D^+ LacZ units – *bkg*) at the *L1* concentration vs. spacing. Errors are 95% confidence limits.

A



B

Parameter	Description	Value
ΔG_{NS}	Binding of CI to non-specific sites	–5.74 ± 0.4 kcal/mol
ΔG_{oct}	CI octamerization and DNA looping	Fitted
ΔG_{tet}	CI tetramerization across the loop	–3.38 ± 0.8 kcal/mol
ΔG_{PRM}	Basal propensity of RNAP to occupy *PRM*	1.18 ± 0.05 kcal/mol
ΔG_{act}	Change in RNAP binding to *PRM* due to CI at *OR2*	–1.33 ± 0.08 kcal/mol
ΔG_{act_loop}	Change in RNAP binding due to CI at *OR2* being part of a DNA looping CI octamer	0.47 ± 1.0 kcal/mol
ΔG_{act_block}	Change in RNAP binding to *PRM* due to CI dimer at *OL3* in loop 4 or to *OL1* in loop 7	0.04 ± 0.6 kcal/mol
max_PRM	Basal *PRM* activity if fully occupied by RNAP	Fitted
bkg_PRM	Background LacZ units from PRM reporter	5 ± 6 LacZ units

C

Separation (bp)	OL^- [CI] = 0	OL^- [CI] = 3.3	OL^+ [CI] = 0	OL^+ [CI] = 3.3	Fitted ?G_{oct}
150	135.2 ± 4.9	482.3 ± 13.3	135.4 ± 17.8	59.7 ± 7.4	0.35
250	133.7 ± 4.5	506.6 ± 34.0	130.2 ± 5.5	29.0 ± 3.9	-0.77
300	135.6 ± 8.1	492.7 ± 71.2	132.1 ± 15.7	39.5 ± 3.2	-0.38
400	133.7 ± 6.1	501.2 ± 34.9	131.6 ± 19.2	36.0 ± 2.8	-0.47
500	129.1 ± 6.2	556.6 ± 39.7	121.1 ± 9.6	37.2 ± 2.5	-0.38
600	127.3 ± 14.6	502.3 ± 21.1	137.8 ± 15.8	39.8 ± 2.1	-0.30
2300	131.9 ± 14.1	493.9 ± 12.9	135.6 ± 7.2	66.5 ± 7.9	0.47
4000	131.3 ± 5.3	461.1 ± 30.9	145.1 ± 9.7	137.2 ± 11.9	1.14
7000	129.8 ± 20.7	505.9 ± 38.4	134.4 ± 17.5	166.4 ± 12.7	1.33
10000	142.2 ± 9.8	428.9 ± 21.0	143.0 ± 5.1	235.8 ± 14.7	1.71

Normalized values ± 95% CL, *n* = 4

Fig. S6. (Continued)

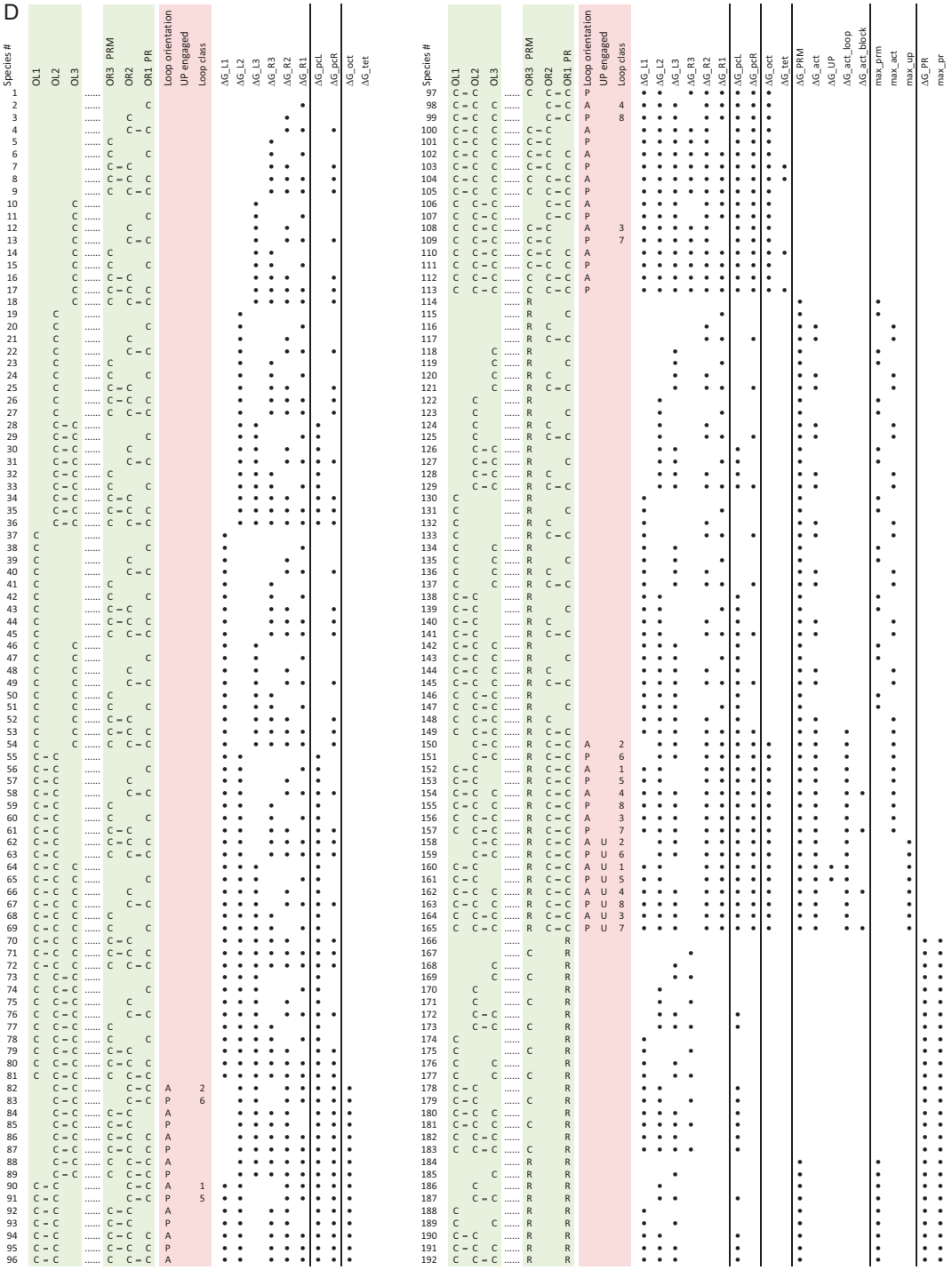


Fig. S6. Analysis of CI looping. (*A*) Raw reporter data. (*B*) Model parameters. (*C*) Normalized reporter data and fitted ΔG_{oct} values. (*D*) Species and assignments in the λ model. UP-containing and pR-containing species were ignored.

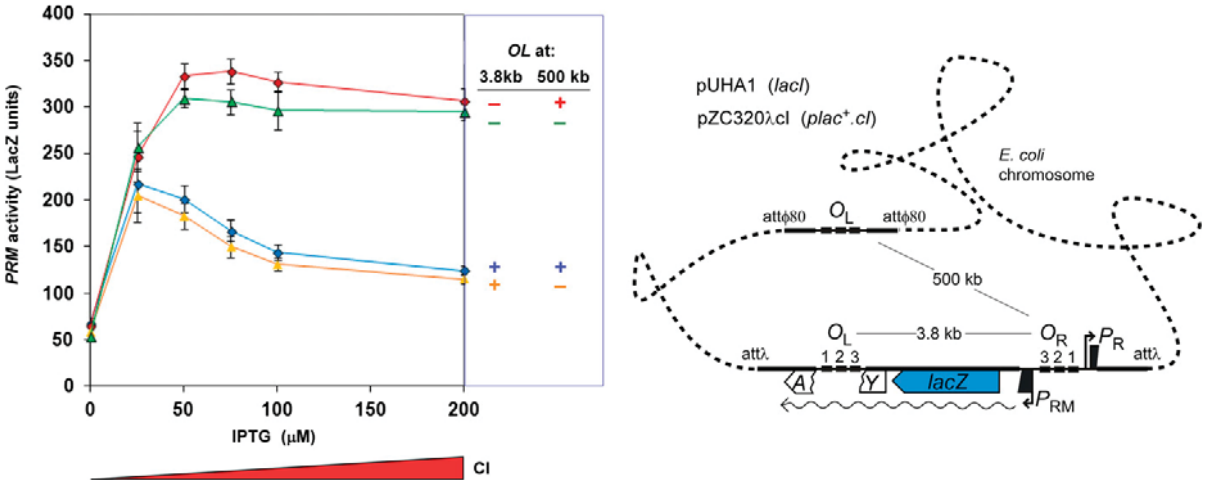


Fig. S7. Lack of effect of *OL* located 500 kb away from *OR.PRM*. The *PRM.lacZ* reporter system was as described (1), with OL^+ (*OL1.2.3* without flanking sequences) or an OL^- sequence placed downstream of *lacZ*, 3.8 kb away from *OR.PRM*, as part of a λRS45 prophage integrated at *attλ* in NK7049 (Δ*lacIZYA*)χ74 *galOP308* StrR Su^-. In addition, pAH167 (2) carrying OL^+ (*OL1.2.3* without flanking sequences) or an *OL*-less sequence, was integrated at *attφ*80 (position 1308595), 502 kb away from *attλ*. λCI repressor was expressed as described (1) from plac+ on the single copy plasmid pZC320λcI, repressed by LacI from the multicopy plasmid pUHA1. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was used to provide up to 3 WLU of CI.

1. Dodd IB, et al. (2004) Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev* 18(3):344–354.
2. Haldimann A, Wanner BL (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol* 183(21): 6384–6393.

A

Separation (bp)	Repression factor	Expression at *L0*	Expression at *L5*	Fitted expression at *L5*	Fitted *j/L5*	*j* (L5=90 nM)
Oid–	105	10000	94.99	93.60		
62	3618	10000	2.76	2.80	66	5940
71	7299	10000	1.37	1.60	119.39	10745
82	6051	10000	1.65	1.60	117.31	10558
92	5270	10000	1.90	2.00	94.18	8476
117	4078	10000	2.45	2.50	73.18	6586
152	1565	10000	6.39	6.50	26.89	2420
207	1176	10000	8.50	8.50	20.32	1829
405	536	10000	18.67	18.70	8.11	730
586	306	10000	32.68	32.90	3.73	336
838	265	10000	37.71	37.80	2.98	268
998	229	10000	43.60	42.70	2.4	216
1495	177	10000	56.38	55.00	1.42	128

B

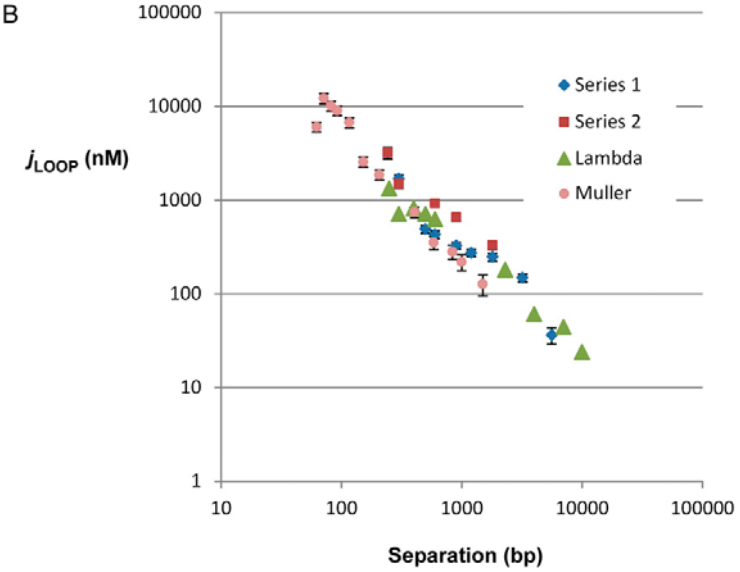
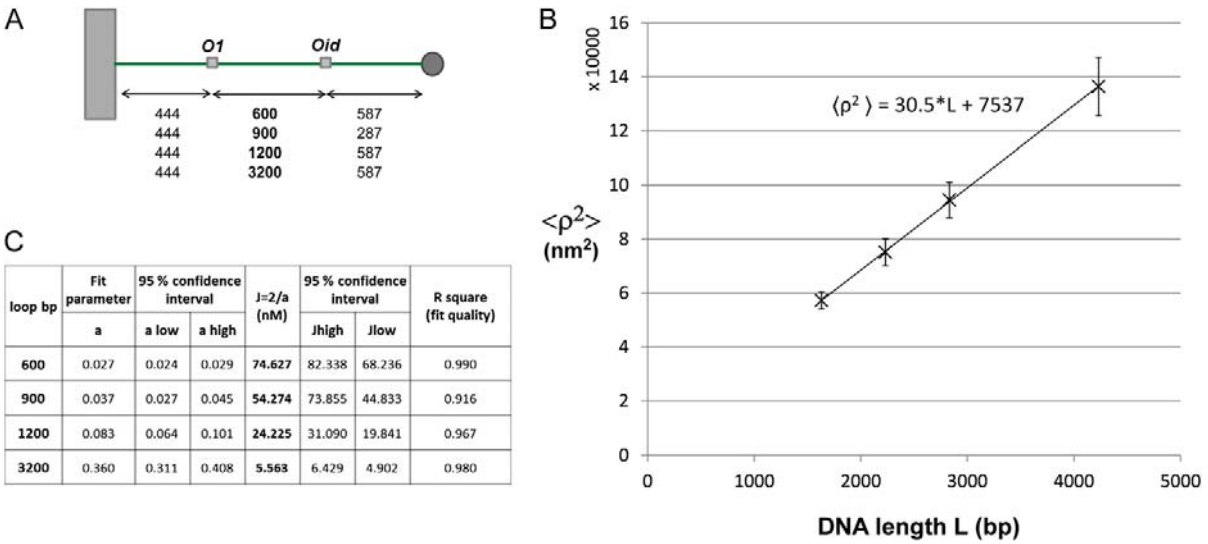


Fig. S8. Analysis of Lac looping vs. separation data of Muller et al. (1). (*A*) The data of Muller et al. (1), giving a repression factor (activity –LacI/activity +LacI) vs. operator separation for *Oid*.spacer.*PlacUV5.O1.lacZ* reporters (and the *Oid*-less control) at fivefold of WT [LacI] were extracted from Fig. 4*B*. The values were converted to arbitrary raw LacZ units by assuming a fixed unrepressed activity of 10,000 units and $bkg = 0$. The model of Fig. 2*C* was used to fit $j_{LOOP}/L5$ values, as described, using the parameters from Fig. 2*D*, except that *bkg* was set to zero, and *Zmax* and *K1* were unfixed. The fitted value for $K1 = 0.144$ is lower than obtained for our data, possibly reflecting a different RNAP affinity (*R*) for the *PlacUV5* promoter used by Muller et al. (1). $L5 = 90$ nM (5×18 nM) was used to calculate j_{LOOP} values. (*B*) Comparison with our j_{LOOP} values.

1. Müller J, Oehler S, Müller-Hill B (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J Mol Biol* 257(1):21–29.



C

loop bp	Fit parameter	95 % confidence interval		J=2/a (nM)	95 % confidence interval		R square (fit quality)
	a	a low	a high		Jhigh	Jlow	
600	0.027	0.024	0.029	74.627	82.338	68.236	0.990
900	0.037	0.027	0.045	54.274	73.855	44.833	0.916
1200	0.083	0.064	0.101	24.225	31.090	19.841	0.967
3200	0.360	0.311	0.408	5.563	6.429	4.902	0.980

Fig. S9. TPM. (*A*) Tether structures. (*B*) TPM calibration. The ρ^2 is linearly related to the DNA tether length in base pairs (18). Time series of *xy* positions of beads tethered by identical DNA molecules in λ buffer were recorded and combined for analysis. Running averages of ρ^2 using a time window of 8 s were computed. The mean and SD of these ρ^2 distributions were plotted vs. the contour lengths of the DNA tethers (marker x). As shown, ρ^2 is a linear function of tether length. This calibration was used to verify loop sizes. (*C*) Fitting statistics for j_{LOOP} measurements of Fig. 5*C* for $y = a(1 + x)$, where y is $p_{unlooped}/p_{looped}$, x is base pairs, and a is $2/j$.

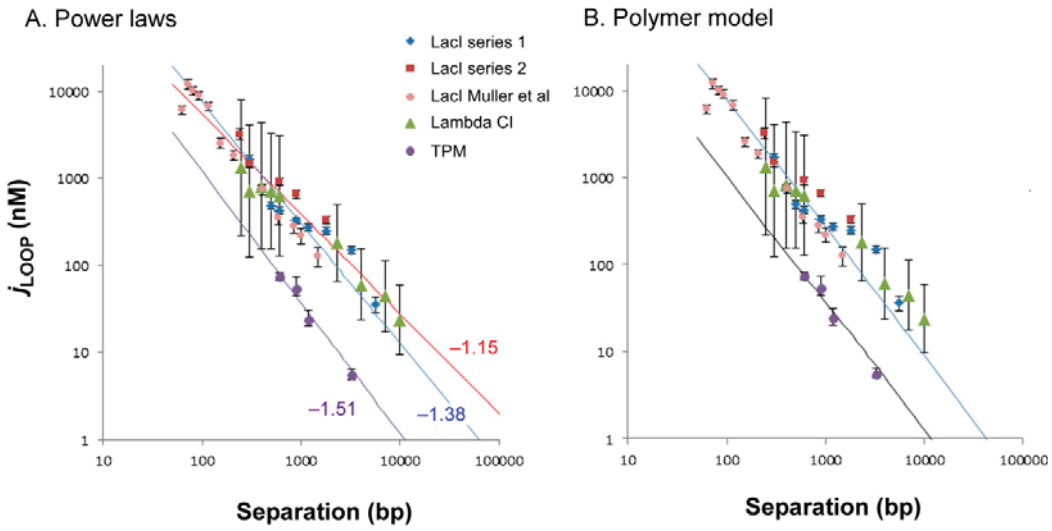


Fig. S10. Fitting the j_{LOOP} vs. separation data with DNA polymer models. Fitting used the Microsoft Excel solver function, comparing the model-generated expected j_{LOOP} values with the observed experimental estimates and minimizing $\Sigma[(observed - expected)^2/expected]$. (*A*) Fitting the in vivo data to a power law $j_{LOOP} = a.bp^{\gamma}$. Fitting of a and γ was to all in vivo data, except the 62-bp point of Muller et al. (1) (blue line; $\gamma = -1.38$), the in vivo data without any Muller et al. data (red line; $\gamma = -1.15$), and the TPM data (purple line; $\gamma = -1.51$). (*B*) Fitting the in vivo (+ Muller data) and in vitro data to a combined worm-like chain and freely jointed chain DNA polymer model. Eq. **3** of Rippe (2): $j_M(n) = 0.53 \times n^{-3/2} \times \exp[(d - 2)/(n^2 + d)] \times l^{-3}$ was used with separation measured in number of Kuhn lengths n = bp × 0.34/l. Kuhn length, l (fitted and allowed to differ in vivo and in vitro): in vivo, l = 45.2 nm = 133 bp; in vitro, l = 165 nm = 486 bp (persistence length = l/2). d (contact distance) = 1.66 nm (fitted but held the same for in vivo and in vitro).

1. Müller J, Oehler S, Müller-Hill B (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J Mol Biol* 257(1):21–29.
2. Rippe K (2001) Making contacts on a nucleic acid polymer. *Trends Biochem Sci* 26(12):733–740.

Chapter 3

Quantitation of interactions between two DNA loops demonstrates loop-domain insulation in *E. coli* cells

David G. Priest[1], Sandip Kumar[2], Yan Yan[2], David Dunlap[2], Ian B. Dodd[1*], Keith E. Shearwin[1]

[1]School of Molecular and Biomedical Science (Biochemistry), University of Adelaide, Adelaide SA 5005, Australia.

[2]Department of Cell Biology, Emory University, Atlanta GA 30322, USA.

Manuscript re-submitted under review at PNAS USA

Statement of Authorship

Title of Paper	Quantitation of interactions between two DNA loops demonstrates loop-domain insulation in E. coli cells
Publication Status	○ Published, ○ Accepted for Publication, ◒ Submitted for Publication, ○ Publication style
Publication Details	David G. Priest, Sandip Kumar, Yan Yan, David Dunlap, Ian B. Dodd, Keith E. Shearwin. Resubmitted manuscript under review at PNAS USA.

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	David Priest		
Contribution to the Paper	Designed, constructed and assayed DNA looping constructs. Made constructs for TPM experiments, assisted collaboration correspondence, helped prepare figures for paper, helped write paper in collaboration with Ian Dodd.		
Signature		Date	

Name of Co-Author	Sandip Kumar		
Contribution to the Paper	Designed and performed TPM experiments, analysed data and assisted in writing the manuscript.		
Signature		Date	

Name of Co-Author	Yan Yan		
Contribution to the Paper	Designed and performed TPM experiments, analysed data and assisted in writing the manuscript.		
Signature		Date	

Name of Co-Author	David Dunlap		
Contribution to the Paper	Designed experiments, analysed data, and assisted in writing the manuscript.		
Signature		Date	

Statement of Authorship

Title of Paper	Quantitation of interactions between two DNA loops demonstrates loop-domain insulation in E. coli cells
Publication Status	○ Published, ○ Accepted for Publication, ◉ Submitted for Publication, ○ Publication style
Publication Details	David G. Priest, Sandip Kumar, Yan Yan, David Dunlap, Ian B. Dodd, Keith E. Shearwin. Resubmitted manuscript under review at PNAS USA.

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)			
Contribution to the Paper			
Signature		Date	

Name of Co-Author	Ian B. Dodd		
Contribution to the Paper	Supervised project, designed experiments, analysed data, performed modelling, wrote correspondence, helped initiate collaboration, wrote bulk of manuscript in collaboration with David Priest.		
Signature		Date	

Name of Co-Author	Keith E. Shearwin		
Contribution to the Paper	Supervised project, designed experiments, analysed data and assisted in writing the manuscript. Initiated project, coordinated collaboration, and corresponded with collaborators.		
Signature		Date	

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Quantitation of interactions between two DNA loops demonstrates loop-domain insulation in E. coli cells

David G. Priest[1], Sandip Kumar[2], Yan Yan[3], David Dunlap[2], Ian B. Dodd[1]*, Keith E. Shearwin[1]

[1] School of Molecular and Biomedical Science (Biochemistry), University of Adelaide, Adelaide SA 5005, Australia [2] Departments of Biology, Cell Biology and [3]Physics, Emory University, Atlanta GA, USA.

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Eukaryotic gene regulation involves complex patterns of long-range DNA-looping interactions between enhancers and promoters, but how these specific interactions are achieved is poorly understood. Models that posit other DNA loops, that aid or inhibit enhancer-promoter contact, are difficult to test or quantitate rigorously in eukaryotic cells. Here, we use the well-characterized DNA looping proteins Lac repressor and phage λ CI to measure interactions between pairs of long DNA loops in E. coli cells in the three possible topological arrangements. We find that side-by-side loops do not affect each other. Nested loops assist each other's formation consistent with their distance-shortening effect. In contrast, alternating loops, where one looping element is placed within the other DNA loop, inhibit each other's formation, thus providing clear support for the loop-domain model for insulation. Modeling shows that combining loop assistance and loop interference can provide strong specificity in long range interactions.

enhancer | lambda CI | Lac repressor | tethered particle motion | statistical mechanical modeling

Transcription of genes is regulated by promoter-proximal DNA elements and distal DNA elements that together determine condition-dependent gene expression. In eukaryotic genomes, enhancers can be many hundreds of kilobases away from the promoter they regulate (1-3) and the intervening DNA can contain other promoters and other enhancers (4-7). How the regulatory influence of distal elements is exerted efficiently and specifically at the correct promoters is poorly understood.

Enhancers are clusters of binding sites for transcription factors and chromatin modifying enzymes, and activate promoters by directly contacting them via DNA looping (8-12). Enhancer trap approaches and mapping of transcription factor binding and chromatin modifications have identified tens of thousands of enhancer elements in metazoan genomes (7, 13-16). Chromatin capture studies show that enhancers and promoters are connected in highly complex condition-dependent patterns (6, 15, 17). . Although core enhancer and promoter elements can provide some specificity (18), enhancers are often able to activate heterologous promoters if they are placed near to each other. Indeed, this lack of specificity is the basis for standard enhancer assays and screens (7, 14, 19). Thus, additional mechanisms are clearly needed to target enhancers to the correct promoters over long distances and to prevent their interaction with the wrong promoters. Dedicated DNA looping elements that can either assist or interfere with enhancer-promoter looping are thought to play a major role.

In theory, any DNA loop that brings the enhancer and promoter closer together should assist their interaction (Figure 1A), since the efficiency of contact increases as the length of the DNA tether between the sites shortens (20-24). Promoter-tethering elements in *Drosophila* that allow activation by specific enhancers over long distances are proposed to form DNA loops between sequences near the enhancer and the promoter (18, 25). In the mouse β–globin locus, the Ldb1 protein binds to proteins at the locus control region and at the promoter and appears to form a bridge necessary for efficient enhancer-promoter contact (26). In bacteriophage λ, the CI protein forms a 2.3 kb DNA loop that brings a distal stimulatory site close to RNA polymerase at the *PRM* promoter (27). Enhancer-promoter targeting has also been demonstrated on plasmid constructs using heterologous looping proteins, for example with λ CI in human cells (28) and the *Drosophila* GAGA protein in human cells and in yeast (29, 30).

DNA looping also seems to be able to inhibit enhancer-promoter contact. Enhancer-blocking insulators are defined by their ability to prevent enhancer activation of the promoter when placed between the enhancer and the promoter. A large body of evidence is consistent with the idea that insulators work by binding proteins that form DNA loops to other insulators (31-35). This mechanism of insulator action is rationalized by the loop domain model (36), which proposes that the formation of a DNA loop creates a separate topological domain that inhibits interaction between any site within the loop and any site outside the loop (Figure 1A). This model is the only one that can currently explain the requirement that an insulator must be between the enhancer and promoter to block activation, as well as observations that two insulators between the enhancer and promoter sometimes do not block activation (37, 38). The loop domain model is also a potential explanation (35) of the topologically associated domains (TADs) revealed by genome-wide mapping of DNA contacts by chromatin capture methods in genomes from mice to bacteria (39-45). Individual DNA sites between two domain boundaries interact more frequently with each other than they do with individual DNA sites in other TADs, that is, domain boundaries act like insulators. Consistent with the loop domain model, domain boundaries interact with each other at high frequency and often contain known DNA looping elements,

Significance

Genes are frequently regulated by interactions between proteins that bind to the DNA near the gene and proteins that bind to DNA sites located far away, with the intervening DNA looped out. In eukaryotic genomes, genes and their distant sites are intermingled in complex ways and it is not understood how the correct connections are formed. Using two pairs of DNA-looping sites in bacterial cells, we tested the idea that one DNA loop can either assist or interfere with the formation of another DNA loop. By measuring the strength of these interactions between loops, we showed that this mechanism is capable of directing a distant site to the correct gene and preventing it contacting the wrong gene.

Reserved for Publication Footnotes

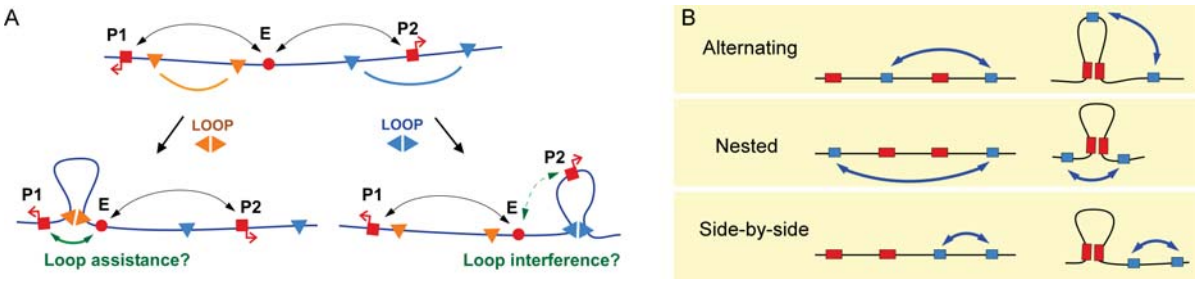


Fig. 1. Interactions between DNA loopsA. DNA-looping interactions between sites on the DNA, for example between an enhancer (E) and promoters (P1 and P2), are proposed to be affected by other DNA loops. Specific interactions between looping elements (triangles) can either assist enhancer-promoter looping by bringing the enhancer and promoter closer together (orange triangles), or are thought to interfere with enhancer-promoter looping by placing them in separate loop domains (blue triangles).B. The three possible topological arrangements of two pairs of interacting sites on DNA. We tested in each case whether the formation of a loop between one pair of sites affects the propensity of the other sites to interact (and vice versa).

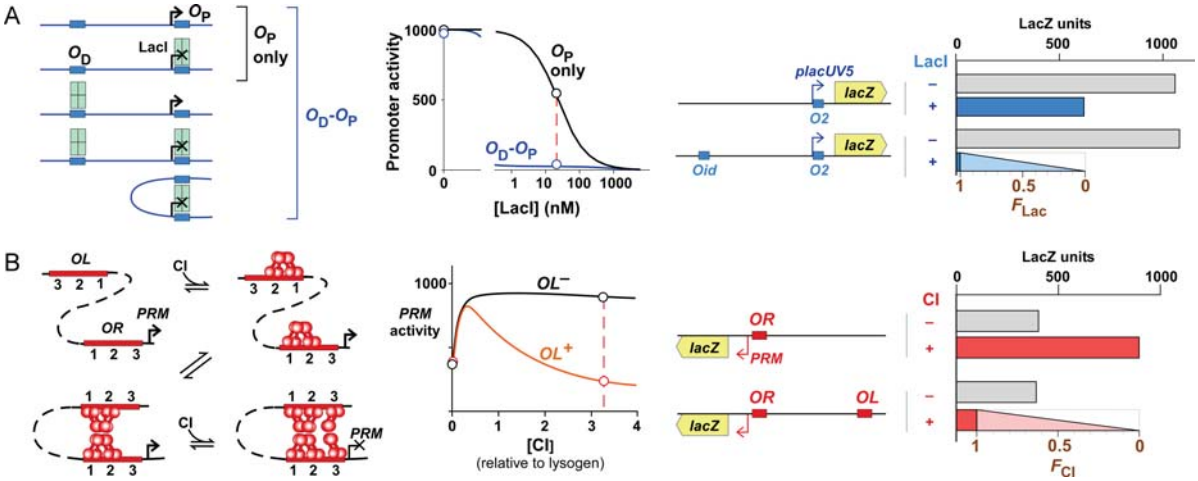


Fig. 2. Measuring looping in vivoA. LacI. (Left panels) A LacI tetramer represses the promoter by binding to a proximal lac operator (O_P). Repression can be improved by DNA-looping between O_P and a distal lac operator (O_D) because a tetramer bound at O_D constitutes a high local concentration at O_P due to DNA tethering. The strength of the O_D effect depends on the operator combination, the LacI concentration and the length of the spacer DNA. (Right panels) The effect of O_D on reporter expression can be used to measure the fraction of time that the system spends in the looped state, F_{Lac}. In the absence of looping, F_{Lac} = 0, activity is the same with or without O_D. If the system were fully looped, F_{Lac} = 1, the activity of the O_D-O_P reporter would be reduced to background.B. CI. (Left panels) Repression of the *PRM* promoter is dependent on DNA looping between *OR* and *OL*, as *OR3* is too weak to be occupied by a CI dimer at physiological concentrations. (Right panels) As with LacI, the effect of *OL* on reporter expression can be used to measure the fraction of time that the system spends in the looped state, F_{CI}.

being enriched for insulator protein binding sites, as well as for active promoters and enhancers (44).

Despite the fundamental significance of the loop domain model in explaining insulator action and the formation of TADs, unequivocal tests of the model in vivo have been hampered by the complexity of eukaryotic genomes and their gene regulatory elements. Evidence from more defined systems has supported the loop domain model. In a plasmid transfection system, a 344 bp DNA loop formed by a Tet repressor derivative around the SV40 enhancer inhibited its activation of a promoter 2 kb away (46). However, the small loop may have affected enhanceosome assembly in these experiments. In an in vitro system with *E. coli* proteins and supercoiled plasmids, a 630 bp DNA loop formed by the Lac repressor (LacI) around the NtrC enhancer element inhibited its activation of the *glnA* promoter 2.5 kb away (47). However, this effect has not been tested in vivo. In both studies, the lack of information about DNA looping efficiencies prevents a quantitative analysis of loop interference.

In order to clearly test the loop domain model in vivo and to rigorously quantitate loop interaction effects, we measured interactions between large DNA loops formed in the *E. coli* chromosome by the two best characterized DNA looping proteins, LacI and bacteriophage λ CI. Previously (24), we quantitated looping efficiency of single DNA loops in vivo by assaying DNA loop-mediated LacI or CI repression of a reporter gene, and in vitro by the single molecule technique, tethered particle motion (TPM). Here, we have combined LacI and CI DNA loops in each of the three possible topologies (Figure 1B) to show that: alternating DNA loops interfere, nested DNA loops assist, and side-by-side loops do not affect one another's formation. Fitting our data to a general statistical-mechanical model of loop interaction allowed calculation of the strength of looping assistance or interference.

Results

Quantitation of interactions between DNA loops

There are three topological ways to arrange two pairs of DNA looping sites: alternating, nested and side-by-side (Figure 1B; note that we ignore the parallel/anti-parallel orientation of the strands at the loop clamps). The expectation from the loop domain model is that in the alternating arrangement, the formation of one loop will interfere with the formation of the second loop. For nested loops, we expect that the formation of one loop would

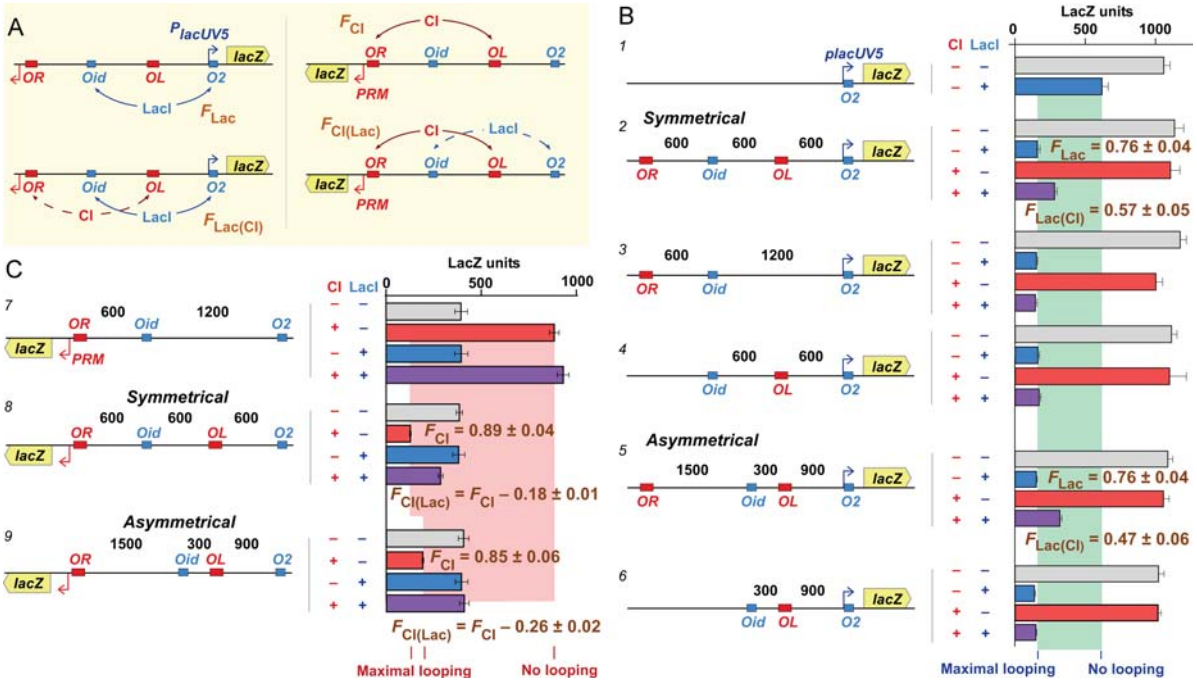


Fig. 3. Loop interference by alternating loops A. Experimental strategy for measuring interactions between alternating loops. A *lacZ* reporter is placed downstream of *PlacUV5* to measure the fraction of LacI-mediated *Oid-O2* looping in the absence of CI (F_{Lac}) or in the presence of CI ($F_{Lac(CI)}$). Similarly, the reporter is placed downstream of *PRM* to measure the fraction of CI-mediated *OL-OR* looping in the absence of LacI (F_{CI}) or in the presence of LacI ($F_{CI(Lac)}$).B. Effect of CI loops on LacI looping. Chromosomal reporters (distances in bp; Materials and Methods; Figure S1A and S2) were exposed to fixed levels of LacI (18 nM) or CI (3.3 wild-type lysogenic units) expressed from other chromosomal constructs (Materials and Methods; Figure S1B;(24)). Histograms show steady-state LacZ units and 95% confidence intervals (*n*=5-12). Construct 1 shows weak repression in the absence of looping to the distal operator. The presence of *Oid* 1200 bp upstream improves repression in both the symmetrical constructs (2, 3 and 4) and the asymmetrical constructs (5 and 6) but this improvement is diminished when *Oid* is flanked by *OL* and *OR* and CI is present (constructs 2 and 5). The F_{Lac} and $F_{Lac(CI)}$ values were calculated by fitting the four data points for each construct with a statistical-mechanical model of LacI repression (24); Materials and Methods). Errors for *F* values are standard deviations. C. Effect of LacI loops on CI looping. Experimental details are as in (B) except *n*=9 for the LacZ units. In the absence of *OL*, CI activates *PRM* but does not repress it (construct 7). Repression by CI is improved in the presence of *OL* in the symmetrical construct (8) and the asymmetrical construct (9); as expected, repression is stronger for the 1200 bp compared to the 1800 bp CI loop (24). This improved repression is diminished when *OL* is flanked by *Oid* and *O2* and LacI is present (additional controls for the symmetrical construct are given in Figure S3). The F_{CI} and $F_{CI(Lac)}$ values were calculated by fitting the four data points for each construct and the four data points of construct 7 with a statistical-mechanical model of CI repression (24); Materials and Methods). $F_{CI(Lac)}$ values are expressed relative to the corresponding F_{CI} values (errors are standard deviations).

assist the formation of the second. If the interior loop forms first, then the linear distance (along the DNA), between the exterior loop sites becomes shorter. When the exterior loop is formed, the distance between the interior loop sites may or may not shorten (depending on the geometry), but they are nevertheless likely to become more spatially constrained because they become linked by a DNA tether on each side. We do not expect side-by-side loops to affect each other.

Previously we used in vivo reporter assays and statistical-mechanical modeling to extract key DNA looping parameters for LacI and CI DNA loops (24), allowing calculation of the fraction of the time (averaged over many cells) that the DNA is in a looped state, which we term *F*. The rationale of this approach is shown in Figure 2.

For LacI, repression of a *PlacUV5* promoter with a single proximal *lacO2* operator is relatively inefficient at low LacI concentrations. Repression is more efficient when a strong distal operator (*Od*) is present and DNA looping between the two operators can occur, aiding binding of the LacI tetramer to the proximal site (Figure 2A). The fractional DNA looping *F* can be extracted from measurement of the efficiency of repression in the presence of the distal operator (24); Materials and Methods).

For CI, the fraction of DNA looping can be obtained from measurements of CI regulation of the phage *PRM* promoter (24, 27, 48). At low CI concentrations, CI tetramers assemble at the *OL* and *OR* sites, and these DNA-bound tetramers can form an *OL-OR* DNA loop by CI octamerization (Figure 2B). *PRM* is activated by binding of CI to *OR2* (49), both in unlooped and looped states (27). However, repression of *PRM* by CI, which occurs at higher CI concentrations, is dependent on looping because repressive CI binding at the very weak *OR3* operator relies on interactions with CI bound at stronger sites at *OL* (Figure 2B). Again, the efficiency of repression of *PRM* in the presence of *OL* can be used to measure the fractional DNA looping *F* (24); Materials and Methods).

Alternating loops interfere in vivo

We made *lacZ* reporters with an alternating arrangement of binding sites for LacI (*O2* and *Oid*) and CI (*OR* and *OL*). For measuring LacI looping, *lacZ* was placed downstream of *PlacUV5.O2*; for measuring CI looping, *lacZ* was placed downstream of *OR.PRM* (Figures 3A, S1A and S2). Two different alternating arrangements were made: symmetrical and asymmetrical (Figure 3B and 3C), with loop sizes ranging from 1200 to 1800 bp. The smallest distance between operators was 300 bp which is well

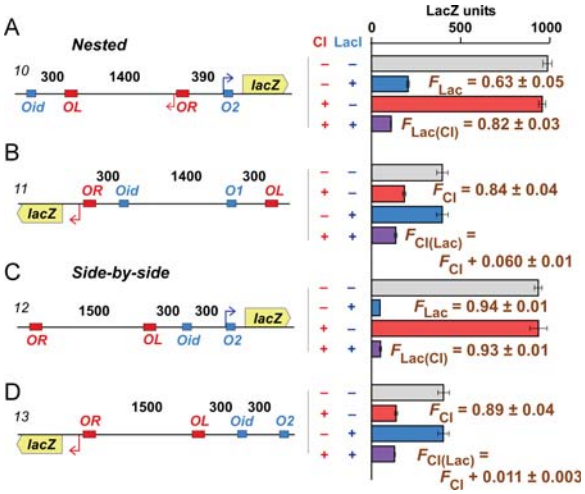


Fig. 4. Loop assistance by nested loops and no effect of side-by-side loopsExperimental details are as in Figure 3. Histograms show steady-state LacZ units and 95% confidence intervals (n=4-9). F_{Lac} and $F_{Lac(CI)}$ values for each construct (10 and 12) were calculated by fitting the four data points shown. The F_{CI} and $F_{CI(Lac)}$ values were calculated by fitting the four data points for each construct (11 or 13) and the four data points of construct 7 (Figure 3). $F_{CI(Lac)}$ values are expressed relative to the corresponding F_{CI} values. Errors for F values are standard deviations. A and B. Nested loops. Addition of the protein for the internal loop increases looping of the external loop.C and D. Side-by-side loops. Addition of CI has no effect on LacI looping and vice versa.

beyond the apparent 20-45 bp persistence length of DNA in *E. coli* (24, 50).

We measured *lacZ* expression in the presence or absence of fixed concentrations of LacI (18 nM) and CI (3.3 wild-type lysogenic units), expressed from chromosomally integrated constructs (27),(24)(Figure S1B).

In the absence of the distal operator, LacI repression was weak (Figure 3B, construct 1). Repression improved substantially with a distal *Oid* operator 1200 bp away (Figure 3B, constructs 2-6), with the magnitude of the improvement indicating looping 76±4 % of the time (errors are 95% confidence limits). When the distal *Oid* was flanked by CI operators and CI was present, then LacI repression was reduced to levels indicating 57±5 % and 47±6 % looping for the symmetrical and asymmetrical constructs, respectively (Figure 3B, constructs 2 and 5). Importantly, both CI operators were required for this loop inhibition (Figure 3B, constructs 3, 4 and 6), implying that CI looping sequestered the distal *lac* operator.

A reciprocal effect was seen by assaying CI looping (Figure 3C). In the absence of the distal *OL* operator, CI activated *PRM* some 2-fold (Figure 3C, construct 7). The presence of *OL* 1200 or 1800 bp upstream enabled CI repression, indicative of 89±4 % or 85±6 % looping, respectively (Figure 3C, constructs 8 and 9). Addition of LacI did not affect *PRM* regulation in the absence of *OL* (construct 7) but inhibited repression when *OL* was present and flanked by *lac* operators, with looping measured at 71±5 % or 59±8 % (constructs 8 and 9). Again, a single *lac* operator within the loop or outside the loop had no effect (Figure S3).

These results show a loop interference effect for loops in the alternating arrangement, with CI or LacI sites effectively acting as insulators.

Nested loops give loop assistance

We made two different nested loop constructs, one with a 1400 bp CI loop inside a 2090 bp LacI loop (Figure 4A, construct 10), and one with a 1400 bp LacI loop (an *Oid-O1* loop) inside a 2000 bp CI loop (Figure 4B, construct 11). Addition of CI improved LacI looping from 63±5 % to 82±3 % for the CI-inside construct. Addition of LacI improved CI looping from 84±4 % to 90±5 % for the Lac-inside construct.

Thus, in a nested arrangement, formation of the inside loop improved formation of the outside loop. Our reporter approach prevented us measuring the fraction of looping for the internal loop.

This experiment also replicates the 'insulator bypass' experiment, which provides such strong support for the loop domain model of insulator action. Placement of *two* insulators between an enhancer and promoter can relieve the enhancer blocking caused by a single insulator (37, 38). Under the loop domain model, looping between the tandem insulators would leave the enhancer and promoter in the same domain. In our experiments also, a single CI site placed between two LacI sites can inhibit their interaction (Figure 3; as long as it can loop to an outside site), while two CI sites do not block interaction of the LacI sites, in fact they stimulate it.

Side-by-side loops do not interact We also made reporters with a 1500 bp CI loop and a 300 bp LacI loop placed next to each other and separated by 300 bp. We found that the presence of CI did not affect LacI looping (Figure 4C), and the presence of LacI did not affect CI looping (Figure 4D), confirming our expectation that side-by-side loops do not interact.

Model to quantitate loop interactions

Although it is clear from our data that alternating and nested loops interact, it is difficult to appreciate the strength of the interaction because none of the loops forms with 100% efficiency (i.e. F=1). However, the measured F values obtained for each of the loop arrangements provide sufficient information to calculate the strength of the loop interference or assistance by use of a simple statistical mechanical model of loop interaction.

If we consider just one pair of operators, then the DNA can exist in either a looped or an unlooped state. This equilibrium is determined by the nature and concentration of the looping protein, the length and nature of the DNA between the operators, as well as the chemical environment. The propensity of loop formation relative to the unlooped ground state under these fixed conditions can be simply defined by a statistical weight W. The fraction looped is a function of this weight $F=W/(1+W)$. Each LacI loop and CI loop has its own weight, W_{Lac} or W_{CI} that determines F_{Lac} or F_{CI} in the absence of the other protein (Figure 5). Note that for LacI and CI, the looped and unlooped states are each mixtures of species (see Figures 2A and 2B), and for the purposes of the loop interaction analysis it is not necessary to distinguish these.

In the case where there are two pairs of operators, there are four loop states: all sites unlooped, only LacI sites looped, only CI sites looped, or both pairs of sites looped (Figure 5). If the loops form independently of each other, that is, they do not interact, then the statistical weight of the double-looped state is just the product of the individual loop weights W_{Lac}•W_{CI}. However, if the loops do interact, a loop interaction factor α (Figure 5) can be used to quantitate the direction and strength of the loop interaction, with the weight of the double looped species represented by α•W_{Lac}•W_{CI}. Thus, when α<1, there is loop interference, the double looped species forms less frequently than expected; when α>1, there is loop assistance, the double looped species forms more frequently than expected.

For each of the alternating loop arrangements, our in vivo measurements provide us with four F values: F_{Lac}, F_{CI} (for looping in the absence of the other loop), $F_{Lac(CI)}$ and $F_{CI(Lac)}$ (for looping in the presence of the other loop). By using a Monte-Carlo fitting procedure, we were able to obtain convergent estimates for W_{Lac}, W_{CI} and α that closely reproduced these F values (Figure 5; Materials and Methods).

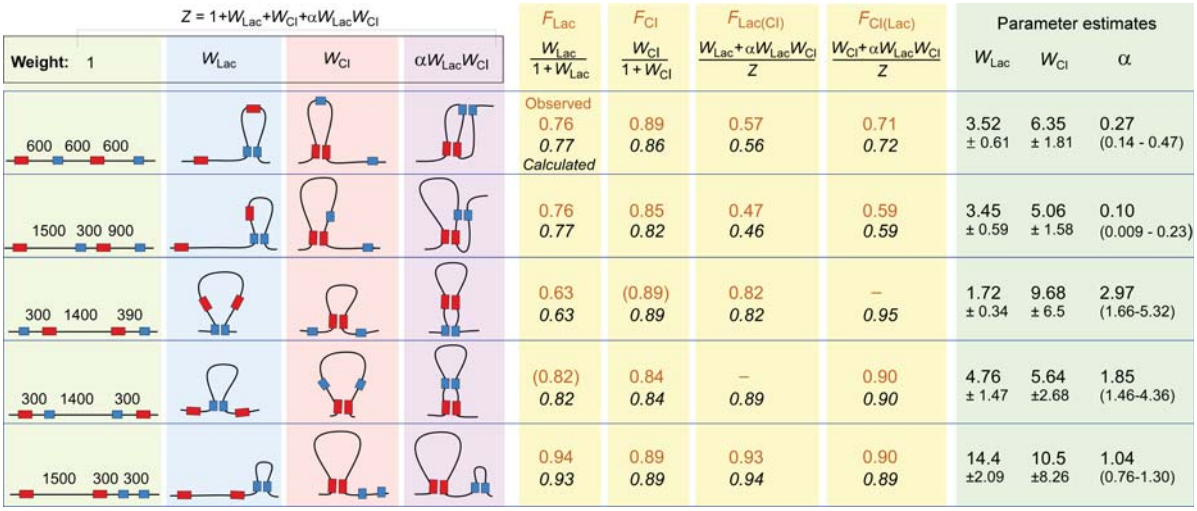


Fig. 5. Modeling loop interactionThe four left panels show for each construct tested, the four possible looped species in the presence of both LacI and CI and their relative statistical weights. The fraction of each species is given by its weight divided by the partition sum, *Z*. α is the loop interaction factor: α>1 indicates loop assistance (the double looped species forms more often than the product of individual loop weights); α<1 indicates loop interference (the double looped species forms less often than the product of individual loop weights). The central four panels list the observed *F* values (brown) from Figures 3 and 4, and the *F* values (italic) calculated from the equations shown using the fitted loop weights and α for each construct (right panel). Observed *F* values in parentheses were estimated from other measurements (Figure S3). W_{Lac}, W_{CI} and α were fitted to minimize the difference between observed and calculated *F* values (Materials and Methods). *W* errors are standard deviations and the ranges for the α estimates are the 2.5-97.5 percentiles (>900 fitting runs).

The obtained values for α, 0.27 (95% confidence interval 0.14-0.47) and 0.10 (0.01-0.23) for the symmetrical and asymmetrical alternating arrangements, respectively, indicate that the formation of one loop inhibits the formation of the other loop by 3.7- or 10-fold (1/α). Whether there is a real difference between the two geometries is not clear, as the errors in these estimates overlap.

For each of the nested loop arrangements, our measurements provide us with only two *F* values: F_{Lac}, $F_{Lac(CI)}$ (for the CI-inside case) or F_{CI} and $F_{CI(Lac)}$ (for the LacI-inside case). However, we can interpolate between other *F* measurements to estimate F_{CI} for the CI-inside case and F_{Lac} for the Lac-inside case (Figure S4). For the CI-inside case where the CI loop is 1400 bp, we used the F_{CI} values obtained for the 1200 bp (construct 8, Figure 3C), 1500 bp (construct 13, Figure 4D), 1800 bp (construct 9, Figure 3C) and 2000 bp (construct 11, Figure 4B) CI loops to give an estimate of F_{CI}=0.89 (Figure S4A). For the LacI-inside case, where we have no direct comparisons for the *Oid-O1* looping, we interpolated between the j_{LOOP} values obtained from model fitting for the 300 bp (construct 12, Figure 4B), the 1200 bp (Figure 3B, constructs 2 and 5) and 2090 bp (construct 9, Figure 4B) LacI loops (Figure S4B). j_{LOOP} is the effective concentration of one site on the DNA relative to another site on the same DNA molecule, and is a critical parameter in determining loop efficiency (51). From this estimate of j_{LOOP} for a 1400 bp loop, and our previous measurements of the dissociation constants for LacI at *O1*, it is possible to calculate F_{Lac}=0.82 for the 1400 bp *Oid-O1* loop {Priest, 2014 #5) (SI Materials and Methods). Having three *F* values for each nested construct allowed us to obtain convergent estimates for W_{Lac}, W_{CI} and α in each case (Figure 5). The values for α were 3.0 (1.7-5.3) and 1.9 (1.5-4.4) for the LacI-inside and CI-inside arrangements, respectively, indicating moderate loop assistance.

Analysis of the side-by-side loop data gave α∼1 (0.76-1.30; Figure 5), indicating a lack of loop interaction in this arrangement.

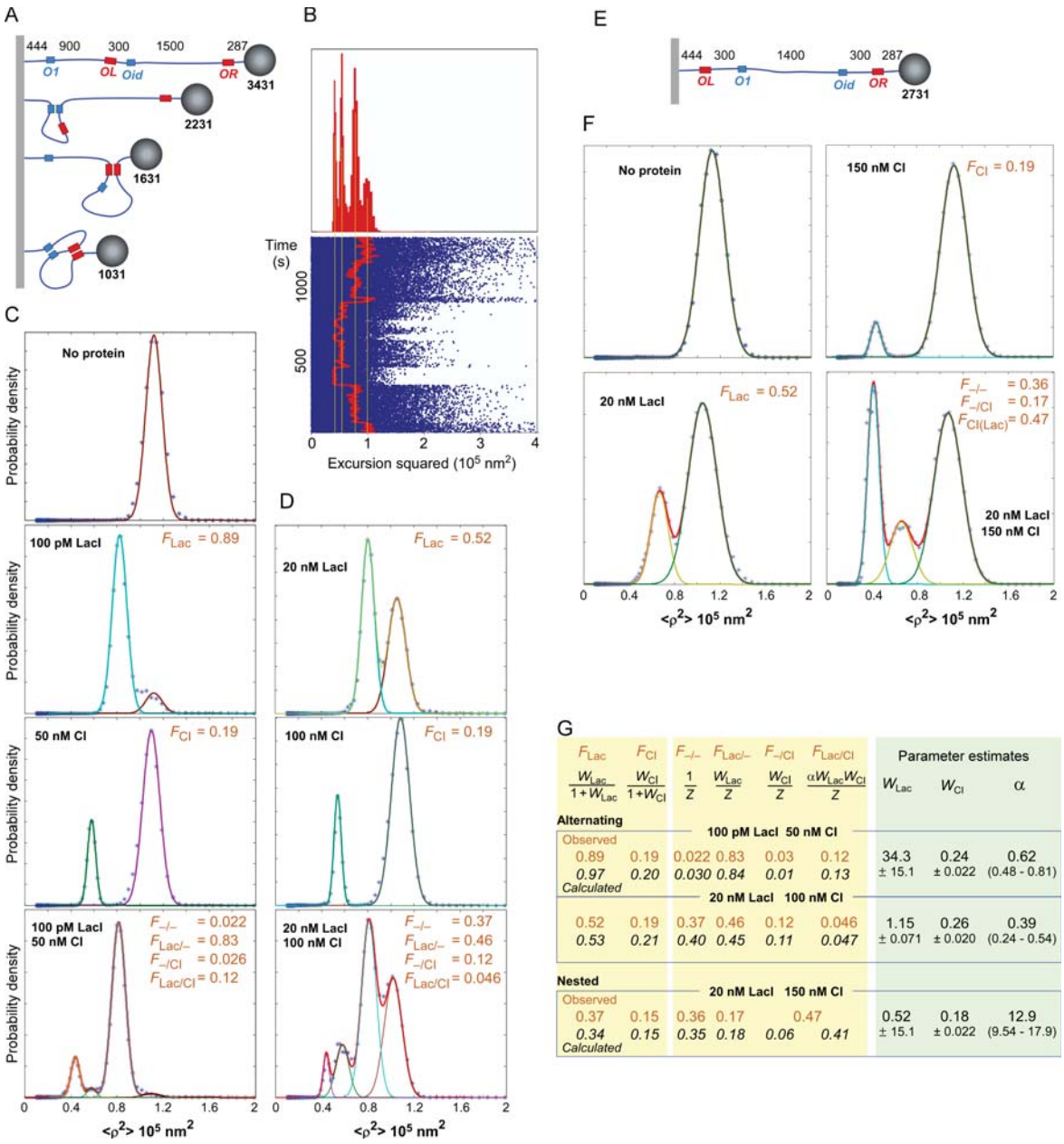
Loop interference and assistance in vitro

We used the TPM technique to detect and measure DNA looping in vitro (24, 52-54). For loop interference, we used a DNA tether with the asymmetrical looping arrangement that we used in vivo (Figure 3B, construct 5, except *Plac*$^{+}$*.O2* was replaced with *PLac*$^{-}$*.O1*). This arrangement had been designed so that LacI or CI or both concurrent loops could be distinguished by their effect on the tether length (Figure 6A), and therefore on the mean displacement of the attached bead. Recordings of the tether lengths versus time for individual tethers exposed to both proteins exhibited discrete stepping between values expected for the unlooped and all three looped forms (Figure 6B). By following multiple beads over time in the presence of LacI or CI, histograms of the probability versus tether length were compiled (Figure 6C and 6D; Materials and Methods). Two combinations of LacI and CI concentrations (100 pM LacI + 50 nM CI and 20 nM LacI + 100 nM CI) were used, and each protein was also used alone at the same concentration. Estimates of the fraction of each looped species were obtained by fitting the histograms to Gaussian curves. Because it was possible to resolve all the looped species, six *F* values were obtained for each condition (Figure 6B and 6C).

These data were analysed using the statistical-mechanical loop interaction model in order to extract loop weights and α (Figure 6G). An α of 0.62 (0.48-0.81) results indicate ∼1.5-fold loop interference for the 100 pM LacI, 50 nM CI data and an α of 0.39 (0.24-0.54) indicates ∼2.5-fold interference for the 20 nM LacI, 100 nM CI data, with overlapping 95% confidence intervals. We have more confidence in the 2.5-fold interference estimate because the four loop species were better balanced in the 20 nM LacI, 100 nM CI condition and the match between data and prediction was better.

This interference effect was some 4-fold weaker than the in vivo results with the same DNA, and the 95% confidence interval for the in vivo α estimate lies completely above the 95% interval for the in vitro estimate (Figures 5 and 6G).

Loop assistance was examined by TPM with the Lac-inside nested construct (Figure 6E; *Plac*$^{+}$*.O2* was replaced with



G

F_{Lac}	F_{CI}	$F_{-/-}$	$F_{Lac/-}$	$F_{-/CI}$	$F_{Lac/CI}$	Parameter estimates		
$\frac{W_{Lac}}{1+W_{Lac}}$	$\frac{W_{CI}}{1+W_{CI}}$	$\frac{1}{Z}$	$\frac{W_{Lac}}{Z}$	$\frac{W_{CI}}{Z}$	$\frac{\alpha W_{Lac} W_{CI}}{Z}$	W_{Lac}	W_{CI}	α
Alternating								
		100 pM LacI 50 nM CI						
Observed								
0.89	0.19	0.022	0.83	0.03	0.12	34.3 ± 15.1	0.24 ± 0.022	0.62 (0.48 - 0.81)
0.97	*0.20*	*0.030*	*0.84*	*0.01*	*0.13*			
Calculated								
		20 nM LacI 100 nM CI						
0.52	0.19	0.37	0.46	0.12	0.046	1.15 ± 0.071	0.26 ± 0.020	0.39 (0.24 - 0.54)
0.53	*0.21*	*0.40*	*0.45*	*0.11*	*0.047*			
Nested								
		20 nM LacI 150 nM CI						
Observed								
0.37	0.15	0.36	0.17	0.47		0.52 ± 15.1	0.18 ± 0.022	12.9 (9.54 - 17.9)
0.34	*0.15*	*0.35*	*0.18*	*0.06*	*0.41*			
Calculated								

Fig. 6. TPM analysis of alternating loopsA and E. Setup of the TPM experiments to measure loop interference or assistance in vitro (not to scale). Distances in bp. DNA was attached to the coverslip by digoxygenin and to the bead by streptavidin (Materials and Methods).B. An example of switching between tether lengths corresponding to unlooped, LacI- or CI-looped, or both CI and LacI looped states for one tether with 20 nM LacI and 100 nM CI over a 1350 s observation interval. The lower panel shows observed values of excursion squared (blue) along with an 8 s moving average (red) during an observation interval of 1350 s. The upper panel shows a histogram for the entire observation in which four states corresponding to, from left to right, the doubly looped, CI-looped, LacI-looped and unlooped tether appear.C and D. Histograms of bead displacement $<\rho^2>$ for the alternating arrangement. *F* values were obtained as the area under each peak by Gaussian fitting. Histograms were compiled from analysis of 12-61beads (total 109-1332 min) under each condition.F. As C and D, but for the nested arrangement. CI-His6 was used instead of CI. Histograms were compiled from analysis of 25-45beads (total 224-877 min) under each condition.G. Statistical-mechanical model of loop interaction (as Figure 5). The observed *F* values (brown) from C, D and F, and the calculated *F* values (italic) using the equations shown and the fitted loop weights and α for each construct (right panel). W_{Lac}, W_{CI} and α were fitted to minimize the difference between the observed and calculated *F* values. *W* errors are standard deviations and the ranges for the α estimates are the 2.5-97.5 percentiles (>900 fitting runs).

PLac⁻.O1). Again, data was collected in the presence of LacI alone, CI alone or both proteins (Figure 6F). The fraction of the CI-looped state was substantially increased in the presence

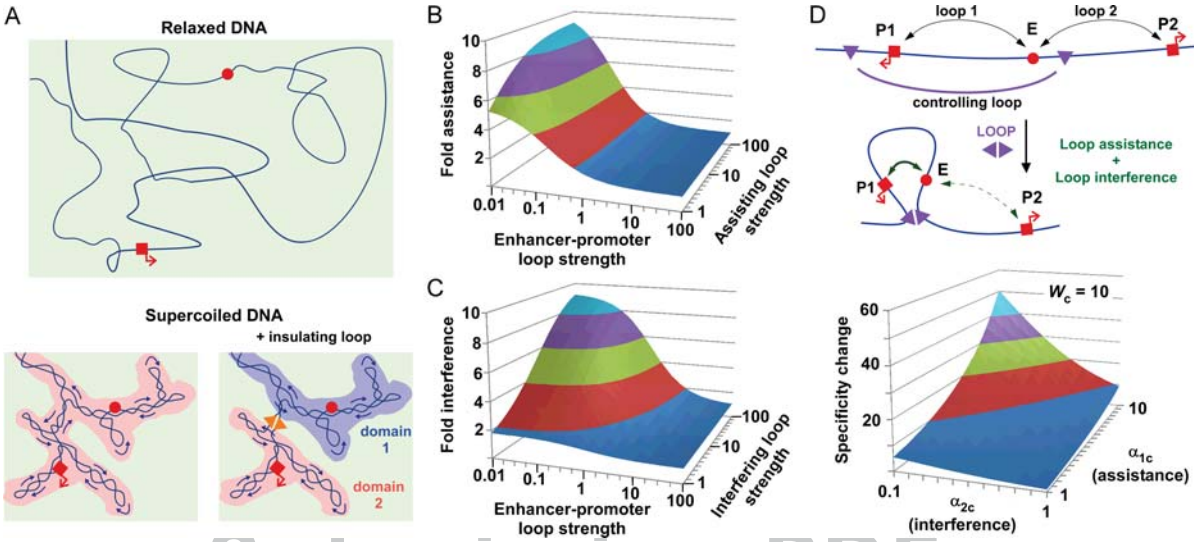


Fig. 7. Effect of DNA supercoiling and DNA loops on contact efficiency and specificity.A. In unconstrained DNA (upper panel) the red sites must find each other in a large, 3-dimensional common search volume (light green shading). DNA supercoiling (lower panels) compacts the DNA and reduces this 3D space but also creates a reduced dimensionality common search volume due to the dynamic interwinding of DNA strands (darker shading). A protein-mediated DNA clamp (right panel) prevents the supercoiling-induced interwinding between strands inside and outside the loop, separating the low dimensionality search volumes so that the sites must again search for each other in the larger 3-D space.B and C. The strength of loop assistance and loop interference in controlling enhancer-promoter (E-P) looping is dependent on the strength of the E-P loop (its weight) and the strength (weight) of the assisting or interfering loop. Derivations are given in Figure S5.

of LacI. Note that this state is a mixture of two forms, with the state of the internal Lac loop invisible to TPM.

Fitting to the loop interaction model gave a large α value of ~13 (9.5-18), indicative of strong loop assistance (Figure 6G).

Discussion

Mechanism of loop interference

Our results provide important support for the loop domain model and show that insulation is not restricted to complex regulatory elements in metazoan genomes but can occur by loop interference between relatively simple DNA-looping protein binding sites. Further experiments will be needed to test whether the insulation effects we see for 1.2 and 1.8 kb loops extend to much longer loops.

The asymmetrical alternating construct gave ~10-fold loop interference in vivo, ~4-fold stronger than the interference seen for the same construct in vitro, with non-overlapping 95% confidence intervals for the in vitro and in vivo estimates of α. This implies an important role of some in vivo factor that affects DNA structure. Our favoured explanation is DNA supercoiling, which would be present in our in vivo assays but absent in TPM. Brownian dynamics simulations show that DNA supercoiling compacts DNA such that the 'search volume' for any two sites on the DNA to find each other is considerably reduced, perhaps by 10-100 fold (55, 56) (Figure 7A). Enhancement of protein-mediated DNA looping by supercoiling has been shown in vitro (56-58) and also stimulates recombination between distant sites in vivo (59). Looping enhancement by supercoiling is also consistent with our measurements of a 5- to 10-fold increase in the efficiency of long-range LacI looping of DNA in vivo compared with relaxed DNA in vitro (24). Much of the enhancement of search volume by DNA supercoiling is likely to be lost when the sites are in separate topological domains (Figure 7A), such as formed by a protein-mediated loop (60). Indeed, LacI loop inhibition of NtrC-promoter contact was dependent on DNA supercoiling of the plasmid template (47). Further experiments are needed to confirm the involvement of DNA supercoiling in the loop interference we measured in vivo. Of particular interest is whether the prevalent DNA supercoiling in eukaryotic genomes (61) plays a role in the efficiency and specificity of enhancer-promoter contact.

We did not expect to see any loop interference in vitro with relaxed DNA, so the ~2.5-fold loop interference effect is intriguing. Whether this supercoiling-independent effect is due to specifics of the TPM setup (for example, the inability of the bead to pass through a loop) or contributes to loop interference in vivo is not clear. It has been proposed that entropic effects can drive DNA circles apart when they are in a confined volume (62).

Mechanism of loop assistance

The loop assistance we observed in vivo is consistent with the distance-shortening effect of nested loops. The statistical weight of a looped species is directly related to the factor j_{LOOP}, the effective concentration of one looping site relative to the other, which quantitates the DNA-tethering effect (24, 51). We have shown that j_{LOOP} for LacI and CI loops over the 600-10000 bp range in *E. coli* cells is roughly inversely proportional to the DNA separation (24). On this basis, j_{LOOP} for external sites of nested loops should increase by a factor proportional to the fold-change in separation. For the LacI-CI-CI-LacI arrangement (construct 10, Figure 4A) this is 2090/690=3.0 (ignoring any length addition due to the loop junction), while for the CI-LacI-LacI-CI arrangement (construct 11, Figure 4B) this is 2000/600=3.3, reasonably close to the fitted α = 3.0 and 1.9 values, respectively (Figure 5). Note that the statistical-mechanical model implies that loop assistance is reciprocal – the formation of the inner loop is also stimulated by the formation of the outer loop. The formation of the external loop does not change the direct DNA tether between the internal sites, but provides a second DNA tether (containing the loop junction) that connects the internal sites. In our constructs, this second DNA tether provided a significant shortening of the distance between CI sites, being 2.3-fold (1400/600) shorter than the unlooped distance.

The loop assistance in vitro, α = 13, was ~2-fold greater than we expected on the basis of distance-shortening. In previous TPM experiments with LacI, we found that j_{LOOP} decreased with distance to the power 1.5 (24). Since the internal loop brings the CI sites 3.3-fold closer together (2000/600 bp), we expected that the distance-shortening effect alone would give $\alpha = 3.3^{1.5} = 6$. The higher assistance seen may be due to specific angles imposed on the two 300 bp DNA 'arms' as they exit the internal LacI tetramer (63), which, combined with the relative stiffness of DNA in vitro (persistence length ~150 bp (22)) may tend to juxtapose the CI sites.

Maximizing loop interaction effects

The loop interactions in our in vivo experiments had relatively mild effects on regulation of transcription, with the largest expression changes roughly 2-fold. This is primarily because the LacI and CI loops were similar in strength, so that the primary gene regulatory loop – the one regulating the promoter – either resists the effects of the interfering loop or forms reasonably efficiently without help from the assisting loop. Modeling shows that the effect of a second interfering or assisting loop on a primary gene regulatory loop is maximized when the primary loop is weak ($W_1 << 1$) and the interfering/assisting loop is stronger than the primary loop $W_2 >> W_1$ (Figures 7B, 7C and S5). The maximum fold assistance is α and the maximal fold interference is 1/α. For nested loops, keeping the distance between each pair of different sites small and the length of the internal loop large should maximize α and at the same time help keep the primary loop weak. However, a long internal loop would also tend to make the assisting loop weak as well, so that mechanisms to strengthen the assisting loop, such as using strongly interacting proteins or multiple interacting sites, would help to maximize loop assistance. Thus, DNA looping elements that target an enhancer to a promoter would be most effective if the enhancer-promoter interaction by itself is weak, if the targeting elements are located close to the promoter and enhancer, and if the targeting elements loop strongly to each other. Long distances between the enhancer and promoter may be useful in minimizing promoter activity and making it strongly dependent on loop assistance.

Similarly, the maximal regulatory effect of an interfering loop would be achieved if the primary loop is naturally weak and the interfering loop is very strong. Systematic testing will be required to determine whether different geometries of the alternating arrangement can increase loop interference beyond the 10-fold effect (α=0.1) we saw. However, although looping efficiencies are not known, a 10-fold effect is potentially sufficient to account for observed insulator effects (33-35) and to explain the reductions in contact probability seen across TAD boundaries (39, 41).

Achieving enhancer-promoter specificity by loop interactions

Our analysis suggests that large changes in enhancer-promoter specificity could be caused by a single DNA loop that combines loop assistance and loop interference. We imagined an enhancer that is able to interact with either one of two promoters, with this contact regulated by a controlling loop that simultaneously assists the enhancer to loop to one promoter and interferes with looping to the other (Figure 7D). We define a specificity change factor *S*, which is the fold change in promoter preference due to the presence of the controlling loop. Using a three-loop model, we can calculate that high *S* values (large changes in the specificity) are possible with moderate α factors and strengths of controlling loops (Figures 7D and S5). Interestingly, the strengths of the enhancer-promoter loops are not important in determining the specificity change; the critical parameters are the α values for the assisting and interfering effects of the controlling loop on the E-P loops, (α_{1c} and α_{2c}), as well as the strength of the controlling loop (W_c). The maximal specificity change obtainable is given by the ratio of the assisting and intefering α values (α_{1c}/α_{2c}), which is approached as the controlling loop gets stronger. Thus, interactions between DNA loops provide a potentially powerful mechanism for regulating enhancer-promoter specificity.

Materials and Methods

Strain constructions

The parent strain for all reporter assays was E4643, which was constructed from BW30270 (CGSC7925) MG1655 *rph*$^+$ by precise deletion of *lacIZYA* (27). The *PlacUV5.lacO2* and λ*OR.PRM* promoters and the *lacOid* and λ*OL* sequences were inserted into pIT-HF-CL.lacZ*O2*$^-$, a modular DNA looping *lacZ* reporter chassis (24); Fig S1A and S2), and integrated into the λ*attB* site. LacI at ~18 nM was expressed from a *PlacI.lacI*$^+$ unit in pIT3-SH-lacI$^+$ integrated at φHK022 *attB* (24); Fig S1A). Lambda CI was expressed from a PRM.cI pIT3-TO-λcI-OL3-4 construct integrated at 186 *attB*, that produces 3.3 ± 0.53 wild-type lysogenic units (WLU) of CI (27); Figure S1B). Strains not expressing LacI or CI contained integrated empty vectors.

LacZ assays

A microtitre plate-based kinetic assay (64) was used, with strains grown at 37 °C to late log phase in minimal medium (1 x M9 salts [10×M9 salts = 67.8g of NaH_2PO_4, 30.0g of KH_2PO_4 and 5g NaCl/L H_2O], 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.01 mM $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$, 0.4 % glycerol).

TPM

TPM experiments were performed as described previously (53). The DNA fragment was prepared using PCR with biotin- and digoxigenin-labeled primers and was attached to an anti-digoxygenin coated overslip and streptavidin-coated polystyrene beads (160 nm radius, Spherotech Inc., Lake Forest, USA). The motion of multiple beads exposed to purified LacI (obtained from Prof. Kathleen Matthews), CI or CI-His6 (65) in 10 mM Tris–HCl pH 7.4, 200 mM KCl, 5% DMSO, 0.1 mM EDTA, 0.2 mM DTT, 0.1 mg/ml α-casein was recorded. Drift correction used immobile beads and only symmetrically moving beads were analysed further. The mean square excursion of each bead was calculated using the formula $\langle \rho^2 \rangle_{8s} = \langle ((x - \langle x \rangle_{8s})^2 + (y - \langle y \rangle_{8s})^2) \rangle_{8s}$. Histograms were compiled from combined recordings of multiple beads and the fractions of each species determined by Gaussian peak fitting.

Modeling

In vivo F values were obtained by fitting the reporter data with previously described statistical-mechanical models of LacI- and CI-mediated DNA looping regulation of *PlacUV5* and λ*PRM* to obtain the key DNA-looping parameters j_{LOOP} (for LacI) or ΔG_{oct} (for CI) for each construct (24). These and the fixed system parameters allow the fractions of all species to be determined and thus the fraction that are looped to be calculated.

For the loop interaction modeling, a Monte-Carlo fitting procedure was used to find values for W_{lac}, W_{CI}, and α that minimized Σ((*observed*–*expected*)^2/*expected*). Varied *F* values were used in repeated fittings to propagate the uncertainty in the *F* estimates.

Further details are provided in SI Materials and Methods and Figures S1 and S2.

Acknowledgments.

We thank Laura Finzi, Iain Murchland, Julian Pietsch and other members of the Shearwin, Dunlap and Finzi laboratories and Kim Sneppen, for discussions, and Kathleen Matthews for LacI. Support was from a Human Frontiers Scientific Program grant (RGP0051; KES and DD), a University of Adelaide PhD scholarship (DGP), the Australian Research Council (DP110100824 and DP11010470), the WH Elliott Fellowship in Biochemistry (IBD) and the Australian NHMRC (GNT1025549; IBD), and the National Institutes of Health (RGM084070A), the Center for Pediatric Nanomedicine in the Department of Biomedical Engineering, Georgia Institute of Technology and Children's Healthcare of Atlanta, GA (DD).

1. Lettice LA, *et al.* (2002) Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A* 99(11):7548-7553.
2. Nobrega MA, Ovcharenko I, Afzal V, & Rubin EM (2003) Scanning human gene deserts for long-range enhancers. *Science* 302(5644):413.
3. Jin F, *et al.* (2013) A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*.
4. Maeda RK & Karch F (2003) Ensuring enhancer fidelity. *Nat Genet* 34(4):360-361.
5. Li G, *et al.* (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148(1-2):84-98.
6. Kieffer-Kwon K-R, *et al.* (2013) Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* 155(7):1507-1520.
7. Marinić M, Aktas T, Ruf S, & Spitz F (2013) An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev Cell* 24(5):530-542.
8. Ptashne M (1986) Gene regulation by proteins acting nearby and at a distance. *Nature* 322(6081):697-701.
9. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, & de Laat W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* 10(6):1453-1465.
10. Carter D, Chakalova L, Osborne CS, Dai Y-f, & Fraser P (2002) Long-range chromatin regulatory interactions in vivo. *Nat Genet* 32(4):623-626.
11. Bulger M & Groudine M (2011) Functional and mechanistic diversity of distal transcription

enhancers. *Cell* 144(3):327-339.
12. Krivega I & Dean A (2012) Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev* 22(2):79-85.
13. Bellen HJ, *et al.* (1989) P-element-mediated enhancer detection: a versatile method to study development in Drosophila. *Genes Dev* 3(9):1288-1300.
14. Ruf S, *et al.* (2011) Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat Genet* 43(4):379-386.
15. Shen Y, *et al.* (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature* 488(7409):116-120.
16. Dunham I, *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57-74.
17. Nord AS, *et al.* (2013) Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* 155(7):1521-1531.
18. Kwon D, *et al.* (2009) Enhancer-promoter communication at the Drosophila engrailed locus. *Development* 136(18):3067-3075.
19. Visel A, *et al.* (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457(7231):854-858.
20. Higgins NP, Yang X, Fu Q, & Roth JR (1996) Surveying a supercoil domain by using the gamma delta resolution system in Salmonella typhimurium. *J. Bacteriol.* 178(10):2825-2835.
21. Ringrose L, Chabanis S, Angrand PO, Woodroofe C, & Stewart AF (1999) Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *EMBO J* 18(23):6630-6641.
22. Rippe K (2001) Making contacts on a nucleic acid polymer. *Trends Biochem Sci* 26(12):733-740.
23. Lieberman-Aiden E, *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950):289-293.
24. Priest DG, *et al.* (2014) Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and λ repressors. *Proc Natl Acad Sci U S A* 111(1):349-354.
25. Calhoun VC, Stathopoulos A, & Levine M (2002) Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. *Proc Natl Acad Sci U S A* 99(14):9243-9247.
26. Deng W, *et al.* (2012) Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149(6):1233-1244.
27. Cui L, Murchland I, Shearwin KE, & Dodd IB (2013) Enhancer-like long-range transcriptional activation by λ CI-mediated DNA looping. *Proc Natl Acad Sci U S A* 110(8):2922-2927.
28. Nolis IK, *et al.* (2009) Transcription factors mediate long-range enhancer-promoter interactions. *Proc Natl Acad Sci U S A* 106(48):20222-20227.
29. Mahmoudi T, Katsani KR, & Verrijzer CP (2002) GAGA can mediate enhancer function in trans by linking two separate DNA molecules. *EMBO J* 21(7):1775-1781.
30. Petrascheck M, *et al.* (2005) DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Res* 33(12):3743-3750.
31. Bell AC, West AG, & Felsenfeld G (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* 291(5503):447-450.
32. Bushey AM, Dorman ER, & Corces VG (2008) Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Mol Cell* 32(1):1-9.
33. Hou C, Zhao H, Tanimoto K, & Dean A (2008) CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci U S A* 105(51):20398-20403.
34. Gohl D, *et al.* (2011) Mechanism of chromosomal boundary action: roadblock, sink, or loop? *Genetics* 187(3):731-748.
35. Chetverina D, Aoki T, Erokhin M, Georgiev P, & Schedl P (2014) Making connections: Insulators organize eukaryotic chromosomes into independent cis-regulatory networks. *Bioessays* 36(2):163-172.
36. Corces VG (1995) Chromatin insulators. Keeping enhancers under control. *Nature* 376(6540):462-463.
37. Cai HN & Shen P (2001) Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. *Science* 291(5503):493-495.
38. Muravyova E, *et al.* (2001) Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science* 291(5503):495-498.
39. Dixon JR, *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485(7398):376-380.
40. Nora EP, *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485(7398):381-385.
41. Sexton T, *et al.* (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* 148(3):458-472.
42. Hou C, Li L, Qin ZS, & Corces VG (2012) Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell* 48(3):471-484.
43. Sofueva S, *et al.* (2013) Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J* 32(24):3119-3129.
44. Phillips-Cremins JE & Corces VG (2013) Chromatin insulators: linking genome organization to cellular function. *Mol Cell* 50(4):461-474.
45. Akhtar W, *et al.* (2013) Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* 154(4):914-927.
46. Ameres SL, *et al.* (2005) Inducible DNA-loop formation blocks transcriptional activation by an SV40 enhancer. *EMBO J* 24(2):358-367.
47. Bondarenko VA, Jiang YI, & Studitsky VM (2003) Rationally designed insulator-like elements can block enhancer action in vitro. *EMBO J* 22(18):4728-4737.
48. Dodd IB, *et al.* (2004) Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev* 18(3):344--354.
49. Meyer BJ & Ptashne M (1980) Gene regulation at the right operator (OR) of bacteriophage lambda. III. lambda repressor directly activates gene transcription. *J Mol Biol* 139(2):195-205.
50. Becker NA, Kahn JD, & Maher r, L James (2005) Bacterial repression loops require enhanced DNA flexibility. *J Mol Biol* 349(4):716-730.
51. Han L, *et al.* (2009) Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS One* 4(5):e5621.
52. Nelson PC (2007) Colloidal particle motion as a diagnostic of DNA conformational transitions. *Curr Opin Colloid In* 12(6):307-313.
53. Kumar S, *et al.* (2014) Enhanced tethered-particle motion analysis reveals viscous effects. *Biophys J* 106(2):399-409.
54. Biton YY, Kumar S, Dunlap D, & Swigon D (2014) Lac repressor mediated DNA looping: monte carlo simulation of constrained DNA molecules complemented with current experimental results. *PLoS One* 9(5):e92475.
55. Jian H, Schlick T, & Vologodskii A (1998) Internal motion of supercoiled DNA: brownian dynamics simulations of site juxtaposition. *J Mol Biol* 284(2):287-296.
56. Polikanov YS, *et al.* (2007) Probability of the site juxtaposition determines the rate of protein-mediated DNA looping. *Biophys J* 93(8):2726-2731.
57. Liu Y, Bondarenko V, Ninfa A, & Studitsky VM (2001) DNA supercoiling allows enhancer action over a large distance. *Proc Natl Acad Sci U S A* 98(26):14883-14888.
58. Norregaard K, *et al.* (2013) DNA supercoiling enhances cooperativity and efficiency of an epigenetic switch. *Proc Natl Acad Sci U S A* 110(43):17386-17391.
59. Rovinskiy N, Agbleke AA, Chesnokova O, Pang Z, & Higgins NP (2012) Rates of gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome. *PLoS Genet* 8(8):e1002845.
60. Leng F, Chen B, & Dunlap DD (2011) Dividing a supercoiled DNA molecule into two independent topological domains. *Proc Natl Acad Sci U S A* 108(50):19973-19978.
61. Naughton C, *et al.* (2013) Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat Struct Mol Biol* 20(3):387-395.
62. Jun S & Mulder B (2006) Entropy-driven spatial organization of highly confined polymers: lessons for the bacterial chromosome. *Proc Natl Acad Sci U S A* 103(33):12388-12393.
63. Villa E, Balaeff A, & Schulten K (2005) Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. *Proc Natl Acad Sci U S A* 102(19):6783-6788.
64. Dodd IB, Perkins AJ, Tsemitsidis D, & Egan JB (2001) Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. *Genes Dev* 15(22):3013--3022.
65. Gao N, Shearwin K, Mack J, Finzi L, & Dunlap D (2013) Purification of bacteriophage lambda repressor. *Protein expression and purification* 91(1):30-36.

SUPPLEMENTARY INFORMATION

SI Materials and Methods

DNA constructions

Strains, expression constructs and reporters were essentially as described in (1).

The parent strain for all reporter assays was E4643, which was constructed from BW30270 (CGSC7925) MG1655 *rph*$^{+}$ by precise deletion of *lacIZYA* (EcoCyc MG1655: 360527-366797) by recombineering (2). EC100D *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80d*lacZ*ΔM15 Δ*lac*X74 *recA1endA1 araD139* Δ(*ara,leu*)7697 *galU galK* λ^{-} *rpsL nupG pir*$^{+}$ (DHFR) (Epicentre, USA) was used for propagation of R6γK *ori* (*pir*-dependent) plasmids.

DNA constructions used PCR (primers from Geneworks, Australia), restriction enzyme based cloning and isothermal Gibson assembly (3). DNA sequences of manipulated regions were confirmed, except for some of the larger spacers, the sizes of which were confirmed by PCR.

The reporter and LacI and CI expression constructs (Fig S1; (1, 4) were made using a plasmid integration system developed from the CRIM plasmids (5), the pZ plasmids (6), and an *O2*$^{-}$ *lacZ* reporter gene (7) preceded by an RNaseIII cleavage site (8). As in the CRIM system, phage integrase proteins were used to integrate the plasmids into phage attachment sites in the bacterial chromosome. Integration was at λ*attB* (EcoCyc (9) MG1655 sequence position: 806551), ϕ80 *attB* (position 1308595), or ϕ186 *attB* site (position: 2783828) using modified CRIM integrase plasmids. PCR was used to screen for correct single-copy integrants. All sequences are available on request.

Reporter constructs

To measure DNA loop-dependent repression, a modular DNA looping *lacZ* reporter 'chassis' was designed to be amenable to both restriction enzyme-based cloning and Gibson isothermal assembly (Fig S1A). The various chassis structures were inserted into pIT-HF-CL.lacZ and were integrated into the λ*attB* site. LacZ expression was from *PlacUV5.O* or *OR.PRM* modules, with various *lac* or λ CI operator modules (Figure S2) located elsewhere on the chassis. The chassis is invertible; when module 1 points towards *lacZ,* LacI looping is measured by repression of *PlacUV5*, whereas in the inverted orientation, module 4 points towards *lacZ* and λ CI looping can be measured by repression of *PRM*.

Spacer DNA between the operator/promoter modules (Fig S1A) was made up of sequences from within the *E. coli* genes *ftsK* (position 932456-936438), *rne* (114410-1143589) and *valS* (4479008-4481858) to minimize the likelihood of incorporation of cryptic promoters. The Series 1 spacers of (1) were used for all *PlacUV5* reporters. In all *PRM* reporters, the λtI terminator was replaced with the φ186tW teminator (Figure S1A).

LacI and CI expression constructs

A DNA fragment carrying *lacI* and its natural *PlacI* promoter was obtained from the plasmid pUHA1 (6), by digestion with SalI. The fragment extended from the –78 position of *PlacI* (366837) to 18 bp downstream of the *lacI* stop codon (365634; includes *O3*), and was inserted into the SalI site of pIT3-SH (Fig S1B). Single-copy integrants of pIT3-SH or pIT3-SH-lacI^{+} at the φHK022 att site were used to provide no protein (vector-only control) or LacI respectively.

Lambda CI was expressed from pIT3-TO-λcI-OL3-4 (4) integrated at 186 *attB*. This contains a *PRM.cI.OL* module that produces 3.3 wild-type lysogenic units (WLU) of CI (Figure S1B). The control strain expressing no CI contained the empty vector pIT3-TO at 186 *attB*. The copies of *OL* and *OR* present on the expression construct do not influence the reporter at λ*attB* by DNA looping because they are located 1.98 Mbp away (1). The two CI expression constructs were integrated into the two LacI-expressing strains to give the four LacI and CI-expressing strains used in this study.

LacZ assays

LacZ assays for the DNA looping reporters were carried out using the microtitre plate method (10), with the modification that cultures grown in M9 minimal medium ('M9MM' = 1 x M9 salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.01 mM $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$, 0.4 % glycerol [10× M9 salts = 67.8g of NaH_2PO_4, 30.0g of KH_2PO_4 and 5g NaCl/L H_2O]) at 37 °C to late log phase were added to a combined lysis-assay buffer, with each well of a microtitre plate containing: 50 µL culture + M9MM (usually 20 µL culture + 30 µL M9MM), 150µL TZ8 (100 mM Tris-HCl pH 8.0, 1 mM $MgSO_4$, 10 mM KCl), 40 µL ONPG (o-nitrophenyl-β-D-galactoside 4 mg/mL in TZ8), 1.9 µL 2-mercapoethanol, 0.95µL polymyxin B (20 mg/mL Sigma). Assays were repeated at least 4 times.

Modeling of looping in vivo

Estimating fractional LacI looping from reporter data

The LacI looping model is described in (1). It is a statistical-mechanical model that involves 7 LacI- or RNAP-bound species for the *Oid*–promoter.*O2* DNA. The fractions of all these species at a fixed [LacI], including the looped species, and the resultant LacZ activity expressed from the promoter, can be calculated from 7 parameters. As the strains and conditions were exactly as in (1), we used the same fixed *lac* system parameter values (± SD): dissociation constant for *O2*, *K2* = 3.66 ± 0.27 nM; dissociation constant for *Oid*, *Kid* = 0.118 ± 0.012 nM; [LacI], *L1* = 18 nM; weight for RNAP occupation of the promoter, *R* = 4.97 ± 0.22; background LacZ units of reporter, *bkg* = 18.1 ± 1.14 lacZ units. Values for maximal LacZ units from *PlacUV5.O2, Zmax2,* and the DNA looping factor, j_{LOOP}, were fitted.

Each of the reporter data sets, comprising 4 data points (LacZ units ± LacI/±CI) was fitted by fixing the parameter estimates for *R*, *L1*, *Kid*, *K2* and *bkg* and by fitting a single *Zmax2* value (to allow for day-to-day variation in the LacZ assay) and two j_{LOOP} values, one in the absence and one in the presence of CI. The fitting algorithm minimized $\Sigma((observed–expected)^2/expected)$, comparing the model-generated expected promoter LacZ values with the observed LacZ data points, giving equal weight to all points. In a Monte Carlo algorithm, the three variable parameter values were iteratively varied by successive random steps (usually 10^6 iterations), with each modified parameter set accepted if the fit improved. Fitting convergence was tested by repeating the fitting process with different randomly chosen initial variable parameter values and was found to be highly reliable. After the fit was obtained, the fraction looped in the absence or presence of CI, F_{Lac} or $F_{Lac(CI)}$, respectively was calculated by summing the fractions of looped and unlooped species.

To gain an appreciation of the error in the estimates of the *F* values, we repeated the fitting process many times, each with slightly different values for the fixed parameters and the LacZ data, a process we term 'jiggling' (1). Each of the 4 data points used in the fitting was changed randomly based on the *t*-distribution, according to its measured standard error of the mean (SEM) and number of observations (*n*). We also incorporated the uncertainty in the fixed parameter estimates of Priest et al. by randomly jiggling their values based on their standard deviations and the normal distribution. In this way, we sampled how variation in the data, the fixed parameters and the fitting procedure combine to affect the *F* estimates. The LacI *F* values listed in the figures are the means and standard deviations from 100 good-scoring jiggled fits.

Estimating fractional CI looping from reporter data

The statistical-mechanical CI looping model is described in (1), and is a simplification of the 192-species model of Cui et al. (2013) due to the absence of the *OL* UP element. We used the same parameter values used by (1), with one exception. The basal activity of *PRM* was higher in this study compared to previously (1), possibly because we assayed the CI looping reporters in minimal medium rather than rich medium. This prevented us getting satisfactory fits to the data using the old parameters. We found that good fits could be obtained with a lower value for ΔG_{PRM} (0.42 ± 0.05 rather than 1.18 ± 0.05 kcal/mol). This parameter defines the basal propensity of RNAP to bind to *PRM* and thus the basal activity of *PRM*.

The CI *F* estimates were obtained similarly to the LacI *F* values. Each fitting used eight data points: four LacZ units (±CI/±LacI) from the OL^- construct (construct 7, Figure 3C), and four LacZ units (±CI/±LacI) for the specific OL^+ construct. The +CI concentration was 3.3 WLU. As for LacI, most of the parameters were held fixed, while a single *max_prm* value (the maximal activity of *PRM*) and two ΔG_{oct} values (the free energy for CI octamerization and *OL-OR* looping), one in the absence and one in the presence of LacI, were fitted. After the fit was obtained, the fraction looped in the absence or presence of LacI, F_{CI} or $F_{CI(Lac)}$, respectively, was calculated by summing the fractions of looped and unlooped species. To gain an idea of the error in the *F* estimates, both the fixed parameter values and the data means were jiggled, as for the LacI loop fitting. The CI system differs from the LacI system because one of the fixed parameters ΔG_{tet} (the free energy of tetramerization between a CI dimer at *OL* and a CI dimer at *OR*) has a strong effect on DNA looping, and jiggling of ΔG_{tet} results in roughly parallel changes in F_{CI} or $F_{CI(Lac)}$ within a single fit. Thus, the CI *F* values listed in the figures are the means and standard deviations of the F_{CI} values from at least 50 good-scoring jiggled fits, and the $F_{CI(Lac)}$ values calculated as the mean and standard deviation of the *increment* from the mean F_{CI} value.

Estimating loop weights and α from the obtained F values

The fractions of the four species in the loop interaction model (Figure 5) are determined by three unknown parameters: W_{lac}, W_{CI}, and α, which are specific to each particular site arrangement and LacI and CI concentration. Expected values for each of the experimentally observable loop fractions can be simply calculated from these parameters (Figures 5 and 6D). Thus a Monte-Carlo fitting procedure similar to that used for determining loop fractions (above) was used to find values for W_{lac}, W_{CI}, and α that minimized $\Sigma((observed–expected)^2/expected)$.

As before, the fitting runs were repeated with the observed *F* values jiggled according to their standard deviations and the normal distribution. To obtain jiggled in vivo $F_{CI(Lac)}$ values, the increment value was jiggled and added to the jiggled F_{CI} value for the run.

The different in vivo data and the in vitro data provide different combinations of observed *F* values that can be used for fitting (Figures 5 and 6D). For the alternating and side by side constructs, F_{CI}, $F_{CI(Lac)}$, F_{Lac}, and $F_{Lac(CI)}$ were used. For the nested constructs, we could not measure the *F* values for the internal loop. We made an estimate of the internal F_{CI} value for the *OL-OR* loop by interpolation from other *OL-OR* loop measurements (Figure S4A). The F_{Lac} value for the internal 1400 bp *Oid-O1* loop was estimated by interpolation of the $j_{LOOP}/L1$ parameter values obtained from the model fitting of the data from four *Oid-O2* loops (Figure S4B). This $j_{LOOP}/L1$ estimate for the 1400 bp loop allows calculation of the expected F_{Lac} for the 1400 bp Oid-O1 loop, using our statistical-mechanical model and parameters previously obtained for the LacI system (1). The TPM experiments provided *F* measurements for LacI and CI alone (F_{CI} and F_{Lac}) and also for each of the four species in the presence of both proteins (unlooped $F_{-/-}$, LacI looped only $F_{Lac/-}$, CI looped only $F_{-/CI}$, both looped $F_{Lac/CI}$).

The W_{lac}, W_{CI}, and α values listed in the figures are the *W* means and standard deviations, and the median and 2.5% and 97.5% percentile α values, from at least 800 good-scoring jiggled fits.

Programs were written in Fortran 77 and run in a Cygwin environment on a laptop computer, and are available on request.

TPM methods

DNA preparation

DNA fragments were prepared using the polymerase chain reaction (PCR) with plasmid DNA as templates, deoxyribonucleotides (Fermentas-Thermo Fisher Scientific Inc., USA), biotin- and digoxigenin-labeled primers (Life Technologies Corporation, Grand Island, NY or Integrated DNA Technologies Inc., Coralville, Iowa) with Taq polymerase (New England BioLabs Inc., Ipswich, MA). All amplicons were purified by using silica-membrane based purification kits (Qiagen, Germantown, MD) and the lengths were checked by gel electrophoresis. The details of the plasmids and primers are available on request.

TPM experiments

TPM experiments were performed as described previously (11); Biton et al., 2014). Briefly, chambers were assembled between a coverslip and a microscope slide using a parafilm spacer. To attach tethers to the glass surface, 50-100 pM of DNA labeled with digoxigenin on one end and biotin on the other end (Figure 5A) were incubated in the chamber filled with 10 mM Tris-HCl pH 7.4, 200 mM KCl (binding buffer) for 15 minutes. Then 100 µL of 20 mg/mL antidigoxigenin (Roche, Basel, Switzerland) in phosphate-buffered saline was introduced into the chamber and incubated at high humidity for ~2 h at room temperature or overnight at 4°C. Then the chamber was gently flushed with three volumes of 10 mM Tris-HCl pH 7.4, 200 mM KCl, 0.5mg/ml α-casein (wash buffer) and a solution of 0.5-1 nM streptavidin-coated polystyrene beads (160 nm radius, Spherotech Inc., Lake Forest, USA) in binding buffer was introduced for a 10 minute incubation. The chamber was gently flushed with three volumes of wash buffer, two of 10 mM Tris–HCl pH 7.4, 200 mM KCl, 5% DMSO, 0.1 mM EDTA, 0.2 mM DTT and 0.1 mg/ml α-casein (λ buffer) and then with 2 volumes of LacI or CI in λ buffer to make sure that the chamber had the desired protein and buffer composition. The chamber was incubated for 30 minutes on the microscope and then 3-5 immobile beads and 3-20 mobile beads were tracked.

Data preprocessing – drift calculations and symmetry selection

Immobile beads that were successfully tracked for the entire recording were used for drift calculations. A 4 sec moving average (center of mass) for selected immobile beads in a video frame was calculated and subtracted from each bead in the same field of view to remove the drift. Immobile beads exhibit very low amplitude motion and a very short time averaging window (4s as opposed to 40s for mobile beads) can be used to determine the anchor point of the bead very accurately. This increases the cutoff frequency of drift motion that can be subtracted while minimizing attenuation of the motion of tethered beads. The drift-corrected recordings of tethered beads were then analyzed for symmetry before selection for further analysis.

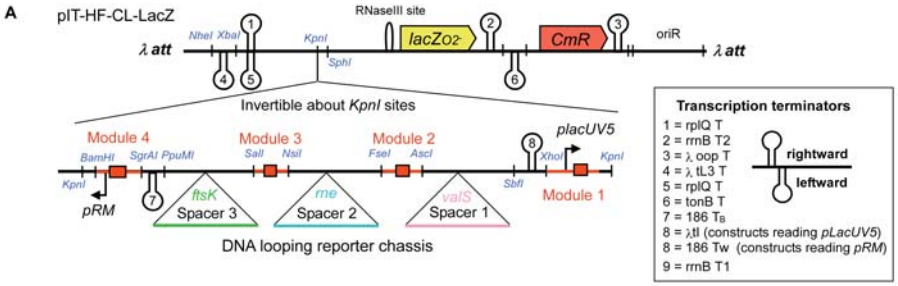
Analysis

The mean square excursion of the bead was calculated using the formula . In plots of and versus time for individual beads, steps in the amplitude signal loop formation and breakdown, and the histogram of the mean excursion shows the time spent in each state. A calibration curve (1) was used to verify the magnitude of the looped and unlooped excursion for each recording. CI can form stable non-specific loops in vitro that shorten the tether length arbitrarily. To avoid including these in our

analysis, any recordings that contained stable excursions of unexpected magnitudes were discarded. All selected recordings in the same experimental condition were concatenated and an overall histogram of tether lengths was plotted (Figures 6C, 6D and 6F). The looping probabilities were determined by fitting the peaks in histograms of the measured tether lengths for the concatenated traces of one condition.

SI References

1. Priest DG*, et al.* (2014) Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and λ repressors. *Proc Natl Acad Sci U S A* 111(1):349-354.
2. Datsenko KA & Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* 97(12):6640-6645.
3. Gibson DG*, et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343-345.
4. Cui L, Murchland I, Shearwin KE, & Dodd IB (2013) Enhancer-like long-range transcriptional activation by lambda CI-mediated DNA looping. *Proceedings of the National Academy of Sciences of the United States of America* 110(8):2922-2927.
5. Haldimann A & Wanner BL (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol* 183(21): 6384-6393.
6. Lutz R & Bujard H (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* 25(6):1203-1210.
7. Müller J, Oehler S, & Müller-Hill B (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J Mol Biol* 257(1):21-29.
8. Linn T & St. Pierre R (1990) Improved vector system for constructing transcriptional fusions that ensures independent translation of *lacZ*. *J. Bacteriol.* 172(2):1077-1084.
9. Keseler IM*, et al.* (2011) EcoCyc: a comprehensive database of Escherichia coli biology. *Nucleic Acids Res* 39(Database issue):D583-590.
10. Dodd IB, Perkins AJ, Tsemitsidis D, & Egan JB (2001) Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. *Genes Dev.* 15(22):3013-3022.
11. Kumar S*, et al.* (2014) Enhanced tethered-particle motion analysis reveals viscous effects. *Biophys J* 106(2):399-409.



Construct	Function	Module 4	Spacer 3 (bp)	Module 3	Spacer 2 (bp)	Module 2	Spacer 1 (bp)	Module 1
Constructs assaying LacI looping (reading from *pLacUV5*)								
Fig 3B, 1	LacI distal operator minus	O^-	300	O^-	300	O^-	300	*pLacO2*
Fig 3B, 1	LacI distal operator minus	O^-	300	O^-	1400	O^-	1800	*pLacO2*
Fig 3B, 2	CI loop interference on LacI symmetrical	*OR+(pR-)321*	600	*Oid*	600	*OL+(pL-)123*	600	*pLacO2*
Fig 3B, 3	CI *OR* outside LacI loop symmetrical	*OR+(pR-)321*	600	*Oid*	600	O^-	600	*pLacO2*
Fig 3B, 4	CI *OL* inside LacI loop symmetrical	O^-	600	*Oid*	600	*OL+(pL-)123*	600	*pLacO2*
Fig 3B, 5	CI loop interference on LacI asymmetrical	*OR+(pR-)321*	1500	*Oid*	300	*OL+(pL-)123*	900	*pLacO2*
Fig 3B, 6	CI *OL* inside LacI loop asymmetrical	O^-	1500	*Oid*	300	*OL+(pL-)123*	900	*pLacO2*
Fig 4A, 10	CI loop nested within LacI loop	*Oid*	300	*OL+(pL-)123*	1408	*OR+(pR-)321*	387	*pLacO2*
Fig 4C, 12	LacI and CI loops side-by-side	*OR+(pR-)321*	1500	*OL+(pL-)123*	300	*Oid*	300	*pLacO2*
Constructs assaying CI looping (reading from *pRM*, *inverted about KpnI site*)								
Fig 3C, 7	CI distal operator minus (OL^-)	*OR+(pR-)321*	600	*Oid*	600	O^-	600	*pLacO2*
Fig 3C, 8	LacI loop interference on CI symmetrical	*OR+(pR-)321*	600	*Oid*	600	*OL+(pL-)123*	600	*pLacO2*
Fig 3C, 9	LacI loop interference on CI asymmetrical	*OR+(pR-)321*	1500	*Oid*	300	*OL+(pL-)123*	900	*pLacO2*
Fig S3	LacI *O2* outside CI loop symmetrical	*OR+(pR-)321*	600	O^-	600	*OL+(pL-)123*	600	*pLacO2*
Fig S3	LacI *Oid* inside CI loop symmetrical	*OR+(pR-)321*	600	*Oid*	600	*OL+(pL-)123*	600	*pLacO*$^-$
Fig 4B, 11	LacI loop nested within CI loop	*OR+(pR-)321*	300	*Oid*	1400	*O1*	300	*OL+(pL-)123*
Fig 4D, 13	LacI and CI loops side-by-side	*OR+(pR-)321*	1500	*OL+(pL-)123*	300	*Oid*	300	*pLacO2*

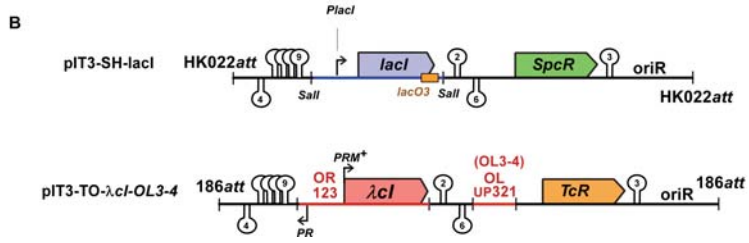


Fig S1. Reporter and expression constructs

A. Reporter vector and chassis. The basic structure of the integrated reporter vector and inserted looping chassis is shown, together with relevant restriction sites and transcription terminators. Modules contained operators (red boxes) for either LacI or CI or a standard operator minus sequence (O^-). CmR is chlorampenical resistance. The table gives the structure of the individual reporter constructs. Spacer lengths are measured between the centres of the operators within the modules (*OL2* and *OR2* for λ).

B. Structure of chromosomally integrated expression constucts for LacI and CI. SpcR is spectinomycin resistance; TcR is tetracycline resistance. Terminators as in A.

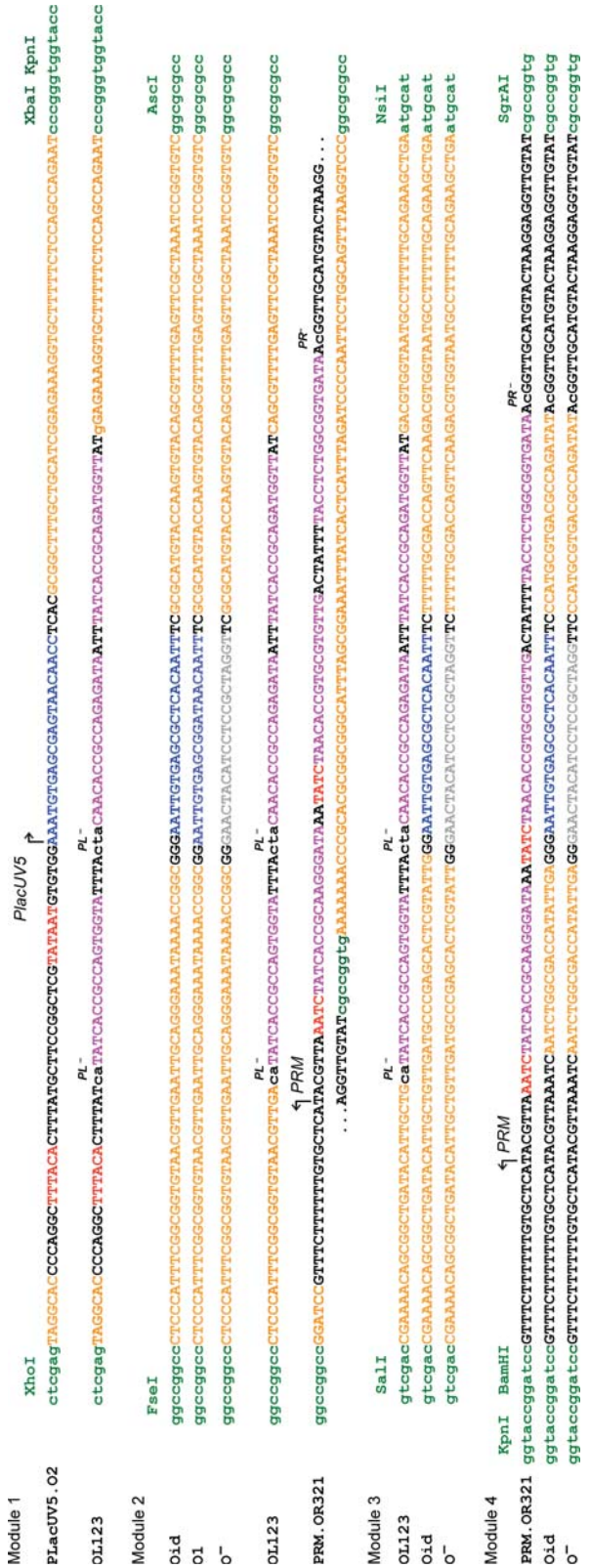


Fig S2. Module sequences
Each 120 or 121 bp module is listed in the orientation shown in Fig S1 and flanked by chassis restriction sites (lowercase green). Non-*lac* or λ sequences are in brown. *Lac* operators are shown in blue (*lacO*$^{-}$ in grey) and CI operators in purple. Promoter sequences (–10 and –35) for *PlacUV5* and *PRM* are in red; lowercase bases are mutations used to inactivate promoters.

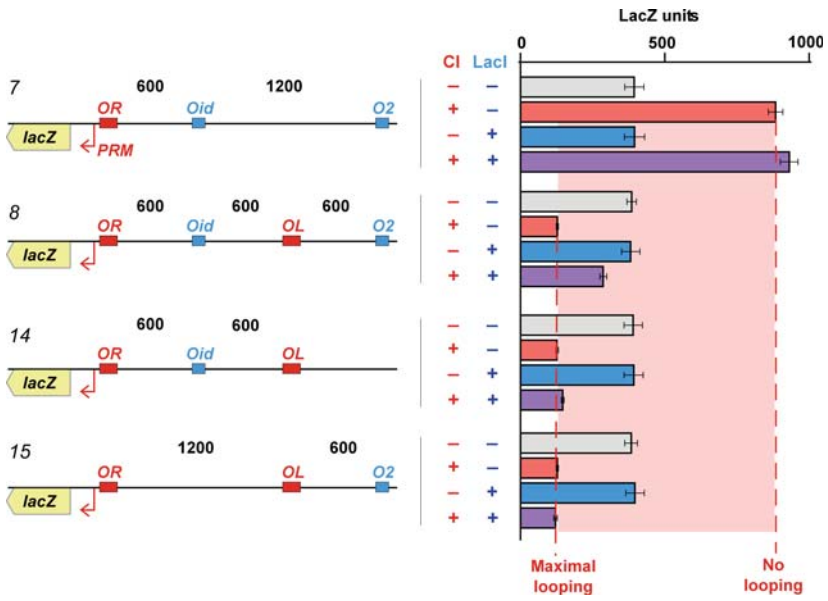


Fig S3. Controls for alternating loop constructs.
Controls for Figure 3, showing that LacI inhibition of CI looping requires two lac operators flanking the CI *OL* site. Details as in Figure 3. Histograms show steady-state LacZ units and 95% confidence intervals (*n*=9).

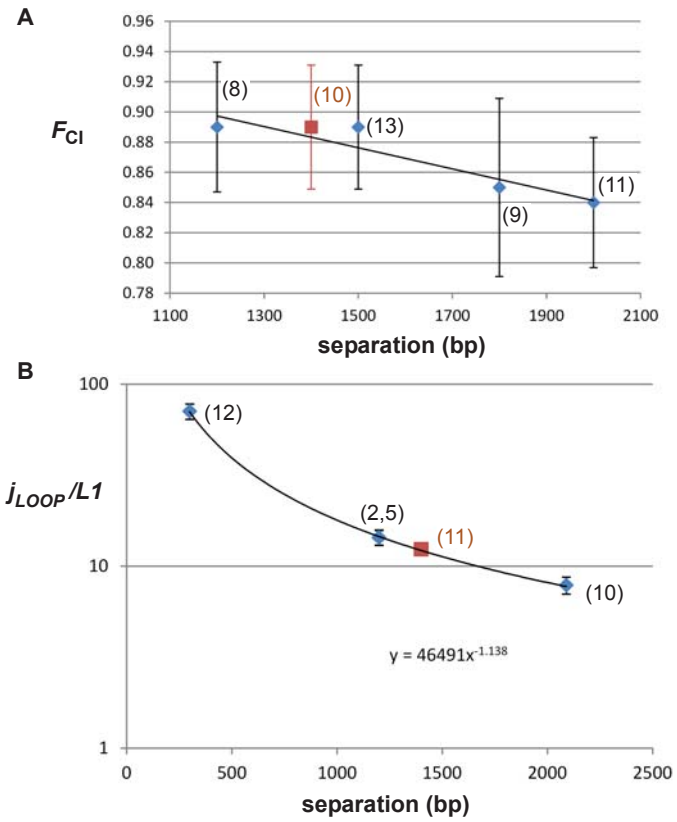


Fig S4. Estimating *F* for internal loops for the statistical-mechanical fitting in Figure 5.
A. Interpolation between measured *F* values for CI looping (see Figures 2, 3 and 4; construct numbers indicated) was used to estimate the *F* value for the internal 1400 bp CI loop in construct 10 (Figure 4A).

B. Interpolation of $j_{LOOP}/L1$ values obtained from model fitting of data for LacI *Oid-O2* looping (see Figures 2, 3 and 4; construct numbers indicated) gives $j_{LOOP}/L1$ = 12.4 for a 1400 bp loop. As $L1$=18 nM, $j_{LOOP}(1400)$ = 223.2 nM. .

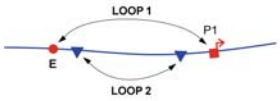
Two loops:

In the absence of Loop 2 there are 2 species:

Species	Weight
unlooped	1
loop1	W_1

The fraction loop 1 in the absence of loop 2 is:

$$F_{no\,2} = W_1/(1+W_1)$$

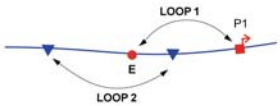


In the presence of Loop 2 there are 4 species:

Species	Weight
unlooped	1
loop1	W_1
loop2	W_2
loop1+2	$\alpha.W_1.W_2$

The fraction loop 1 in the presence of loop 2 is:

$$F_2 = (W_1 + \alpha.W_1.W_2)\,/(1+W_1+W_2+\alpha.W_1.W_2)$$



So the **assistance** effect *A* (the fold increase in loop 1) due to loop2 is:

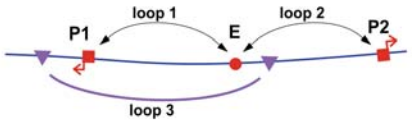
$$A = F_2 / F_{no2} = \frac{(W_1 + \alpha.W_1.W_2)}{(1+W_1+W_2+\alpha.W_1.W_2)} \times \frac{1+W_1}{W_1}$$

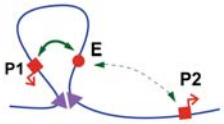
$$= \frac{(1 + \alpha.W_2)(1+W_1)}{(1+W_1+W_2+\alpha.W_1.W_2)}$$

And the **interference** effect *I* (the fold decrease in loop 1) due to loop2 is:

$$I = F_{no2} / F_2 = \frac{(1+W_1+W_2+\alpha.W_1.W_2)}{(1 + \alpha.W_2)(1+W_1)}$$

Controlling specificity:





We assume that Loop 1 and Loop 2 are mutually exclusive i.e. the enhancer can only loop to one promoter at a time

In the absence of Loop 3 there are 3 species:

Species	Weight
unlooped	1
loop1	W_1
loop2	W_2

The fractions looped are:

$$F_1 = W_1/(1+W_1+W_2) \qquad F_2 = W_2/(1+W_1+W_2)$$

The **preference** for loop1 over loop2, $P_{no3} = F_1/F_2 = W_1/W_2$ (1)

In the presence of Loop 3 there are 6 species:

Species	Weight
unlooped	1
loop1	W_1
loop2	W_2
loop3	W_3
loop1+3	$\alpha_{13}.W_1.W_3$
loop2+3	$\alpha_{23}.W_2.W_3$

α_{13} and α_{23} are the loop interaction factors. For the above arrangement, $\alpha_{13} > 1$ (assistance) and $\alpha_{23} < 1$ (interference).

$$F_1 = (W_1 + \alpha_{13}.w1.W_3)\,/(1+W_1+W_2+W_3+\alpha_{13}.W_1.W_3 + \alpha_{23}.W_2.W_3)$$

$$F_2 = (W_2 + \alpha_{23}.W_2.W_3)\,/(1+W_1+W_2+W_3+\alpha_{13}.W_1.W_3 + \alpha_{23}.W_2.W_3)$$

The **preference** for loop1 over loop2 in the presence of loop 3 is:

$$P_3 = F_1/F_2 = (W_1 + \alpha_{13}.W_1.W_3) / (W_2 + \alpha_{23}.W_2.W_3) \qquad (2)$$

Combining (1) and (2) gives the specificity change *S* (the fold change in loop 1 vs 2 preference) due to loop3:

$$S = P_3 / P_{no3} = \frac{(W_1 + \alpha_{13}.W_1.W_3)}{(W_2 + \alpha_{23}.W_2.W_3)} \times \frac{W_2}{W_1} = \frac{(1 + \alpha_{13}.W_3)}{(1 + \alpha_{23}.W_3)} \qquad (3)$$

Fig S5. Modeling loop assistance and loop interference

Chapter 4

Methods Development

4.1 Introduction

Chapters 2 and 3 have centred on two manuscripts examining long-range DNA looping by the LacI repressor (Chapter 2) and combining LacI and λCI looping to study the loop domain model (Chapter 3). The technical core of these studies was to build *lacZ* reporter constructs, integrate them into the chromosome of *Escherichia coli* strains expressing different combinations of LacI and λCI and then perform LacZ assays. This chapter will concern the development of these tools; including the generation of a suitable reporter strain, development of a chromosomal integration system, design and construction of a modular 'DNA looping reporter chassis' and developments made to an assay to measure LacZ levels. Chapter 6 contains extended materials and methods and other reference material such as DNA sequence of PCR primers.

4.2 Plasmids for integration into the *E. coli* chromosome

E. coli has been proven over and again to be a useful model organism in which to test genetic constructs and perform genetic screens [Shuman and Silhavy, 2003]. DNA can be maintained in the *E. coli* cell either as an independent replicon (such as a plasmid) or within the host chromosome (where temperate prophage reside). Plasmids can be maintained at high copy number and are easy to retrieve, suiting them to applications such as protein expression and genetic screens. However copy number variation leads to transcriptional noise across the population [Kaern et al.,

2005], which can affect applications where precise control of protein expression is required, or where precise measurements of reporter gene expression must be made. In these cases, chromosomal integration is preferred since DNA can be stably maintained without antibiotic selection at a similar copy number between cells. As the manuscripts in chapters 2 and 3 show, the reduced variability in expression of LacI, λCI and *lacZ* afforded by chromosomal integration allowed, via model fits, determination of the fraction of time a reporter construct was in the looped state, a parameter assumed by some to be unattainable from *in vivo* systems [Schleif, 1992].

Recombineering [Datsenko and Wanner, 2000] can be used for inserting DNA into the *E. coli* chromosome, however it seems to have found greater application in making small insertions, mutations and deletions [Wang et al., 2009] (however a recent study [Sabri et al., 2013] used recombineering to make larger insertions). Another system specifically for integrating DNA cassettes into the *E. coli* chromosome, known as the conditional-replication, integration, and modular (CRIM) plasmids, was originally developed by Haldimann and Wanner [2001], but has seen multiple rounds of development and improvement over the years in the Shearwin-Egan lab, and part of this development will be detailed in this section and the accompanying manuscript [St-Pierre et al., 2013]. The CRIM system achieves integration by employing bacteriophage integration machinery; the integrase for a particular bacteriophage attachment site (attB site) is expressed from a temperature-sensitive helper plasmid, which drives integration of a subsequently-transformed integration plasmid, which contains the DNA of interest and a phage attP site. The integration plasmid has a conditional replication origin (R6Kγ) and can only be propagated in *pir*$^+$ strains (such as EC100D), and therefore successful integrants are selected by the antibiotic resistance encoded by the integration plasmid. Excision is prevented by curing the helper plasmid through growth at the restrictive temperature and absence of the specific excisionase. Haldimann and Wanner [2001] developed different CRIM plasmid pairs encoding attP sites and integration machinery of a variety of bacteriophage and this allows serial integration of multiple DNA cassettes into the one *E. coli* strain (one integrant at each attB site for e.g. λ, Φ80, HK022, and P21).

The improvements made to the system in our lab can be divided into three main parts (1) initial work done by Ian Dodd developing the pIT3 system from the CRIM plasmids (where 'IT' stands

for integrating and terminating) by introducing more terminator sequences to transcriptionally insulate the DNA of interest from its chromosomal neighbourhood, (2) work done by Ian Dodd and myself to expand the pIT3 series to include multiple combinations of phage attachment sites and antibiotic resistance genes and (3) a large subsequent effort done mainly by Cui Lun in collaboration with François St-Pierre at Stanford University, culminating in the combination of the integration and helper functions onto one plasmid, yielding a system where chromosomal integration is as easy as plasmid transformation [St-Pierre et al., 2013].

4.2.1 pIT3 integration vector series

Ian Dodd developed the backbone for the pIT3 series of integration plasmids by combining together sequences from several different sources; (1) the R6Kγ origin and phage attachment site from the CRIM plasmids (pAH series) [Haldimann and Wanner, 2001], (2) a modular antibiotic resistance cassette from Lutz and Bujard [1997], (3) four *rrnB* terminators from pRS45 (from Rob Simons) and (4) a multiple cloning site (MCS) + RNAseIII site from pTL61T [Linn and St Pierre, 1990]. Cleavage of the reporter mRNA at the RNAseIII site removes any influence on the LacZ units arising from the 5‘ leader sequence, which differs for different promoters. Figure 4.1 shows a plasmid map of pIT3–CL, where ‘C’ indicates the antibiotic resistance (chloramphenicol) and ‘L’ the phage attachment site (lambda). Ian Dodd also cloned pIT3–TP (Tetracycline, ϕ80), pIT3–TO (186 att) and pIT3–TL. I completed the pIT3 series by generating the remaining 12 plasmids with four antibiotics (Spec, Tet, Chlor and Kan) and four attachment sites (lambda, ϕ80, HK022 and 186).

4.2.2 pIT3–CL–LacZtrim

It was decided that all reporter constructs for this study would be integrated at the λ*att* site, and Ian Dodd had cloned a vector (pIT3–CL–LacZ), where any DNA of interest can be cloned upstream of a *lacZ* reporter gene. This plasmid however was already 7.5 kb without any insert DNA and since integration efficiency seems to deteriorate when plasmid sizes reach ~ 9 kb, we sought to remove any superfluous DNA. Therefore a 1268 bp restriction fragment (*Psp*OMI/*Nde*I) containing unnecessary DNA downstream of the *lacZ* gene was replaced by a 287 bp fragment

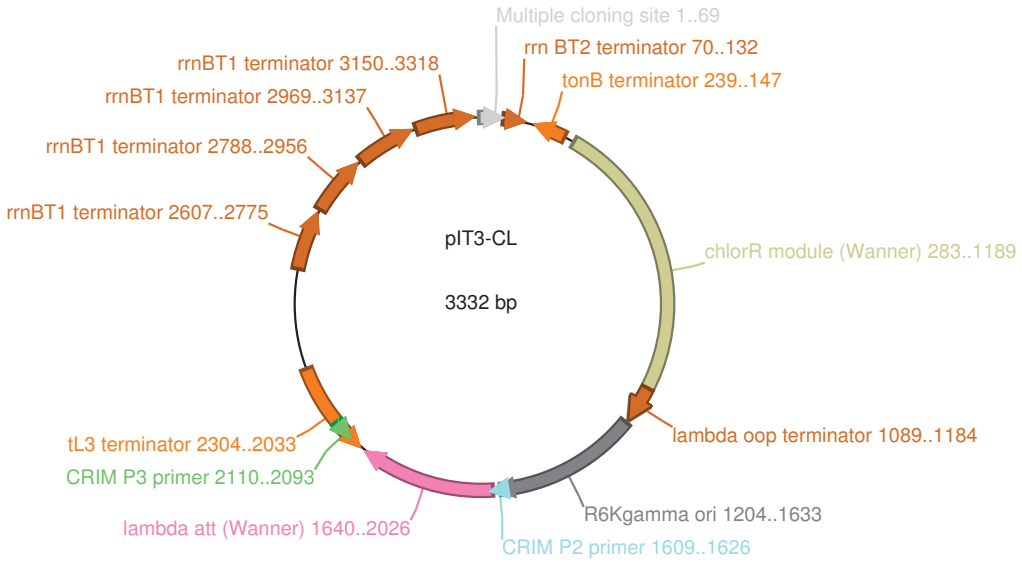


Figure 4.1: Plasmid map of pIT3-CL integration vector generated with ApE software.

(of synthetic origin), which reconstituted the 3' end of *lacZ*. The resulting plasmid (pIT3–CL–LacZtrim) served as the vector carrying DNA looping constructs into the *E. coli* chromosome. If this study were to be repeated, we may have decided to use the OSIP system [St-Pierre et al., 2013], however given that we were integrating a large number of constructs into relatively few strains, it was not much extra work to make batches of competent cells of each strain harbouring the pλINT helper plasmid.

4.2.3 Generating a reporter strain

We sought an *E. coli* reporter strain as close as possible to the wild-type K-12 strain, so we turned to MG1655 *rph*$^+$ (CGSC 7925). Since our main task was to test *lacZ* reporters, we needed to delete the *lac* operon from this strain to prevent interference from host-encoded *lacZ* and *lacI*. This was achieved through recombineering [Wang et al., 2009], using a 90 bp single-stranded oligonucleotide homologous to two regions either side of the *lac* operon, and a successful deletion was found by screening for white colonies on X-gal plates. The resulting strain, E4643 (MG1655 Δ*lac*IZYA, (Δ360,527–366,797)), served as the base strain into which all reporter constructs were subsequently integrated.

4.3 Manuscript: One-step cloning and chromosomal integration of DNA

The following manuscript describes the One Step Integration Plasmid (OSIP) system, the most recent development made to the CRIM system in the Shearwin lab, the primary development being combination of the integration helper and integration vector onto one plasmid. Other improvements include the introduction of a toxic gene to prevent propagation of the unwanted parental vector in cloning strains, as well as an efficient system employing FLP recombinase to remove the bulk of the integration plasmid (save the DNA of interest) after it is integrated into the host chromosome.

One-step cloning and chromosomal integration of DNA.

Francois St-Pierre[†,‡,☯], Lun Cui[§,☯], David G. Priest[§], Drew Endy[†], Ian B. Dodd[§], and Keith E. Shearwin[*,§]

[†]Department of Bioengineering, Stanford University, California, CA 94305, United States.

[‡]Department of Pediatrics, Stanford University, California, CA 94305, United States.

[§]School of Molecular and Biomedical Science (Biochemistry), University of Adelaide, Adelaide SA 5005, Australia.

[☯]These authors contributed equally to this work

ACS Synthetic Biology, 2013. 2 537-541. doi: 10.1021/sb400021j.

Statement of Authorship

Title of Paper	One-Step Cloning and Chromosomal Integration of DNA
Publication Status	◉ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., and Shearwin, K. E. (2013) One-Step Cloning and Chromosomal Integration of DNA. ACS Synthetic Biology DOI: 10.1021/sb400021j Publication Date (Web): May 6, 2013

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	LUN CUI		
Contribution to the Paper	Performed experiments, designed experiments, analyzed data and wrote the manuscript. (Co-first author)		
			11/1/13

Name of Co-Author	François St-Pierre		
Contribution to the Paper	Performed experiments, designed experiments, analyzed data and wrote the manuscript.		
Signature		Date	Sept22, 2013

Name of Co-Author	David G. Priest		
Contribution to the Paper	Performed experiments on early version of OSIP plasmids.		
Signature		Date	3/9/14

Name of Co-Author	Drew Endy		
Contribution to the Paper	Helped to evaluate and edit the manuscript.		
Signature		Date	18-SEP-2013

Statement of Authorship

Title of Paper	One-Step Cloning and Chromosomal Integration of DNA
Publication Status	◉ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., and Shearwin, K. E. (2013) One-Step Cloning and Chromosomal Integration of DNA. ACS Synthetic Biology DOI: 10.1021/sb400021j Publication Date (Web): May 6, 2013

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Co-Author	Ian B. Dodd		
Contribution to the Paper	Performed experiments on early version of OSIP plasmids, designed experiments and assisted in writing the manuscript.		
Signature		Date	23/12/2013

Name of Co-Author	Keith E. Shearwin		
Contribution to the Paper	Supervised the development of the work, designed experiments, analyzed data and wrote the manuscript.		
Signature		Date	23/12/13

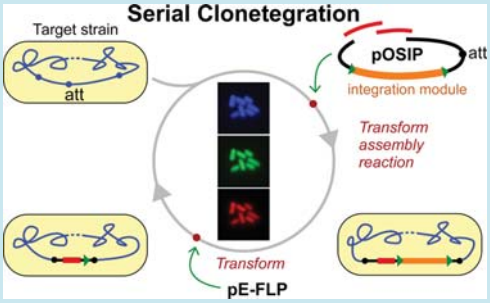
One-Step Cloning and Chromosomal Integration of DNA

François St-Pierre,[†,‡,⊥] Lun Cui,[§,⊥] David G. Priest,[§] Drew Endy,[†] Ian B. Dodd,[§] and Keith E. Shearwin*[,§]

[†]Department of Bioengineering and [‡]Department of Pediatrics, Stanford University, California 94305, United States
[§]Department of Biochemistry, School of Molecular and Biomedical Sciences, The University of Adelaide, SA 5005, Australia

Supporting Information

ABSTRACT: We describe "clonetegration", a method for integrating DNA into prokaryotic chromosomes that approaches the simplicity of cloning DNA within extrachromosomal vectors. Compared to existing techniques, clonetegration drastically decreases the time and effort needed for integration of single or multiple DNA fragments. Additionally, clonetegration facilitates cloning and expression of genetic elements that are impossible to propagate within typical multicopy plasmids.



KEYWORDS: *genome engineering, chromosomal integration, DNA assembly, genetic parts, genetic engineering, clonetegration*

Heterologous expression from a host cell chromosome, rather than plasmids, can reduce metabolic burdens and obviate the need for selectable markers in maintaining designer DNA sequences within an evolving bacterial population. Chromosomal integration of DNA is thus critical in synthetic biology, biotechnology, and metabolic engineering.[1,2] However, while existing techniques such as CRIM,[3] recombineering,[4−7] and Tn7-based integration[8,9] are useful and popular, they are also time-consuming. For example, CRIM requires multiple rounds of DNA transformation, while recombineering involves numerous steps and can take 1−2 weeks to complete.[4,6] We simplified the cloning of DNA sequences into prokaryotic chromosomes via three types of improvements to the well-known CRIM system, in which integration is mediated by bacteriophage integrases. The resulting method is quick and simple, enabling the bacterial chromosome to be used as a practical and powerful replacement to traditional plasmid vectors.

First, we developed new vectors that reduce the length and complexity of the integration protocol. In the CRIM system (Figure 1a, black trace), the target cell is initially transformed with a helper plasmid expressing a bacteriophage integrase (steps A1-2). A second transformation (step 4) introduces the CRIM plasmid containing the cloned insert and the "*attP*" site, a sequence of DNA necessary for site-specific recombination at the corresponding "*attB*" site on the bacterial chromosome. To test whether efficient integration could be achieved with a single transformation, we combined the integrase-expressing and the integration plasmids into a single vector (Figure 1b). The resulting hybrid vector, One-Step Integration Plasmid (pOSIP), integrates at high efficiency, thus bypassing two protocol steps typically requiring overnight incubation (Figure 1a, left-most green trace).

Restriction sites at key locations enable easy modification of the pOSIP backbone. We took advantage of this architecture to quickly construct five plasmid variants expressing a tyrosine integrase from either phage 186, HK022, lambda, phi80, or P21 or the serine integrase from phiC31. To our knowledge, our integration system is the first to use the efficient integrase from phage 186. pOSIP further contains two useful features not present in the CRIM plasmids (Figure 1b). First, forward and reverse transcription terminators flank the Multiple Cloning Site (MCS) to insulate integrated OSIP plasmids from transcription in chromosomal regions flanking the integration site and vice versa (Supplementary Figure 1). Second, a counter-selectable cassette containing the toxic *ccdB* gene in the middle of the MCS facilitates cloning by eliminating cells transformed only with the parental (unmodified) OSIP vector.[10]

Second, we sought to further simplify the integration protocol. Integration efficiency of most pOSIPs is sufficiently high to produce large numbers of integrants from pOSIP-insert cloning mixtures, instead of purified (miniprepped) plasmid DNA. This new one-step procedure, which we call "clonetegration" (simultaneous cloning and integration), effectively treats chromosomes as large cloning vectors. Like standard plasmid cloning, clonetegration requires the initial assembly of one or multiple DNA fragments into pOSIP using one's favorite cloning technique, such as Gibson assembly,[11] Clonetech In-Fusion(R),[12] or traditional restriction digest and ligation. The

Received: March 5, 2013
Published: May 6, 2013

dx.doi.org/10.1021/sb400021j | *ACS Synth. Biol.* 2013, 2, 537−541

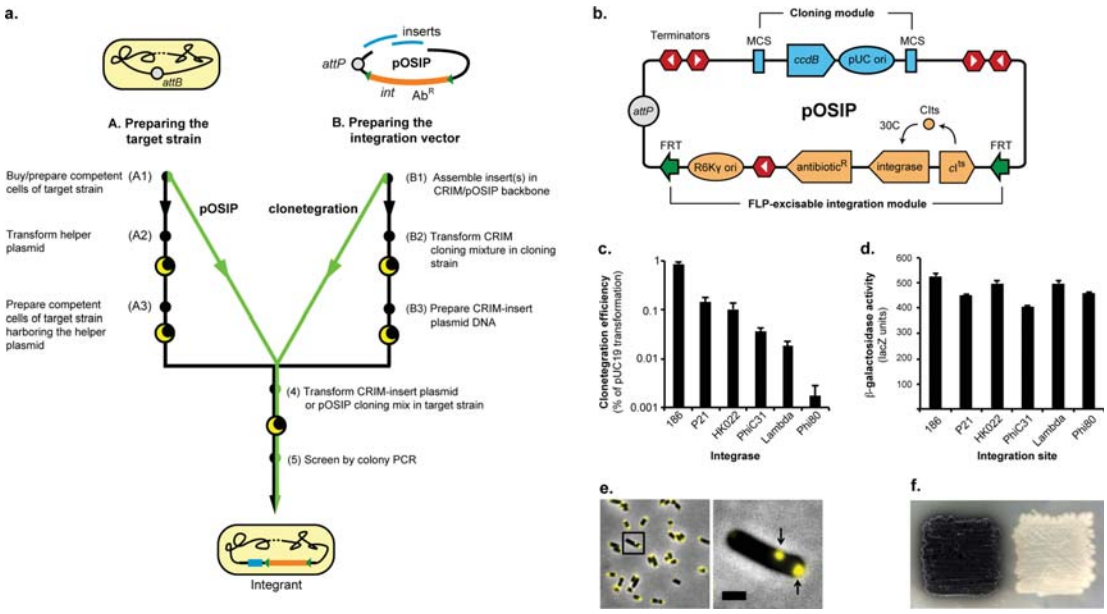


Figure 1. One step cloning and integration ("clonetegration"). (a) Clonetegration with pOSIP (green trace) requires fewer steps and is faster compared to chromosomal integration by CRIM (black trace). (b) pOSIP is composed of two functional modules. Cloning results in the replacement of the "insert module" with the DNA fragment to be integrated. The heat-inducible "integration module" allows expression of the integrase and selection of integrants by antibiotic resistance. The integration module can be removed postintegration by expressing FLP recombinase. (c) Clonetegration enables integration of a ∼4.6-kb lacZ cassette at high efficiency (mean ± SEM; n = 3 independent cultures per integrase/integration site). (d) Clonetegration of *lacZ* cassette into MG1655 Δ*lacIZYA* at different chromosomal loci results in qualitatively similar β-galactosidase activity (mean ± SEM; n = 4 independent cultures per integration site). (e,f) Sequences that cannot be cloned on medium-copy number plasmids can be assembled directly in the chromosome. Clonetegration of a *tsr-venus* expression cassette results in bright, membrane-localized fluorescent foci in NEB 5-alpha (e, arrows). Scale bar = 1 μm. Clonetegration of a *vioE-ABDE* expression cassette results in dark pigmentation (f, left), seen here on bacterial cells streaked on a square area on an LB-plate; in comparison, the parental strain produces a beige color (f, right).

resulting cloning mixtures are then directly transformed into chemically or electro-competent cells (Figure 1a, right-most green trace). Integration of properly assembled pOSIP-insert molecules occurs during post-transformation outgrowth. Verification of integration can be performed by colony PCR using sets of primers we optimized from the original CRIM set (Supplementary Tables 1, 2; Supplementary Figure 1).

We quantified the efficiency of clonetegration using a 4.6kb lacZ-expression cassette assembled into each of our six pOSIP plasmids (Figure 1c). All trials gave *lacZ* positive integrants in MG1655 Δ*lacIZYA* (Supplementary Figure 2). Expression from the integrated constructs is similar across all integration sites, suggesting that surrounding chromosomal sequences had minimal influence on our terminator protected insert (Figure 1d). The phage 186 integrase performed best, with a clonetegration efficiency ∼500-fold better than that of the poorest-performing integrase, Phi80. Phage 186 integrase might therefore be suggested for the most challenging sequences targeted for chromosomal integration. Of note, we observed that reactions catalyzed by phage 186 integrase result in integration not only within the *tRNIleY* gene[13] but sometimes into an alternative site within the *tRNIleX* gene (39% probability; n = 72 colonies; Supplementary Figure 3). No colonies harbored integrants at both chromosomal loci. We designed new PCR validation primers for users needing to distinguish between these two integration locations (Supplementary Tables 1, 2). Finally, we successfully performed clonetegration in *Salmonella typhimurium*, a bacterium receiving renewed interested in synthetic biology,[14] demonstrating that clonetegration is generalizable to other prokaryotes (Supplementary Figure 4).

In the CRIM protocol, one must first clone target sequences in a multicopy plasmid prior to integration (Figure 1a, steps B1−3). By removing this plasmid intermediate step, clonetegration can also enable integration of sequences that are lethal when present at multiple copies within a cell. We encountered such toxic sequences ourselves while working on separate research projects. In one such endeavor, we wanted to use the protein fusion Tsr-Venus as a sensitive, single-molecule fluorescent reporter of gene expression.[15] However, we were unable to clone a *tsr-venus* expression cassette in pOSIP or in the popular pBR322 vector (Methods in Supporting Information). In contrast, clonetegration of our toxic Tsr-Venus expression cassette into NEB 5-alpha was successful, resulting in cells with the expected pattern of yellow fluorescence foci at polar and midcell membrane positions[15] (Figure 1e). As a second example, we were exploring the construction of colorimetric reporters using a large (∼6 kb) four-gene dark green pigment-producing cassette (*ABDE*) from the *vioE* operon. While we were unable to clone a *vioE-ABDE* expression cassette in pOSIP using standard methods,

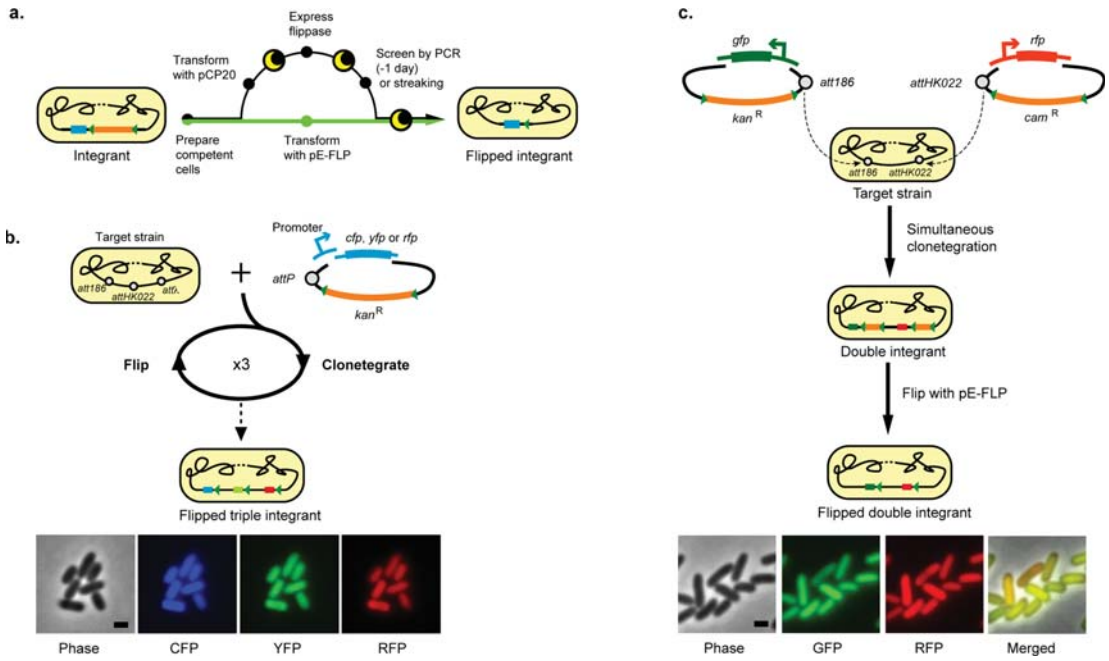


Figure 2. Serial and parallel clonetegration. (a) The pE-FLP plasmid enables simple one-step excision of the integration module from pOSIP integrants (green trace). In contrast, excision with pCP20 requires multiple steps performed over 2−3 days (black trace). (b) Clonetegration is scalable to multiple integration sites. Three fluorescent protein (FP) expression cassettes were clonetegrated in successive rounds into NEB 5-alpha. The integration module and kanamycin resistance marker, flanked by FRT sites, were excised using pE-FLP after each round of clonetegration. The resulting bacterial strain expresses mTurquoise CFP (blue), Venus YFP (green), and mRuby2 RFP (red) from phages 186, HK022, and lambda integration sites, respectively. A phase contrast image is shown for reference. Scale bar = 1 μm. (c) Clonetegration can be performed concurrently with two independent DNA fragments. A Clover GFP (green) expression cassette was clonetegrated into phage 186 integration site in NEB 5-alpha using a pOSIP encoding a kanamycin resistance marker (kan^R). In parallel, an mRuby2 RFP (red) expression cassette was clonetegrated into phage HK022 integration site with a chloramphenicol resistance marker (cam^R). The integration modules and antibiotic resistance markers of both pOSIP vectors, flanked by FRT sites, were removed in a single step using pE-FLP. A phase contrast image and a merged image from all channels are shown for reference. Scale bar = 1 μm.

clonetegration into NEB 5-alpha successfully produced dark green colonies (Figure 1f).

Third, we aimed to improve the ease of constructing strains with multiple integrated sequences. Like CRIM, our system results in integration of the entire pOSIP plasmid. The ability to remove the integration module, including the antibiotic marker (e.g., kan^R), would make our system scalable by allowing integration of additional kanamycin-resistant pOSIPs into other chromosomal loci. To enable construction of such marker-less strains,[16] we engineered all pOSIPs with FRT sites flanking the integration module (Figure 1b), allowing excision of these sequences following expression of the FLP recombinase.[17]

To test FLP-mediated excision, we transformed chromosomal integrants with the popular FLP-expressing plasmid pCP20.[18] We followed the standard pCP20 protocol (Figure 2a, black trace). First, we transformed the target strain with pCP20 and grew cells overnight at 30 °C on ampicillin plates. Next, we restreaked individual colonies and induced FLP expression by overnight growth at 37 °C; like our integrase expression cassette, FLP expression in pCP20 is under control of the thermosensitive transcription factor lambda CIts. Growth is performed on LB without antibiotics because pCP20, which contains a temperature-sensitive replicon, is not propagated at 37 °C. Lastly, we screened the colonies on the resulting LB plates for FLP-mediated excision by PCR and by streaking for loss of antibiotic resistance.

The standard protocol produced excision of pOSIP integration modules in less than 5% of pCP20-transformed cells (Supplementary Figure 5); CIts-controlled modules, present on both pOSIP and pCP20, may be interfering with one another. We thus developed a new FLP expression system that relies neither on temperature induction nor lambda CIts. Because the pCP20 protocol is time-consuming, we also sought to design a new procedure so that excision can be performed in a single day, rather than the 2 or 3 days required when using pCP20.

We achieved both goals by driving FLP expression via pE, a strong constitutive promoter from phage P2[19] (Supplementary Figures 5 and 6; Supplementary Table 3). In contrast to pCP20, pE-FLP successfully catalyzed excision of the integration and propagation modules from integrants with 100% efficiency. Also, pE-FLP is active immediately after transformation, thus avoiding the overnight heat-induction step of pCP20. Given the high efficiency of pE-FLP, screening colonies for excision of the antibiotic marker and neighboring sequences is typically not necessary. Because pE-FLP retains the temperature-sensitive replicon of pCP20, blocking postflip-

dx.doi.org/10.1021/sb400021j | *ACS Synth. Biol.* 2013, 2, 537−541

ping propagation of the plasmid is achieved by simply growing integrants at 37 °C. In short, pE-FLP reduced the excision protocol to a mere transformation step.

We confirmed that our overall protocol, including pE-FLP-mediated vector backbone excision, does indeed enable scalable clonetegration: we successfully integrated expression cassettes for cyan, yellow, and red fluorescent proteins at separate chromosomal loci via three separate rounds of clonetegration (Figure 2b, Supplementary Figure 7). Each time, integration was mediated by a pOSIP vector encoding a kanamycin resistance marker. As needed to allow reuse of the kanamycin resistance marker, we excised the integration module from integrants via flippase expression from pE-FLP. Similarly, we serially clonetegrated four 4.6-kb lacZ expression cassettes, giving 18.4 kb of integrated sequences (Supplementary Figure 8).

In the protocols described above, each expression cassette must be integrated via a separate round of clonetegration. We sought to increase the rate at which multicassette integrants could be constructed. Since each OSIP plasmid expresses its own integrase, cotransformation should allow integration at multiple chromosomal sites simultaneously. Indeed, dual integration of two fluorescent-protein-expressing cassettes was successful in standard commercially available chemically competent cells, without requiring further optimization of the standard protocol (Figure 2c, Supplementary Figure 7b).

While all clonetegration experiments described here were successful, further improvements to clonetegration efficiency could be useful for certain applications such as library construction or integration of very large sequences. Such improvements might be obtained by increasing the expression level of our pOSIP integrases, screening new phage integrases for higher integration efficiency, or transforming DNA using electroporation. As with all cloning procedures, best laboratory practice is to sequence the region of interest from the final clonetegrated strain to ensure that no errors have been introduced.

In part because of the laborious nature of current procedures to integrate DNA fragments into prokaryotic chromosomes, plasmids remain the most popular expression vectors. The technique we describe here, clonetegration using pOSIP, is simple and rapid, can facilitate cloning of toxic sequences, and is amenable to automation. We have integrated up to four expression cassettes at independent chromosomal loci in successive rounds and two cassettes in the same round. We anticipate that clonetegration with pOSIP will become a valuable technique facilitating genetic engineering with difficult-to-clone sequences and rapid construction of synthetic biological systems.

■ ASSOCIATED CONTENT

Supporting Information

Complete Methods, Supplementary Tables 1 to 3, Supplementary Figures 1 to 8 and Supporting References. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +61 08 83135361. E-mail: keith.shearwin@adelaide.edu.au.

Author Contributions

⊥These authors contributed equally to this work. F.S.-P., L.C., I.B.D., and K.E.S. designed research; F.S.-P., L.C., D.G.P., I.B.D., and K.E.S. performed research; F.S.-P., L.C., I.B.D., D.E., and K.E.S. analyzed data; and F.S.-P., L.C., I.B.D., D.E., and K.E.S. wrote the paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the China Scholarship Council (L.C.), the U.S. NSF Synthetic Biology Engineering Research Center (SynBERC; D.E.), Human Frontiers Science Program grant number RGP051 (K.E.S.), Australian Research Council grant numbers DP110100824 and DP11010470 (K.E.S, I.B.D.), and a William H Elliott Biochemistry Fellowship (I.B.D.).

■ REFERENCES

(1) Tyo, K. E., Alper, H. S., and Stephanopoulos, G. N. (2007) Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol. 25*, 132−137.

(2) Keasling, J. D. (2012) Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng. 14*, 189−195.

(3) Haldimann, A., and Wanner, B. L. (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol. 183*, 6384−6393.

(4) Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., and Court, D. L. (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc. 4*, 206−223.

(5) Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A. 97*, 5978−5983.

(6) Kuhlman, T. E., and Cox, E. C. (2010) Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res. 38*, e92.

(7) Zhang, Y., Buchholz, F., Muyrers, J. P., and Stewart, A. F. (1998) A new logic for DNA engineering using recombination in Escherichia coli. *Nat. Genet. 20*, 123−128.

(8) McKenzie, G. J., and Craig, N. L. (2006) Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial chromosomes using Tn7 without need for selection of the insertion event. *BMC Microbiol. 6*, 39.

(9) Sibley, M. H., and Raleigh, E. A. (2012) A versatile element for gene addition in bacterial chromosomes. *Nucleic Acids Res. 40*, e19.

(10) Shetty, R. P., Endy, D., and Knight, T. F., Jr. (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng. 2*, 5.

(11) Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343−345.

(12) Irwin, C. R., Farmer, A., Willer, D. O., and Evans, D. H. (2012) In-fusion(R) cloning with vaccinia virus DNA polymerase. *Methods Mol. Biol. 890*, 23−35.

(13) Reed, M. R., Shearwin, K. E., Pell, L. M., and Egan, J. B. (1997) The dual role of Apl in prophage induction of coliphage 186. *Mol. Microbiol. 23*, 669−681.

(14) Prindle, A., Selimkhanov, J., Danino, T., Samayoa, P., Goldberg, A., Bhatia, S. N., and Hasty, J. (2012) Genetic Circuits in Salmonella typhimurium. *ACS Synth. Biol. 1*, 458−464.

(15) Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X. S. (2006) Probing gene expression in live cells, one protein molecule at a time. *Science 311*, 1600−1603.

(16) Minaeva, N. I., Gak, E. R., Zimenkov, D. V., Skorokhodova, A. Y., Biryukova, I. V., and Mashko, S. V. (2008) Dual-In/Out strategy for genes integration into bacterial chromosome: a novel approach to

step-by-step construction of plasmid-less marker-less recombinant E. coli strains with predesigned genome structure. *BMC Biotechnol. 8*, 63.
(17) Turan, S., and Bode, J. (2011) Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications. *FASEB J. 25*, 4088−4107.
(18) Cherepanov, P. P., and Wackernagel, W. (1995) Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene 158*, 9−14.
(19) Callen, B. P., Shearwin, K. E., and Egan, J. B. (2004) Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol. Cell 14*, 647−656.

dx.doi.org/10.1021/sb400021j | *ACS Synth. Biol.* 2013, 2, 537−541

Supporting Information for *One-step cloning and chromosomal integration of DNA*.

François St-Pierre, Lun Cui, David G Priest, Drew Endy, Ian B Dodd, and Keith E Shearwin.

METHODS

OSIP plasmid construction

OSIP plasmids were constructed using standard cloning techniques. We amplified tL3, *attP*, R6Kγ and antibiotic resistance markers from CRIM plasmids (*1*); the integration module from CRIM helper plasmids (*1*); *ccdB* and pUC ori from BBa_I52002 (*2*); the *rrnB*T2 terminators from pTL61T(*3*) and the *tonB* and *rrnB*T2 terminators from *E. coli*. FRT sites were encoded in assembly primers using the sequence from the Keio collection (*4*).

The pUC origin of replication, located within the multiple cloning site, allows the parental plasmid to replicate at high copy number in strain DB3.1 (Invitrogen) and hence provide ample starting material for cloning. The *pir*-dependent R6Kγ origin is unnecessary for clonetegration, but allows the user to maintain a plasmid version of the pOSIP clone in a *pir* -containing strain.

Sequences of several pOSIP variants are available in GenBank (Accession numbers KF030457 to KF030467). These are named according to the combination of antibiotic resistance marker and *attP* site that they contain. For example, pOSIP-KL carries the kanamycin resistance marker (K) and the lambda (L) *attP* site. Chloramphenicol and tetracycline resistance markers are denoted by (C) and (T), respectively, while the one letter codes for each *att* site are given in Supplementary Table 1.

We constructed pE-FLP by amplifying and assembling three sequences: the pE promoter from phage P2 (*5*), the flippase gene from pCP20 (*6*), and the pKD46 backbone containing the temperature sensitive origin of replication and ampicillin resistance marker (*7*). The pE-FLP sequence is available from GenBank (Accession number KF030456).

Bacterial strains

The base pOSIP vectors, containing a pUC origin of replication (pUC*ori)* and the cytotoxic gene *ccdB*, were propagated using the *ccdB*-resistant strain DB3.1 (Invitrogen). Removal of the *ccdB*-pUC*ori* cassette during cloning results in vectors with the pi-protein conditional origin of replication R6Kγ, which is inactive in most cloning strains. When propagation as multi-copy vectors was required, the resulting plasmids were transformed in the pi-expressing strain TransforMax™ EC100D™ *pir*+ (Epicentre). Standard cloning strains such as DH5α-Z1 were used for propagation of all other plasmids.

Integrations into *E. coli* chromosomal *attB* sites were performed with either NEB 5-alpha (C2987) or MG1655 Δ*lacIZYA*. MG1655 Δ*lacIZYA*, used as the target strain for integration of *lacZ* cassettes, was constructed by using recombineering (*8*) to precisely delete the entire lac operon from MG1655 (*rph*+) (CGSC 7925). Integration with the φC31 integrase was performed on strain MG1655 Δ*lac* lambda *attB*::φC31. To construct this strain, the φC31 *attB* sequence, not normally present in the *E.coli* chromosome, was clonetegrated using a kanamycin-resistant pOSIP into the lambda *attB* site, followed by excision of the integration module. *Salmonella typhimurium* LT2 (LB5010), a *galE*⁻, rough LPS, virulence plasmid negative strain, was a gift from Dr Renato Morona (University of Adelaide).

Media, antibiotics and buffers

E. coli was typically grown in Lysogeny broth (LB) (*per Liter:* 10 g Tryptone, 5 g Yeast Extract, 5 g NaCl) at 30 or 37ºC with rotary shaking at 200 rpm. Recovery after

transformation was performed in SOC (*per Liter:* 20 g Tryptone, 5 g Yeast Extract, 0.5 g NaCl, 0.186 g KCl, 0.952 g $MgCl_2$, 50 mM glucose). For β-galactosidase assays, cells were grown in M9 minimal medium supplemented with glycerol (1xM9 salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% Glycerol, 0.01 mM $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$ in sterile water).
Standard antibiotic concentrations were used for plasmid propagation (*9*). For the low-copy plasmid pE-FLP, we generally used 50 μg/ml of ampicillin, as higher concentrations severely decreased growth rate. For selection of chromosomal integrants, we used the appropriate antibiotic(s), kanamycin and/or chloramphenicol, at a concentration of 15 μg/ml.

Preparing and testing competent cells
We obtained DH5α chemically-competent cells from a commercial source (New England Biolabs, NEB 5-alpha, C2987). We prepared chemically-competent cells of all other strains using the TSS method (*10*), and electrocompetent cells by standard methods (*9*).

We measured the transformation efficiency of competent cells using 50 pg (1 μl at 50 pg/μl) of pUC19 plasmid (NEB N3041). We transformed TSS chemically-competent cells as described in the clonetegration section or as suggested by the manufacturer (NEB 5-alpha). For electro-competent cells, we gently mixed pre-chilled DNA with 20 μl of cells thawed on ice, then transferred the mixture to pre-chilled 1 mm electroporation cuvettes. We transformed cells using a Bio-Rad MicroPulser, with the 'E.coli' pre-set Ec1 program (1.8 kV). SOC medium was added immediately after electroporation, and cells were incubated at 37ºC with rotary shaking at 200 rpm for 1 hour. For both electro- and chemically-competent cells, we plated transformed cells on LB agar supplemented with 100 μg/ml of ampicillin, and incubated overnight at 37ºC.

Clonetegration
An overview of the clonetegration procedure is depicted in Figure 1a of the main text.

LacZ cassette
In our *lacZ* cassette, the promoter from the ampicillin resistance gene (pBla) drives expression of the full-length β-*galactosidase* gene. pBla-*lacZ* inserts were amplified by PCR from pLacatt1-delLacY-pBla-lacZ (*11*) by PCR using KAPA HiFi (Kapa Biosystems) polymerase according to the manufacturer's instructions. pOSIP vector backbones were digested by restriction endonucleases. Digestion and PCR products were gel purified (Zymoclean Gel DNA recovery kit, Zymo Research). The pBla-*lacZ* fragment and the digested pOSIP backbone were assembled using Gibson Isothermal Assembly (*12*). Assembly reactions (20 μl) contained ~60 ng of pOSIP backbone with ~50 ng of pBla-*lacZ*, resulting in approximately equimolar concentration of backbone and insert. Overlaps between insert and backbone were ~40 bp.

Aliquots (5 μl) of unpurified assembly reaction were mixed gently with 0.1 ml TSS chemically-competent cells, incubated on ice for 30min in a pre-cooled Eppendorf tube, heat-shocked for 45 seconds at 42ºC, placed on ice for a further 2 min, and supplemented with 0.9 ml of pre-warmed SOC. Cells were incubated at 37ºC for one hour, spread on LB agar plates with the appropriate antibiotics plus X-gal, and incubated at 30ºC overnight. Clonetegration efficiency was expressed as the percentage of the pUC19 transformation efficiency, normalised by the amount of DNA transformed.

Parallel clonetegration of *lacZ* cassettes were performed as with single clonetegration experiments, except that electrocompetent MG1655 Δ*lacIZYA* were used as target cells. We

selected integrants on LB plates containing 15 µg/ml kanamycin and 15 µg/ml chloramphenicol.

Fluorescent proteins & color pigments
Templates for experiments with fluorescent proteins and color pigments were obtained by gene synthesis or, when indicated by the prefix "BBa_", via the Registry of Standard Biological Parts (*2*).

For experiments with the *vioE* operon, we PCR-amplified *vioE-ABDE* from BBa_K274003, and the strong promoter BBa_R0040 from BBa_J13002. Using In-Fusion cloning (*16*), we assembled these inserts in a pOSIP variant expressing phage HK022 integrase. Clonetegration experiments using this cloning mixture were successful (Fig. 1f). In contrast, transformation of the same cloning mixtures so as to allow plasmid propagation without integration (EC100D*pir*+ strain, 30ºC outgrowth) did not produce pigment-producing colonies. As expected, transformation of an amp^R control plasmid (pUC19) in the same batch of EC100D*pir*+ cells produced ampicillin-resistant colonies.

For experiments with Tsr-Venus, Blue Heron Bio (USA) synthesized the DNA sequence for *tsr-venus* (*13*). Blue Heron Bio attempted to clone *tsr-venus* into the medium-copy plasmid pBR322 (New England Biolabs N3033) downstream of three candidate promoter-RBS fragments: (1) the strong promoter BBa_J23101 followed by a strong 25 bp Ribosome Binding Site (RBS) "TAACTTTAAGAAGGAGCCCTTCACC" extracted from plasmid pVS116Tsr (*13*), (2) BBa_J23101 as above, but immediately followed by the weaker RBS BBa_B0032 and a SpeI site, (3) a strong promoter/RBS from phage lambda pR (lambda gene coordinates 37,941-38,040) followed by a SpeI site. Only cloning with the weaker RBS from (2) was successful, but produced only minimal fluorescence under the microscope. We then assembled the strong promoter-RBS BBa_J13002 upstream of *tsr-venus* in a pOSIP variant expressing phage P21 integrase. Clonetegration experiments following either In-Fusion cloning or traditional restriction digest/ligation cloning were successful (Fig. 1e). In contrast, transformation of the same cloning mixtures so as to allow plasmid propagation without integration (EC100D*pir*+ strain, 30ºC outgrowth) did not produce fluorescent cells. As expected, transformation of an amp^R control plasmid (pUC19) in the same batch of EC100D*pir*+ cells produced ampicillin-resistant colonies.

For all other experiments with fluorescent proteins (FPs), gene expression was driven by a strong 115 bp constitutive promoter-RBS sequence extracted from an in-house vector (pNCS), followed by a 99-111 bp leader sequence containing a 6xHis tag. For experiments with serial clonetegration of FP cassettes, these promoter fragments were amplified from pNCS. Venus YFP (*14*) was amplified from *tsr-venus* (see above), while the genes encoding mTurquoise CFP (*15*) and mRuby2 RFP (*16*) were amplified from in-house vectors. For parallel clonetegration experiments, single promoter-FP fragments were amplified from vectors containing Clover GFP or mRuby2 RFP (*16*) downstream of the *p*NCS promoter-RBS sequence.

All DNA fragments were PCR-amplified with PrimeStar HS (Takara) according to the manufacturer's instructions. pOSIP was digested with EcoRI/PstI (kanamycin resistant variants), or BamHI/PstI (chloramphenicol resistant variants). Digestion and PCR products were gel purified (Zymoclean Gel DNA recovery kit, Zymo Research). We assembled fluorescent cassettes in 2 µl (total volume) reactions using In-Fusion assembly (*17*) with inserts at 1:1 molecular ratio to 20 ng of vector backbone. We transformed and selected

integrants as specified above, except that we used 0.5 µl of assembly reaction to 10 µl (commercially-prepared) or 100 µl (self-prepared) chemically competent cells.

For parallel clonetegration of FP cassettes, we assembled a Clover GFP expression cassette in a kanamycin-resistance *att186* pOSIP variant and an mRuby2 RFP expression cassette in a chloramphenicol-resistance *attHK022* pOSIP variant using In-Fusion assembly as described above. We transformed 0.5 µl of each assembly reaction in 20 µl of NEB DH5α commercial chemically-competent cells. We selected integrants on LB plates containing 15 µg/ml kanamycin and 15 µg/ml chloramphenicol.

PCR verification of integration

We used colony PCR to confirm integration of pOSIP at single copy into the bacterial chromosome. Isolated colonies were picked and resuspended in a 20 µl PCR reaction volume containing 0.5U KAPA 2G Robust Polymerase (Kapa Biosystems), 1 µl of each primer (P1, P2, P3 and P4; Supplementary Table 3), 2 µl 2 mM dNTPs and 4 µl 5X KAPA 2G buffer B (Kapa Biosystems). We followed the following colony PCR program: initial denaturation for 3 minutes at 95ºC; 30 cycles of amplification (denaturation at 95ºC for 20 seconds, primer annealing at 50ºC – or 55ºC for P21 primers – for 20 seconds, extension at 72ºC for 40 seconds); and a final extension at 72ºC for 4 minutes.

Excision of the integration module from pOSIP

An overview of the excision procedure is depicted in Figure 2a of the main text.

We transformed integrants with >50 ng of pE-FLP and selected for successful transformation by growth on ampicillin (50µg/ml) plates at 30ºC for one day. We confirmed the loss of the pOSIP integration module in ampicillin-resistant colonies by colony PCR with primers P1 and FLIP_F (Supplementary Tables 1 and 3), or by streaking on LB agar with and without the appropriate antibiotics, followed by overnight growth at 30ºC. Because all colonies consistently excised their pOSIP integration module, this verification step is typically unnecessary. When excising pOSIP integrants containing a pBla-*lacZ* cassette, we supplemented all plates with 3% X-gal (BioVectra) to confirm continued expression of β-galactosidase.

For initial tests with pCP20 instead of pE-FLP, we followed a similar procedure as with pE-FLP. However, a separate FLP-induction step was necessary (Fig. 2a): we induced FLP expression by streaking pCP20 transformants on LB agar plates without antibiotics, followed by overnight incubation at 37ºC. Resulting colonies were tested for successful excision as described above.

When testing for proper function of integrated fragments is difficult or laborious, we recommend sequencing the region of interest from the final clonetegrated strain to ensure no errors have been introduced.

Minimal medium LacZ Assay

Our modification (*18*) to the original Miller assay uses a microtitre plate format. Fresh colonies were inoculated into wells of a flat-bottom 96-well microtitre plate (Corning Costar 3599) containing 100 µl M9 glycerol minimal medium (and appropriate antibiotics) per well. The plate was sealed to prevent evaporation, and incubated for ~16 hours with shaking (200 rpm) at 37ºC.

Aliquots from each well of the overnight microtitre plate were then diluted 33-fold into 100 µl M9 glycerol in a fresh 96-well plate, and growth at 37ºC was continued, with shaking (450 rpm), to an OD_{600}~0.6-0.8 Multiskan Ascent microtitre plate reader (Thermo Fisher Scientific, Australia). Aliquots of culture from each well were then added to a combined lysis-assay buffer, with each well of the microtitre plate containing: 50 µL culture + M9 glycerol minimal medium (usually 20 µL culture + 30 µL M9 glycerol minimal medium), 150 µL TZ8, 40 µL ONPG (o-nitrophenyl-b-D-galactoside 4 mg/mL, Biovectra, in TZ8), 1.9 µL 2-mercapoethanol, and 0.95 µL polymyxin B (20 mg/mL, Sigma). TZ8 buffer is 100 mM Tris-HCl pH 8.0, 10 mM KCl and 1 mM $MgSO_4$. The increase in absorbance at 414 nm (A_{414}) was followed at 28ºC in a Multiskan Ascent microtitre plate reader. A_{414} was read every 2 minutes for 1 hour with 10 seconds shaking (960 rpm) immediately preceding each reading. The change in A_{414} per minute was determined and β-galactosidase units were calculated as follows:

$$\beta\text{-galactosidase units} = 200{,}000 \times (A_{414}/\text{min})/(OD_{600} \times \text{culture volume in } \mu l)$$

Fluorescence Microscopy

Cells were imaged using a metal-halide light source (EXFO X-Cite 120PC), a cooled CCD camera (Hamamatsu Orca-ER) and a 100x phase-contrast oil-immersion objective (Olympus Fluar 100x 1.3-NA Ph3) on an inverted microscope (Zeiss Axiovert 200M). Fluorescent images were acquired with the following filters (ex = excitation, em = emission): ex:440/20 nm and em:480/40 nm for mTurquoise CFP, ex:485/30 nm and em:530/40 nm for Clover GFP or mVenus YFP, and ex:545/20 nm and em:590/80 nm for mRuby2 RFP.

Integration site	P1 primer	P2 primer	P3 primer	P4 primer
HK022 (H)	GGAATCAATGC CTGAGTG	ACTTAACGGCTGA CATGG	ACGAGTATCGAGA TGGCA	GGCATCAACAGCA CATTC
Lambda (L)	GGCATCACGGC AATATAC	ACTTAACGGCTGA CATGG	GGGAATTAATTCT TGAAGACG	TCTGGTCTGGTAG CAATG
Φ80 (P)	GATTTGAGCGA GCAACTGTACC	GTTCGCAGAGTGT TATGGTT	AAAGAAACAGAG AAGGGCAC	TGGCCTTAACAAA GACATCA
P21 (T)	ATCGCCTGTAT GAACCTG	ACTTAACGGCTGA CATGG	GGGAATTAATTCT TGAAGACG	TAGAACTACCACC TGACC
φC31 (C)	GGCATCACGGC AATATAC	ACTTAACGGCTGA CATGG	ACGAGTATCGAGA TGGCA	ATACCGGAATTCG GATCCTCTAGACA TGCCCGCCGTGAC CGTCG
186 (O) (*E .coli*, primary)	CTCATTCGAAA CCACCCACCG	ACTTAACGGCTGA CATGG	ACGAGTATCGAGA TGGCA	GATCATCATGTTT ATTGCGTGG
186 (*E. coli*, secondary)	TCCGGAATGCC TGCATTG	ACTTAACGGCTGA CATGG	ACGAGTATCGAGA TGGCA	CCCTGGAGCCAAA ATATCC
186 (*S. typhimurium*)	TCTGGAATCGC TTAGCCG	ACTTAACGGCTGA CATGG	ACGAGTATCGAGA TGGCA	CTGGCGCCAGTAT GTCCT

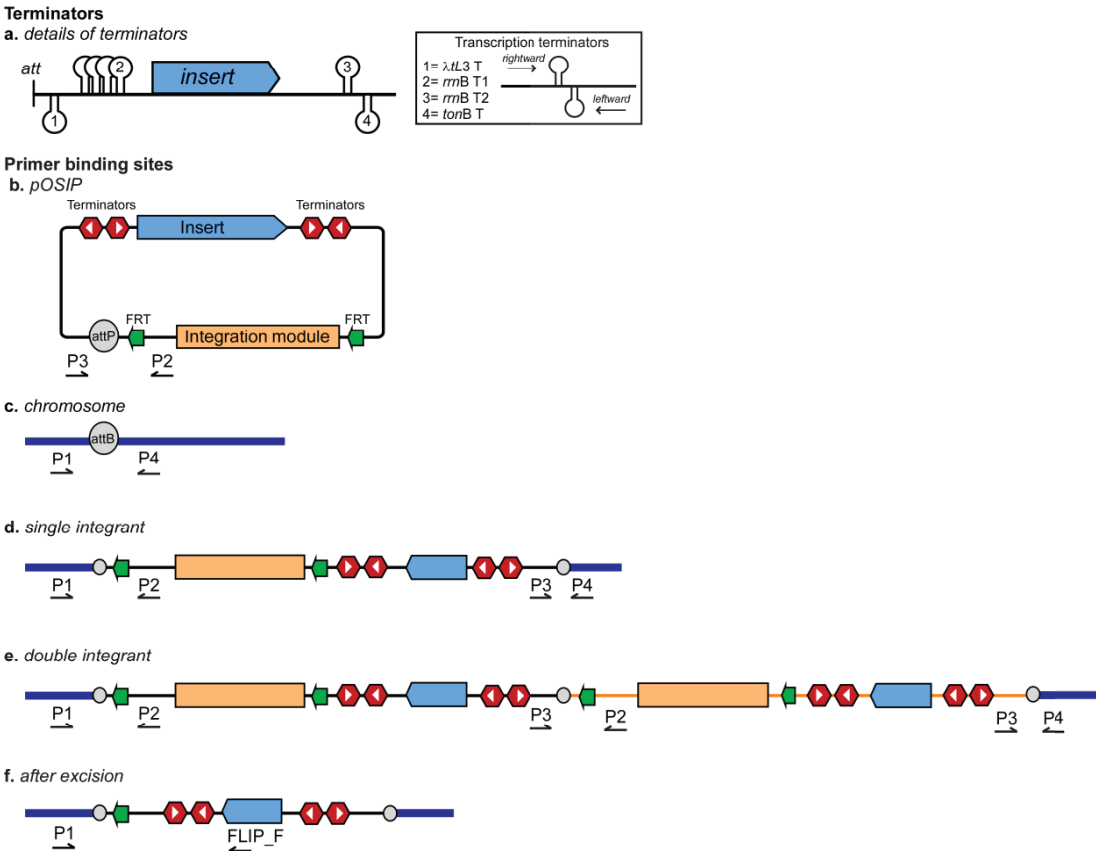
Supplementary Table 1: Primer sequences for colony PCR verification of integration. For each integration site, four primers (P1-P4; see Supplementary Figure 1) were combined in a single colony PCR reaction. All primer sets are for integration into *Escherichia coli* MG1655, unless otherwise indicated. Primers for HK022, Lambda, Φ80, P21 and φC31 are from Haldimann and Wanner (*1*), except for the underlined primers, which were designed to better distinguish between single integration, multiple integration and no integration events (Supplementary Table 2). Verification primers for integration of pOSIP variants expressing 186 or φC31 integrases were also designed for the current work. *Salmonella typhimurium* contains a single naturally occurring 186 *attB* site, equivalent to the secondary site of *E. coli*. However, the chromosomal sequence surrounding the *Salmonella attB* site differs from the *E. coli sequence*, thus requiring a different pair of P1 and P4 primers. The letter in parentheses following each *att* site indicates the abbreviation used for systematic naming of the pOSIP variants.

Integration site	No integration	Single integration	Additional band from multiple tandem integration
HK022	740	343, 824	427
Lambda	741	631, 1156	1046
Φ80	569	303, 499	233
P21	506	622, 1110	1226
φC31	1150	446, 1093	389
186 (*E. coli*, primary)	241	328, 389	476
186 (*E. coli*, secondary)	601	429, 648	476
186 (*S. typhimurium*)	601	429, 648	476

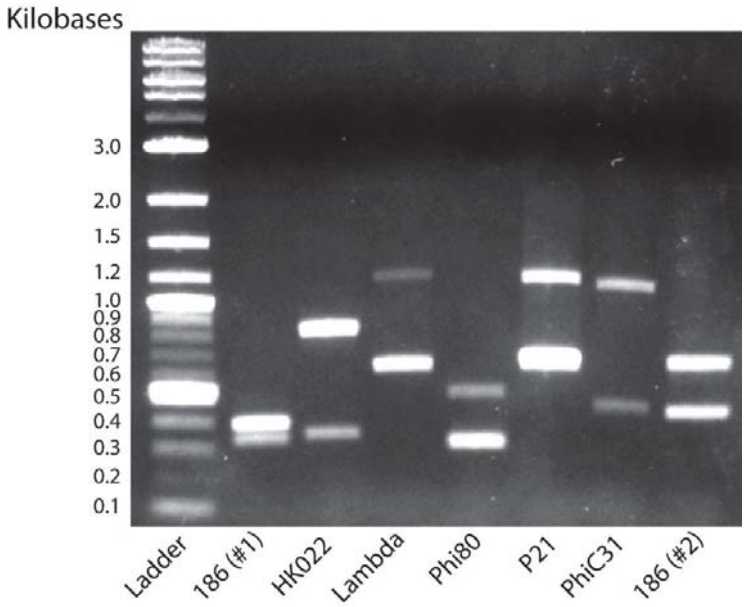
Supplementary Table 2: Predicted sizes of PCR fragments following colony PCR verification of integration. As shown in Supplementary Figure 1, absence of integration produces a single fragment from primers P1 and P4. Single integration produces two fragments, one from P1 and P2, and one from P3 and P4. In multiple integration events, where multiple pOSIP plasmids integrate in tandem, primers P2 and P3 produces an additional fragment. Sizes are listed in base pairs (bp).

Integration site	**Predicted PCR product size (bp)**
HK022	793
Lambda	1081
Φ80	943
P21	1072
φC31	1081
186 (*E. coli*, primary)	839
186 (*E. coli*, secondary)	1098
186 (*S. typhimurium*)	1098

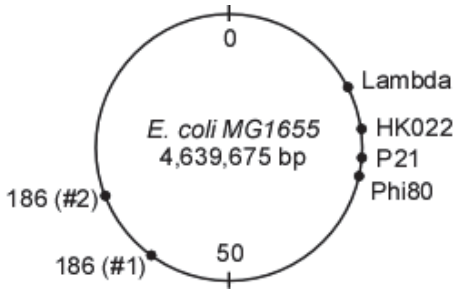
Supplementary Table 3. **Predicted sizes of PCR fragments following colony PCR verification of integration module excision.** Predicted sizes (in base pairs) are listed for PCR reactions with a primer specific to the insert (lacZ; primer FLIP_F: ATCTGGTGCTGGGTCTGGTG) and a primer specific to the integration site (primer P1, Supplementary Table 1). Given the high efficiency of pE-FLP, screening colonies for excision of the antibiotic marker and neighboring sequences is typically not necessary.



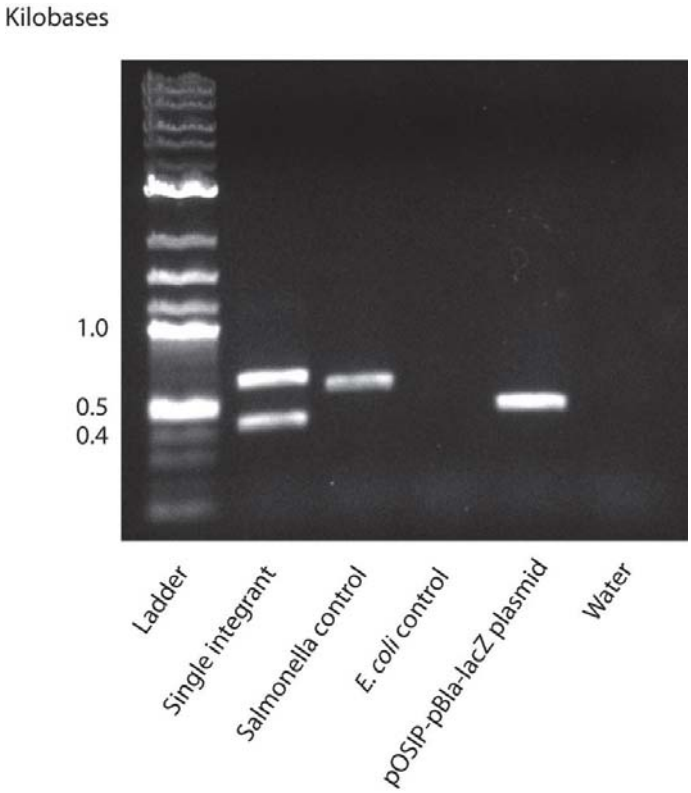
Supplementary Figure 1. Details of terminators and location of primer binding sites for PCR verification of integration and excision. (**a**) The identity and direction of the transcriptional terminators protecting the cloning module of pOSIP are shown. (**b**) We retain the nomenclature of Haldimann and Wanner *(1)* for naming the PCR primers used for verification of successful integration. Colony PCR screening for single integrants uses a mixture of primers P1 to P4 (Supplementary Table 1). The P2-P3 control fragment is produced when using pOSIP plasmid as a template. (**c**) Absence of integration produces a single *attB* fragment from primers P1 and P4. (**d**) Following successful site-specific recombination between the pOSIP *attP* sequence and chromosomal *attB* site, single integration produces two fragments, one from P1 and P2, and one from P3 and P4. (**e**) During rare events where multiple pOSIP integrate in tandem in the same *attB* site, primers P2 and P3 produce an additional fragment. (**f**) To verify correct excision of the pOSIP integration module by FLP, we used the appropriate P1 primer and an insert-specific primer, in this case a primer (FLIP_F) which anneals within the *lacZ* gene. The sizes of DNA fragments expected for all PCR reactions above are given in Supplementary Tables 2 and 3.



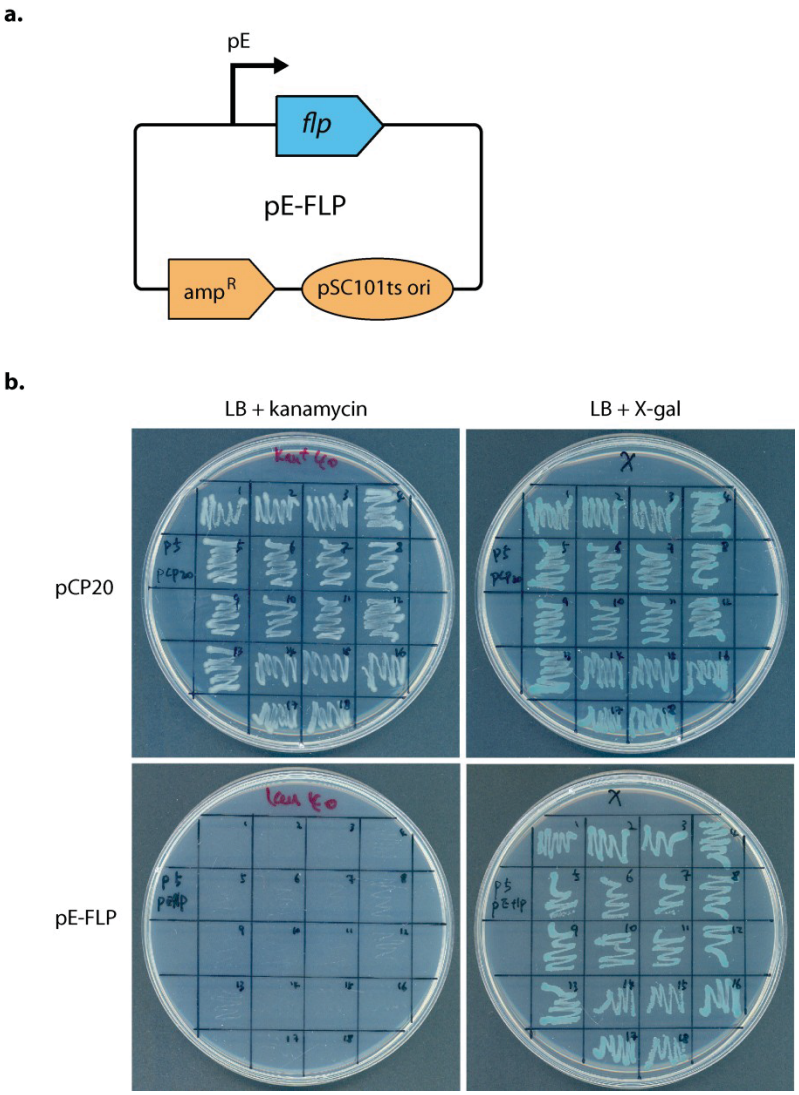
Supplementary Figure 2. PCR verification of integration of a lacZ cassette into six independent chromosomal loci. A pBla-lacZ cassette was clonetegrated using six pOSIP variants, each expressing a different phage integrase. For each construct, a single kanamycin-resistant colony was tested by colony PCR using the primers listed in Supplementary Table 1. The band pattern obtained is consistent with integration of the lacZ cassette at single copy into the bacterial chromosome (Supplementary Table 2). 186(#1) and 186(#2) correspond to the primary and secondary integration sites for the phage 186 pOSIP variant, respectively. A 2-log DNA ladder (N3200, New England Biolabs) provided size markers.



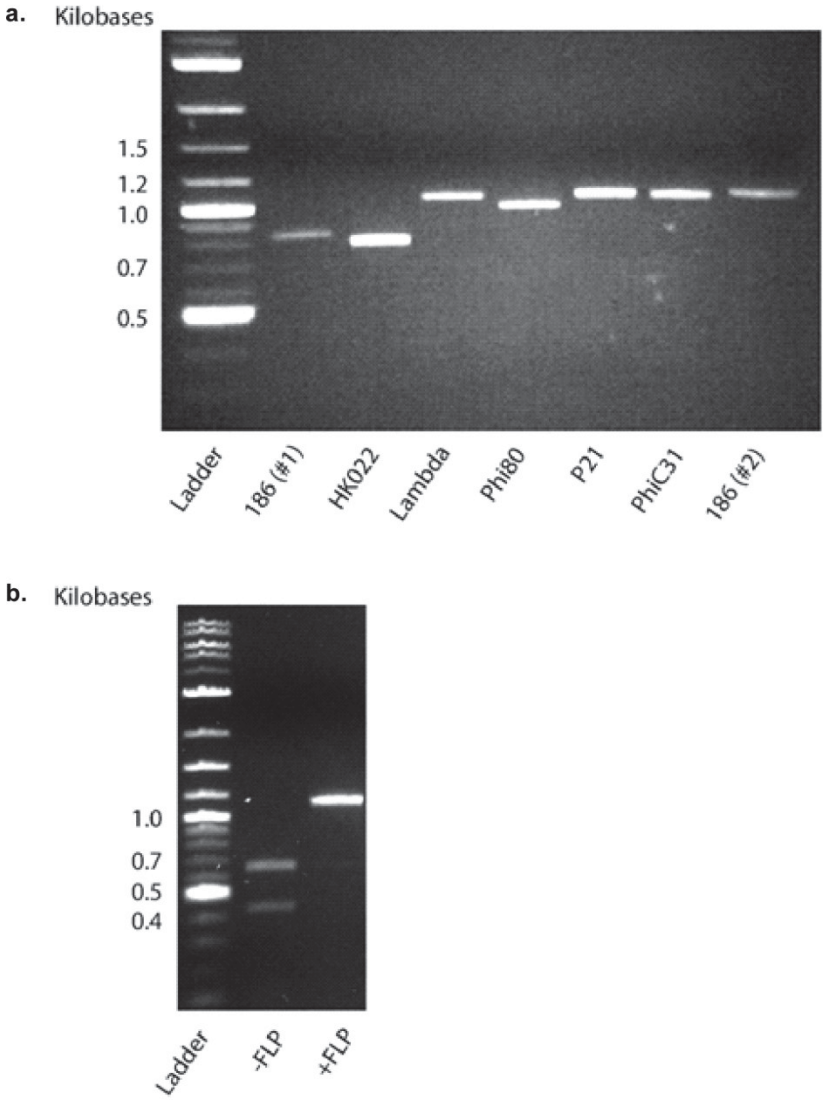
Supplementary Figure 3. Location of integration sites on the *Escherichia coli* chromosome. Positions of the integration sites for phages lambda, HK022, P21 and Phi80 – and corresponding pOSIP variants. Primary (186 #1) and secondary (186 #2) integration sites for the pOSIP variant expressing phage 186 integrase are also shown. '0' and '50' are position markers in minutes. This figure was adapted and expanded from Haldimann and Wanner *(1)*.



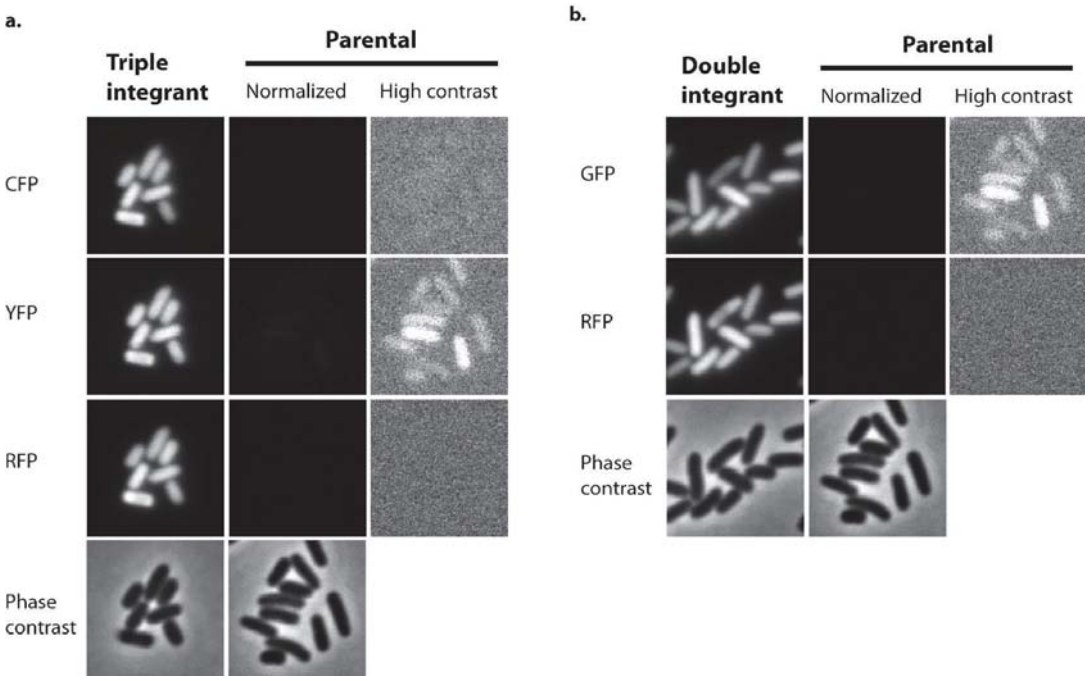
Supplementary Figure 4. **Clonetegration of a lacZ expression cassette in *Salmonella typhimurium*.** A pBla-lacZ expression cassette was clonetegrated into electrocompetent *S. typhimurium* strain LT2, which carries a naturally occurring phage 186 integration site, identical in sequence to the secondary 186 *attB* site of *E. coli*. Clonetegration efficiency, expressed as percentage of pUC19 transformation efficiency (Figure 1c), was 0.005 ± 0.0008 (mean ± s.e.m; n=3 independent transformations). To put this number into context, this efficiency is sufficient to give 140 colonies following transformation of 0.25 µl of a standard 20 µl assembly reaction into electrocompetent (2 x 10^9 cfu/µg pUC19) *Salmonella*. A kanamycin-resistant colony was picked for PCR verification of integration using the primers listed in Supplementary Table 1. The band sizes observed are consistent with a single chromosomal integration (Supplementary Table 2). The following control templates were also submitted to the same PCR verification test: the parental *Salmonella* strain, *E. coli* MG1655, a pOSIP plasmid with lacZ cassette and water. All control templates also produced the predicted band pattern (Supplementary Table 2). Note that the P1 and P4 primers for *Salmonella* lie outside the region of homology with *E. coli*, and give no product on an *E. coli* template. A 2-log DNA ladder (N3200, New England Biolabs) provided size markers.



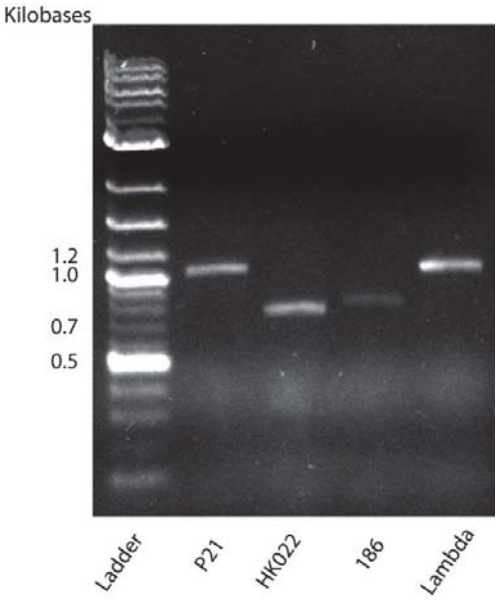
Supplementary Figure 5. **Excision of the pOSIP integrase module using the pE-FLP plasmid.** (**a**) Map of the pE-FLP plasmid. The strong promoter pE from phage P2 drives constitutive expression of flippase (*flp*). pSC101ts is a low-copy, temperature-sensitive origin of replication which is functional at 30ºC but not at >37ºC. (**b**) Comparison of the efficiency of integration module excision by pE-FLP and pCP20. A strain containing a single chromosomal copy of pOSIP-pBla-lacZ (kanR) in the lambda integration site was transformed with either pCP20 or pE-FLP. From each transformation plate, eighteen colonies were picked and streaked onto either LB + kanamycin (15 µg/ml) or LB plates containing X-gal (3%) but no antibiotics. The lack of growth on LB + kanamycin plates seen with all pE-FLP transformants (bottom left plate) indicates successful excision of the integration module. The blue color observed on LB + X-gal plates provides a visual demonstration that the pBla-lacZ cassette is retained and is functional (bottom right plate). In contrast, all 18 colonies from the pCP20 transformations grew on kanamycin (top left plate), suggesting failure to excise the pOSIP integration module.



Supplementary Figure 6. **PCR verification of excision of the pOSIP integrase module.**
(**a**) PCR verification of excision of the pOSIP integration module, including the antibiotic marker. The six strains of pBla-lacZ single integrants (Supplementary Figure 2) were transformed with pE-FLP and grown on ampicillin agar at 30ºC. For each strain, a single colony was selected for colony PCR using the primers listed in Supplementary Table 3. A 2-log DNA ladder (N3200, New England Biolabs) provided size markers. (**b**) A *S. typhimurium* strain with integrated lacZ expression cassette (Supplementary Figure 4) was transformed with pE-FLP. A single colony was selected for colony PCR using the primers listed in Supplementary Table 3. The band pattern obtained suggests successful excision of the integration module (Lane: +FLP; Supplementary Table 3). The band pattern prior to transformation with pE-FLP is shown for comparison (Lane: -FLP). A 2-log DNA ladder (New England Biolabs) provided size markers.



Supplementary Figure 7. Fluorescence of strains expressing fluorescent proteins (FPs) is not due to cellular autofluorescence. The parental strain targeted for serial (**a**) and parallel (**b**) integration of FP expression cassettes was imaged under the same conditions (exposure, lamp power and filter cubes) as the strains with integrated CFP, YFP, and RFP (Fig. 2b) or GFP & RFP (Fig. 2c) expression cassettes. Parental strain images were either scaled identically to the corresponding integrant images ("normalized" set), or scaled so as to produce a high contrast. CFP/GFP/YFP/RFP correspond to the set of excitation and emission filters used during imaging (Methods). Phase contrast images are shown as reference.



Supplementary Figure 8. Serial integration of a lacZ expression cassette into four chromosomal loci. A lacZ expression cassette was clonetegrated into the phage P21, HK022, 186 and lambda integration sites in four successive rounds. In each round, the pOSIP integration module was excised from a successful integrant using pE-FLP. The presence of chromosomally-integrated pOSIPs — with excised integration module — was confirmed for all four integration sites by colony PCR using the primers listed in Supplementary Table 3. A 2-log DNA ladder (New England Biolabs) provided size markers.

References – Supporting Information

1. Haldimann, A., and Wanner, B. L. (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria, *J Bacteriol 183*, 6384-6393.
2. Shetty, R. P., Endy, D., and Knight, T. F., Jr. (2008) Engineering BioBrick vectors from BioBrick parts, *J Biol Eng 2*, 5.
3. Linn, T., and St Pierre, R. (1990) Improved vector system for constructing transcriptional fusions that ensures independent translation of lacZ, *J Bacteriol 172*, 1077-1084.
4. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection, *Mol Syst Biol 2*, 2006 0008.
5. Callen, B. P., Shearwin, K. E., and Egan, J. B. (2004) Transcriptional interference between convergent promoters caused by elongation over the promoter, *Mol Cell 14*, 647-656.
6. Cherepanov, P. P., and Wackernagel, W. (1995) Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant, *Gene 158*, 9-14.
7. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products, *Proc Natl Acad Sci U S A 97*, 6640-6645.
8. Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000) An efficient recombination system for chromosome engineering in Escherichia coli, *Proc Natl Acad Sci U S A 97*, 5978-5983.
9. Sambrook, J., Russell, D. (2001) *Molecular Cloning: A Laboratory Manual*, CSHL Press.
10. Chung, C. T., Niemela, S. L., and Miller, R. H. (1989) One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution, *Proc Natl Acad Sci U S A 86*, 2172-2175.
11. Palmer, A. C., Ahlgren-Berg, A., Egan, J. B., Dodd, I. B., and Shearwin, K. E. (2009) Potent transcriptional interference by pausing of RNA polymerases over a downstream promoter, *Mol Cell 34*, 545-555.
12. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nat Methods 6*, 343-345.
13. Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X. S. (2006) Probing gene expression in live cells, one protein molecule at a time, *Science 311*, 1600-1603.
14. Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications, *Nature biotechnology 20*, 87-90.
15. Goedhart, J., van Weeren, L., Hink, M. A., Vischer, N. O., Jalink, K., and Gadella, T. W., Jr. (2010) Bright cyan fluorescent protein variants identified by fluorescence lifetime screening, *Nat Methods 7*, 137-139.
16. Lam, A. J., St-Pierre, F., Gong, Y., Marshall, J. D., Cranfill, P. J., Baird, M. A., McKeown, M. R., Wiedenmann, J., Davidson, M. W., Schnitzer, M. J., Tsien, R. Y., and Lin, M. Z. (2012) Improving FRET dynamic range with bright green and red fluorescent proteins, *Nat Methods 9*, 1005-1012.

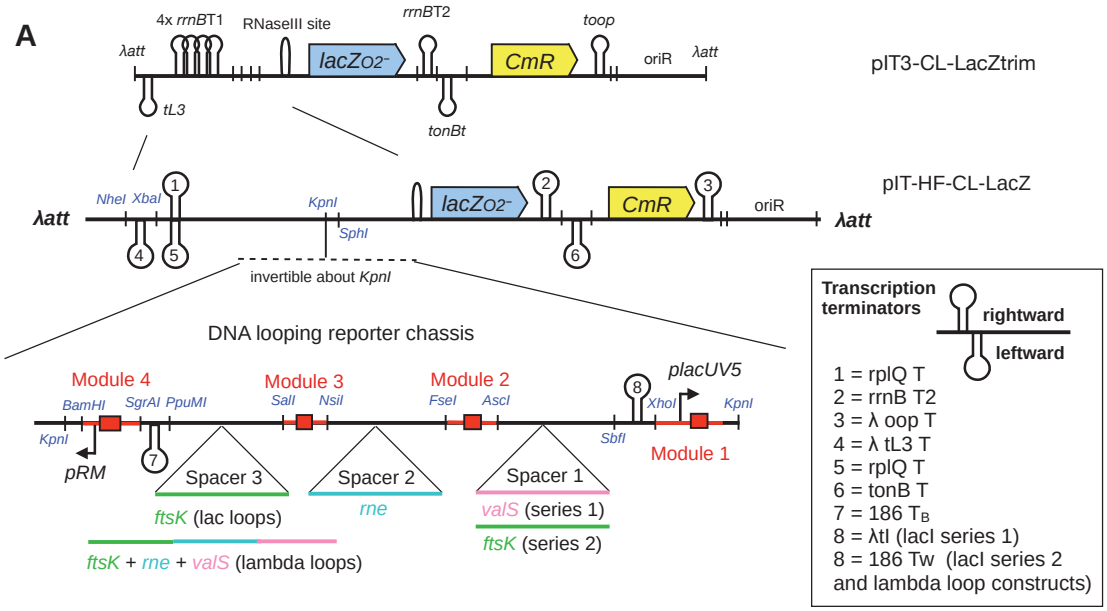
17. Irwin, C. R., Farmer, A., Willer, D. O., and Evans, D. H. (2012) In-fusion(R) cloning with vaccinia virus DNA polymerase, *Methods Mol Biol 890*, 23-35.
18. Dodd, I. B., Perkins, A. J., Tsemitsidis, D., and Egan, J. B. (2001) Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny, *Genes Dev 15*, 3013-3022.

4.4 A DNA looping reporter 'chassis'

To study DNA looping comprehensively, a number of combinations of operators separated by a range of spacings was required. Therefore Ian Dodd designed a modular DNA looping reporter chassis, into which different operators and spacers can be cloned (Figure 4.2A). The chassis consists of four 'modules' into which operators can be cloned, separated by three 'spacers' where variable lengths of spacer DNA can be inserted. Having four modules available allows testing of both lac and λ looping separately or together. The two outer modules can contain $P_{lac^{UV5}}$ (and a lac operator, Module 1) or P_{RM} (within *oR*, Module 4), and since the chassis can be inverted about the *Kpn*I sites, it is possible to measure effectively the same reporter chassis from either promoter. For example with an alternating arrangement of λ*CI* and *lac* operators, it is possible to measure both the interference of Lac looping on λCI looping and vice versa (see Chapter 3).

The chassis was designed initially with restriction enzyme-based cloning in mind, and each modular element is flanked by a unique pair of restriction enzyme sites to allow swapping of operators and spacers. Five initial chassis were ordered as synthetic DNA (from GenScript) (Figure 4.2B), containing different operators in each module position. The modules themselves are of a set length, with space being filled in with DNA from the *A* gene from phage 186, thereby allowing operators to be swapped without changing the spacing between them.

Estimates of DNA looping are based upon measurements of $P_{lac^{UV5}}$ and P_{RM} activity, however there is potential for some of the measured LacZ activity to result from transcription arising from unknown or cryptic promoters. The design of the looping reporter chassis sought to minimise this background in three main ways. Firstly, transcription emanating from the *E. coli* chromosome near the λ*att* site should be prevented from entering the reporter by transcriptional terminators outside the reporter chassis (Figure 4.2A). Secondly, it was reasoned that the best way to minimise incorporation of cryptic promoters in spacer DNA was to source it from within *E. coli* genes (*valS*, *rne* and *ftsK* genes) (however note that there was some evidence that the nature of the spacer DNA affected lac looping (Chapter 2)). Finally, terminators were also placed upstream of $P_{lac^{UV5}}$ and P_{RM} in order to minimise background transcription emanating from within the chassis. Despite these precautions, it was expected that some background would re-



B

Synthetic DNA	*Module4*	*Module3*	*Module2*	*Module1*
pUC57-DS1-Kpn-flip	*O⁻*	*O⁻*	*Oid*	*O3*
Seq2:	*oR⁺(pR⁻)123*	*oL⁺(pL⁻)321*	*O⁻*	*O⁻*
Seq3:	*oR-r1*	*O1*	*oL⁺(pL⁻321)*	*O2*
Seq4:	*Oid*	*oL3-4*	*O1*	*O1*
Seq5:	*O1*	*Oid*	*oL3-4*	*oL⁺(pL-)321*

Figure 4.2: **A.** DNA looping reporter chassis showing arrangement of modules, spacers and restriction sites. The chassis is invertible about the *Kpn*I site, such that the same looping reporter can be read from either the P_{lacUV5} or P_{RM} end. Right: Legend detailing terminators. Note that *λtI* was replaced with *186tW* in some λCI reporters. **B.** Shows the contents of the modules in each of the five synthetic chassis that were ordered by gene synthesis.

main, and a number of approaches were taken to measure this background for $P_{lac^{UV5}}$, which are detailed in Section 4.7.

About halfway through this study the Shearwin lab adopted a new cloning technique called Gibson Assembly [Gibson, 2011], whereby introduction of 30–40 bp of homology at the ends of DNA fragments allows their assembly. This homology can be introduced in the tails of PCR primers, and serendipitously, since each module in the reporter chassis is flanked by unique fragments of the *A* gene, it was possible to design primers to uniquely amplify each module and spacer for Gibson assembly, and therefore we didn't have to redesign the DNA looping chassis. Gibson Assembly dramatically sped up the process of constructing DNA looping chassis because whilst restriction enzyme-based cloning is restricted to swapping one module or spacer at a time, it was found that with Gibson Assembly an entire chassis, consisting of up to seven PCR products (4 modules and 3 spacers) could be assembled from scratch. This was achieved by performing a seven fragment 'linear' assembly (of the 4 modules and 3 spacers), which served as the template for a PCR with the two flanking primers of the chassis, with this PCR thereby selecting the correct chassis from background misassembled products. This PCR product was then gel extracted and assembled in a two fragment assembly with suitably-cut pIT vector backbone and either transformed straight into the target strain ('clonetegration'), or propagated in and prepared from the pir^+ strain (E4644) for integration.

For reference, tables 6.3 and 6.4 detail the contents and construction of the lac looping only (LacLoop (LL)) and the lac and λ (LacLamLoop (LLL)) strains, whilst tables 6.5 and 6.2 detail the Gibson Assembly fragments and the primers used to amplify them respectively. Restriction enzyme-based cloning was used to construct the earlier Lac looping constructs (LL1-18), whereas assembly was used to make most of the remaining reporters.

4.5 An example of restriction enzyme-based cloning of reporters

Instead of explaining the construction of all 44 of Lac Looping constructs, it is instructive to explain the process of creating one, LacLoop2 (LL2), in detail. Further detail on the various techniques employed (such as gel extraction, and transformation) can be found in Section 6.1.2.

LacLoop2 is the distal operator minus control to pair with LacLoop1 (*Oid*.300bp.*O3*), and

comparison of these two constructs reveals the contribution of repression by DNA looping vs simple binding of LacI to the promoter-proximal operator. The synthetic plasmid pUC57–DoddSeq2 was digested with *Nsi*I/*Asc*I (Module 2), giving a 142 bp fragment containing the O^- operator, which was gel extracted and ligated to similarly cut, sapped (antarctic phosphatase treated) and gel extracted pUC57–LacLoop1. Primers 852/853 were used to amplify across Module 2 and subsequently sequence it from primer 853, thus yielding pUC57–LacLoop2.

Next, a miniprep of pUC57–LL2 was digested with *Nhe*I/*Sph*I-HF and cloned into similarly cut pIT3–CL–LacZtrim, with ligations transformed into TSS competent [Chung et al., 1989] E4644 cells. We chose the low copy number *pir*$^+$ strain (E4644) instead of the high copy number *pir116* strain (E4645) to avoid possible problems of having numerous unrepressed $P_{lac^{UV5}}$ promoters in the cell. However even in this strain we encountered some problems, for example with LacLoop5 (*O1*.300bp.*O1*), which suffered frequent deletion about the two *O1* operators, highlighting the selective advantage of removing strong promoters (and their associated products) from the population of cells (See Section 4.8). Transformants of pIT3–CL–LacLoop2–LacZtrim (herein referred to pIT–HF–LacLoop2, where 'HF' stands for Human Frontiers), were screened with primers 462/440 by colony PCR to yield a 1618 bp product, which was subjected to sequencing.

Finally, a miniprep of pIT–HF–LL2 was prepared from E4644 and transformed into TSS-competent E4643 (containing an integrated *lacI* expression module and the pλINT helper plasmid). pIT–HF–LL2 is 6243 bp, and transformation into TSS-competent cells worked well, however later when integrating larger plasmids (such as pIT–HF–LL29, *Oid*.3200bp.*O2*, 9079 bp), it was found that TSS transformation became less efficient. We found that this size limit could be pushed further (to about 12–15 kb) by using electrocompetent cells, but we are unsure what is the limiting step for integration of such large plasmids; e.g. traversal by the DNA of the outer and inner cell membranes, the actual process of integration into the chromosome or some hitherto unknown size-based restriction. Integrants were tested for single integration by colony PCR with the λ*att* primers (primer numbers 467, 468, 466 and 469), which generate different products for no–, single– or double–integrants [Haldimann and Wanner, 2001]. Once integrated, we found that looping reporters were very stable and successful integrants were then tested for

their LacZ activity.

4.6 Developing a LacZ assay protocol

The β-galactosidase (LacZ) protein accumulates in the cell to a steady state level proportional to the activity of the promoter driving its transcript. Any differences in the 5' end of transcript are controlled for by introducing an RNAse III cleavage site just upstream of the *lacZ* open reading frame, such that all transcripts have the same 5' end. The amount of LacZ enzyme per cell (and thus promoter activity) can be assayed by lysing (or permeabilising) a log-phase culture of cells and adding a colorimetric substrate for LacZ (Xgal for plate assays and ONPG for liquid culture assays). The original LacZ assay developed by Miller (detailed in [Miller, 1992]) was conducted in glass test tubes, and is an endpoint assay in that the rate of accumulation of the coloured product is calculated between two time points (start and end). Ian Dodd [Dodd et al., 2001] made a number of improvements to this assay including: (1) adapting the assay to work in plastic 96-well microtitre plates, which (2) allowed the rate of cleavage of ONPG to be measured (by A_{414}) in a plate reader by fitting a straight line to a number of consecutive readings (kinetic instead of end-point assay) and later, (3) combining the lysis buffer and ONPG into one buffer (which we call TZ8+), since it was found that when using the cell permeabilisation reagent polymyxin-B, pre-incubation was not required.

For this study we naturally wanted LacZ measurements of the highest possible precision and to this aim, one aspect of the LacZ protocol to be changed was to switch the growth medium from the poorly defined Lennox Luria Broth (LB) to the well-defined M9 Minimal Medium (M9MM) (`http://openwetware.org/wiki/M9_medium`). By using a well-defined medium, variation in LacZ units due to batch-to-batch differences in growth medium is minimised. Initial attempts to grow and assay strains in M9MM (1x M9 salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% glycerol) showed variable growth and variable LacZ measurements, for example an assay conducted on LacLoop1 (LacI absent) gave an average of 1202 ± 69 (stdev) units in one assay (02/08/2011), but only 554 ± 34 units on another day (30/08/2011). Interestingly however, it appears the units measured on 2/8/11 seemed to be simply scaled down from those of 30/8/11 as similar repression values were obtained. Furthermore, growth rates between the two days were

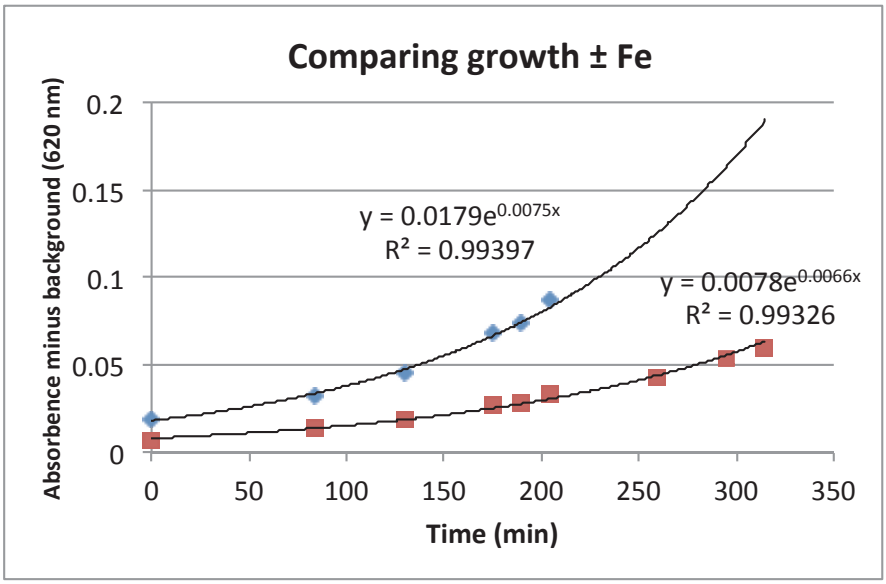


Figure 4.3: Cultures of E4643 with a Lac Looping reporter were grown in the absence (red squares) or presence (blue diamonds) 0.01 mM ferrous ammonium sulphate and their growth was monitored by measuring absorbance at 620 nm.

similar.

This variability in LacZ units and growth rate, combined with the fact that we were using a 'minimal' medium, suggested that the cells were deficient in an essential micronutrient. I located a study that found addition of ferrous (iron(II)) sulphate ($FeSO_4 \cdot 7H_2O$) to M9MM (to 0.01 mM) resulted in a decrease in the doubling time of the strain BL21 from 102 min to 71 min [Paliy and Gunasekera, 2007]. Subsequently, I added ferrous ammonium sulphate ($(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$) to (0.01 mM) and found that the doubling time of E4643 decreased from around 108 min to 90 min (Figure 4.3). Due to the tiny amount of iron needed to boost cell growth, we suspected that our Milli-Q H_2O (Millipore) was 'too pure' and that minor variations in the background concentration of iron introduced variable growth. Thus, as a result all subsequent cultures were grown in the presence of 0.01 mM ferrous ammonium sulphate, and batches of media were prepared and stored at −20°C since a stock solution of $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$) was observed to oxidise when stored at room temperature. The final LacZ M9MM lacZ assay protocol can be found in the extended materials and methods (Section 6.1.5).

4.7 Estimating background units of the looping reporter chassis

When reporting on a promoter in *E. coli* some of the measured activity may result from cryptic transcription arising elsewhere from the promoter being interrogated. Cryptic transcription can arrive at the reporter gene from both directions; the sense-directed background leads to an overestimation of activity whilst antisense transcription reduces the measured signal via transcriptional interference [Callen et al., 2004]. Background can be reduced to minimal levels by judicious use of transcriptional terminators, of which many types exist (Fig 4.2). Nevertheless, residual background poses a problem especially for strongly-repressed promoters, such as $P_{lac^{UV5}}$, for which previous measurements show nearly complete repression to ~ 1–2 LacZ units from $\gtrsim$ 1000 unrepressed units [Müller et al., 1996]. For example, for LacLoop8 (*Oid*.300bp. $P_{lac^{UV5}}$ *O2*) without subtracting any background, the repression factor (the ratio of unrepressed to repressed promoter activity) at the high LacI concentration is $\frac{1159}{37} = 31.3$, whereas subtracting 18 background units yields $\frac{1141}{19} = 60.1$. Despite the suggested near complete repression of $P_{lac^{UV5}}$, in our reporters the lowest activity we measured was ~ 18 units for LacLoop13 (*Oid*.300bp. $P_{lac^{UV5}}$ *Oid*) at the high LacI concentration. This puts a lower limit of ~ 18 units of background in our reporters, and it was eventually decided to subtract this amount from all the data (see below).

The reporter gene for our measurements is the *lacZ* gene, and to ensure that there were no other sources of β-galactosidase activity in the parent strain or assay reagents, an assay was performed with the parental strain that lacked the DNA looping reporter (E4643 with a LacI-expressing integrant). As expected, this strain yielded zero units, thereby confirming that any background must originate from the *lacZ* gene in the looping reporters.

For plasmid-based reporters, background transcription can arise anywhere in the plasmid however chromosomally-integrated reporters have the additional possibility of background originating from the bacterial chromosome near the integration site. The pIT3 series of integrating vectors contain a number of transcriptional terminators to attenuate this transcription (Fig 4.2). As an estimate for the remaining transcription entering the looping reporters, pIT–HF–LL4 was digested with *Acc*65I (an isoschizomer of *Kpn*I) and then religated, effectively creating a pIT–HF–CL–LacZtrim ‘empty-vector’ control. Integrants of this plasmid into the LacI-expressing

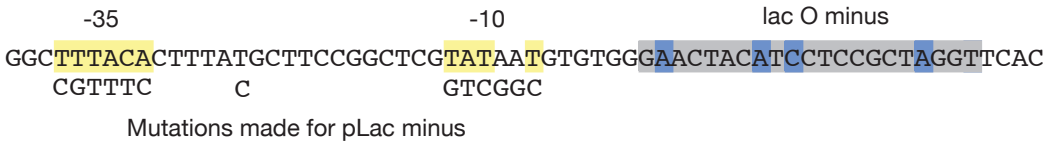


Figure 4.4: Showing the DNA sequence changes to make $P^{-}_{lac^{UV5}}$. Features are annotated on figure.

strains gave 40–50 units (Fig 4.5), however whether this is the true amount of background entering the multiple cloning site of pIT–HF–CL–LacZtrim is unknown since the *Kpn*I deletion may have introduced a cryptic promoter at the deletion breakpoint.

$P_{lac^{UV5}}$ poses a problem for background estimation since the Lac repressor is a known roadblock to transcription [Sellitti et al., 1987] (also submitted manuscript by Shearwin Lab member Nan Hao). Thus the level of background varies depending on the strength of LacI-mediated roadblocking, which depends on the level of LacI present and the number and strength of the *lac* operators. To minimise incorporation of cryptic promoters, spacer DNA in the looping reporters was taken from within *E. coli* genes. Furthermore, to minimise background originating from both upstream and within the DNA looping chassis, a terminator was placed just upstream of $P_{lac^{UV5}}$ (Fig 4.2).

To obtain a background estimate in the most realistic possible context, we decided to mutate $P_{lac^{UV5}}$ by scrambling its −10 and −35 sites (Fig 4.4). Two constructs were assayed, LacLoop10 (O^{-}.300bp. $P^{-}_{lac^{UV5}}$ *O1*) and LacLoop11 (O^{-}.300bp. $P^{-}_{lac^{UV5}}$ O^{-}). LacLoop10 and 11 showed a maximum of 80–100 units (Fig 4.5), thereby placing an upper limit on the background at 80–100 units. It could be argued that our $P^{-}_{lac^{UV5}}$ still has residual promoter activity, which would lead to an overestimation of the background, however the repression factor (at low LacI) of LacLoop10 is ∼ 1.6, much less than that of the equivalent $P^{+}_{lac^{UV5}}$ construct (LacLoop6, repression factor ∼ 4), which is consistent with the fact that roadblocking by LacI has a lesser effect than repression by promoter occlusion (Fig 4.5). The extent to which the repressive effect seen for LacLoop10 is due to repression of residual $P^{-}_{lac^{UV5}}$ promoter activity vs roadblocking of upstream background could be tested by increasing the distance between $P^{-}_{lac^{UV5}}$ and *O1* such that *O1* no longer overlapped the promoter (e.g create a construct such as O^{-}.300bp. $P^{-}_{lac^{UV5}}$ O^{-}.100bp.*O1*), however due to time constraints, such a construct has not yet been made.

LacI binding to the strong distal operator will pose an additional roadblock to any background transcription upstream, however a measure of the contribution of this effect was not obtained.

In conclusion, the level of background is expected to vary between 18–100 units depending on the construct, however we chose to use a value of 18 units for all reporters, which simplifies the modelling, and which we believe is justified for two reasons. Firstly, for the LacI-absent strains unrepressed $P_{lac^{UV5}}$ activity in our assay varies between ~ 1050–1250 units depending on the promoter proximal *lac* operator, and the difference between subtracting 18–100 units of background (~ 2–9 % of promoter activity) has minimal effect. Secondly, background for $P_{lac^{UV5}}$ in the presence of LacI and a promoter-proximal operator will vary between ~ 18 (most repressed) and ~ 60–80 units (e.g. see the value for LacLoop10 at low LacI). Therefore, the level of background will be a slight underestimate for some constructs (such as LacLoop9, O^{-}.300bp. $P_{lac^{UV5}}$ *O2*), however this does not greatly affect the model fits, and furthermore, having to fit multiple measures of background would have reduced the reliability of the fitted values for parameters such as J_{loop}.

4.8 Problems encountered when working with DNA

Readily accessible troughs in the fitness landscape will almost always predominate in a population of *E. coli*, and it is up to the experimenter to identify and circumvent all avenues a cell might pursue to rid itself of selective burdens introduced by the experimental construct. Whilst unfavourable DNA constructs (e.g. containing strong promoters) can readily be co-selected with antibiotic resistance, the danger remains that promoters will be annulled or removed by mutation or recombination. Recombination appears to be favoured when constructs are being maintained on multi-copy plasmids, since in our experience, even repeat-bearing constructs seem to be stable upon chromosomal integration. Note however that Claudia Vickers from the Australian Institute for Bioengineering and Nanotechnology (AIBN) informs us of her experience that large repetitive integrations into *E. coli* can result in large chromosomal deletions of the DNA between the repeats (Claudia Vickers personal communication). As a matter of course, I suggest scanning all DNA constructs with an algorithm to detect repeats before constructing them (such as REPFIND: `cagt.bu.edu/page/REPFIND_submit?`). This section provides some examples of

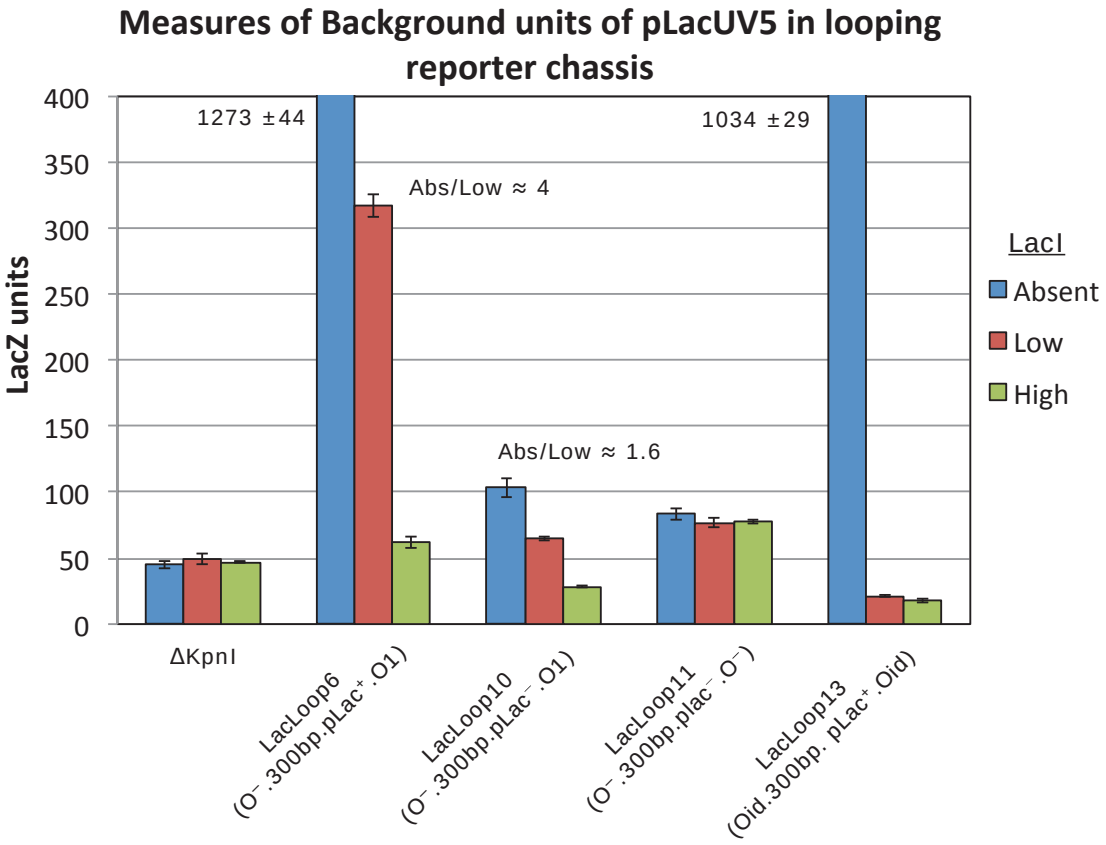


Figure 4.5: Different measures of background units of Lac Looping reporters. LacZ assays of indicated strains were performed at different LacI levels. Y-axis is truncated to show detail at lower values. High values are indicated with their 95%CI. The repression factors indicated on the graph (Abs/Low) are the ratio of LacZ units in the absence to the presence of low LacI levels.

problems we encountered when handling reporter constructs.

4.8.1 Recombination deletes a strong promoter

$P_{lac^{UV5}}$ is a relatively strong repressible promoter, yielding ~ 1000 units in our LacZ assay. Maintaining this promoter on a high copy number plasmid will result in high LacZ levels, which may inhibit cellular growth and therefore provide selective pressure to mutate (or delete) the promoter. To minimise promoter mutations, we initially maintained all pIT–HF–LacLoop plasmids in the low copy *pir*$^{+}$ strain (E4644 in the Shearwin lab stocks). In LacLoop 5 $P_{lac^{UV5}}$ is flanked by two *O1* operators (*O1*.300bp. $P_{lac^{UV5}}$ *O1*), and we feared $P_{lac^{UV5}}$ would be susceptible to deletion by recombination across the *O1* repeats. A PCR across the region (with primers 853/863), luckily showed that the original integrants were fine, however when I streaked out E4644 harbouring pIT–HF–LacLoop5, some colonies had indeed undergone a recombination across the *O1* operators, and this was subsequently confirmed by sequencing the short (deletion) PCR product. LacLoop 6 does not contain such a repeat (O^{-}.300bp. $P_{lac^{UV5}}$ *O1*), and accordingly the short PCR product was not observed after propagation of pIT–HF–LacLoop6. E4644 is *rec*A^{+}, which may explain this susceptibility to recombination. As a further precaution, we decided to transform the LacI expression vector pUHA1 into E4644 (which is LacI negative), to keep $P_{lac^{UV5}}$ repressed and remove the selective pressure (however see below).

4.8.2 Unwanted co-selection of a LacI-expressing plasmid

In order to minimise deletion of $P_{lac^{UV5}}$ from pIT plasmids, subsequent constructs were maintained in E4644 + pUHA1, which would remove the cellular burden through repression of the promoter (see above). DNA looping chassis were made by Gibson Assembly and transformed into E4644 + pUHA1 from which a miniprep was prepared for integration into the LacI-expressing target strains. Upon conducting test LacZ assays, I often observed unexpectedly low units in the LacI absent strain (where we expect unrepressed $P_{lac^{UV5}}$ activity), and initially I suspected this was due to mutations in $P_{lac^{UV5}}$ or the *lacZ* gene. Subsequently however, I re-streaked some of these strains ± Kanamycin, and found that they all harboured pUHA1, even in the absence of any Kanamycin in the growth media. Cells are normally readily cured of

plasmids that are not being selected for by their encoded antibiotic resistance, however in this case we believe pUHA1 was being co-selected due to its ability to repress $P_{lac^{UV5}}$. Attempts to cure pUHA1 from integrant strains by re-streaking proved fruitless, with only about 10% curing rate after multiple re-streaks. This problem was subsequently solved by gel extracting pIT–HF–LacLoop minipreps away from pUHA1 prior to integration. This experience highlights that gel extraction should be a routine practice at most steps in the preparation of DNA.

4.8.3 Recombination about a repeated terminator leads to an inversion

The original set of LacLamLoop constructs reading P_{RM} (LLL4, 5, 6, 13, 14, 17 & 19) had to be re-made due to a deleterious recombination event. We were alerted to a possible error in the original set of constructs by the observation that whereas P_{RM} looping constructs built by Lun Cui were repressed down to ~ 150 units, those made by myself were hardly repressed from the 400 units of constitutive P_{RM} activity (Fig 4.7). This result was hard to understand, however I noticed that Cui also had trouble with his constructs and had fixed the issue by replacing the *λtI* terminator in the chassis with *186tW* (Fig 4.6). Unfortunately, there is another *λtI* terminator within the λ*att* site that was not identified during the design of the chassis and we eventually discovered that in the original set of strains, the DNA flanked by the two *λtI* sequences had inverted (Fig 4.6). The effect of this inversion was to direct $P_{lac^{UV5}}$ inward to the chassis, from which transcription may dislodge λCI from *oL* and thereby reduce λ looping, to produce the lower-than-expected observed P_{RM} repression in these strains (Fig 4.7).

4.8.4 *λtI* inversion explains the original dataset

The problems encountered here highlight the importance in designing experiments that isolate and test one particular variable (e.g. looping interference), yet even the most well thought out experiments can have flaws that introduce confounding factors. In order to draw conclusions, the experimenter must perform as many controls as feasible and be critical of their own data whilst seeking other's opinions. Sometimes all the controls will pass but the data will still not make sense, in which case one has to think hard about the possible cause, which, once discovered may explain the seemingly unexplainable.

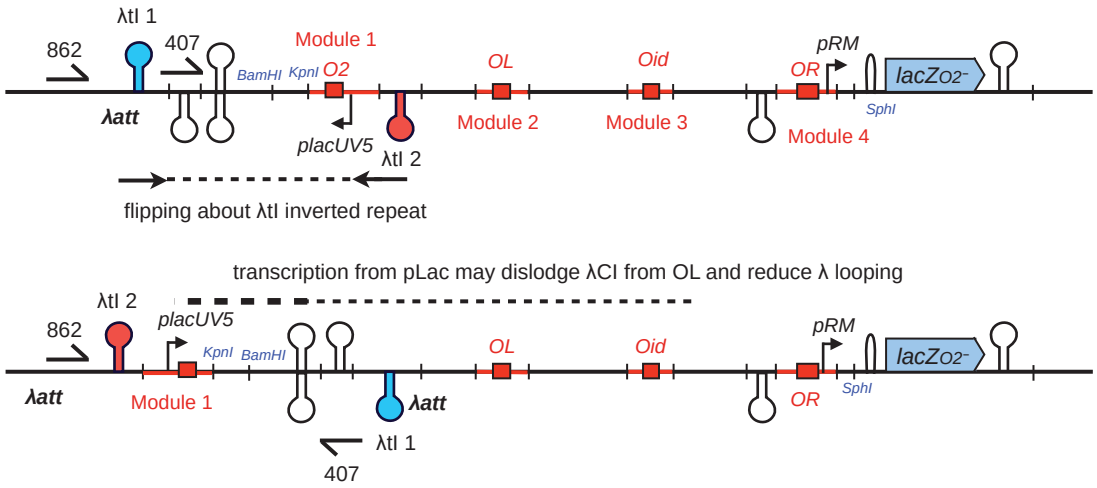


Figure 4.6: Inversion about the *λtI* terminators. The DNA looping chassis in the normal (top) and inverted (bottom) orientations, showing how the inversion directs $P_{lac^{UV5}}$ transcription towards *oL*, where it may dislodge λCI and reduce λ looping. Primers 862 and 407 yield a defined product only in the inverted orientation.

For example in the case of the inversion about *λtI*, the only original construct that had normal repressed P_{RM} levels was LacLamLoop17 (the Lac within Lambda nested construct). Whilst all other constructs have $P_{lac^{UV5}}$ between the λ*att* terminator repeat, LLL17 has *oL*, and therefore the inversion simply placed *oL* in the opposite orientation, which as shown previously [Cui et al., 2013] has no effect on λ looping.

LacLamLoop19 (Lac and Lambda side-by-side, reading P_{RM}) was hard to explain since it had high unrepressed P_{RM} only in the absence of LacI. Initially I thought the (LacI [High], No λCI) strain was erroneous, however this result makes sense in the context of the inversion since in the presence of LacI, $P_{lac^{UV5}}$ will be repressed and not affect λCI looping. Indeed this LacLamLoop19 strain isolates the inversion effect to a single variable since it was later shown with LacLamLoop19new that side-by-side loops have no effect on one-another (Chapter 3)).

4.9 LacI protein work

As discussed in Priest et al. [2014] supplementary information (Chapter 2), model fits to lac looping LacZ data were aided by knowing the concentration of LacI in the Low (L1) and High (L2)

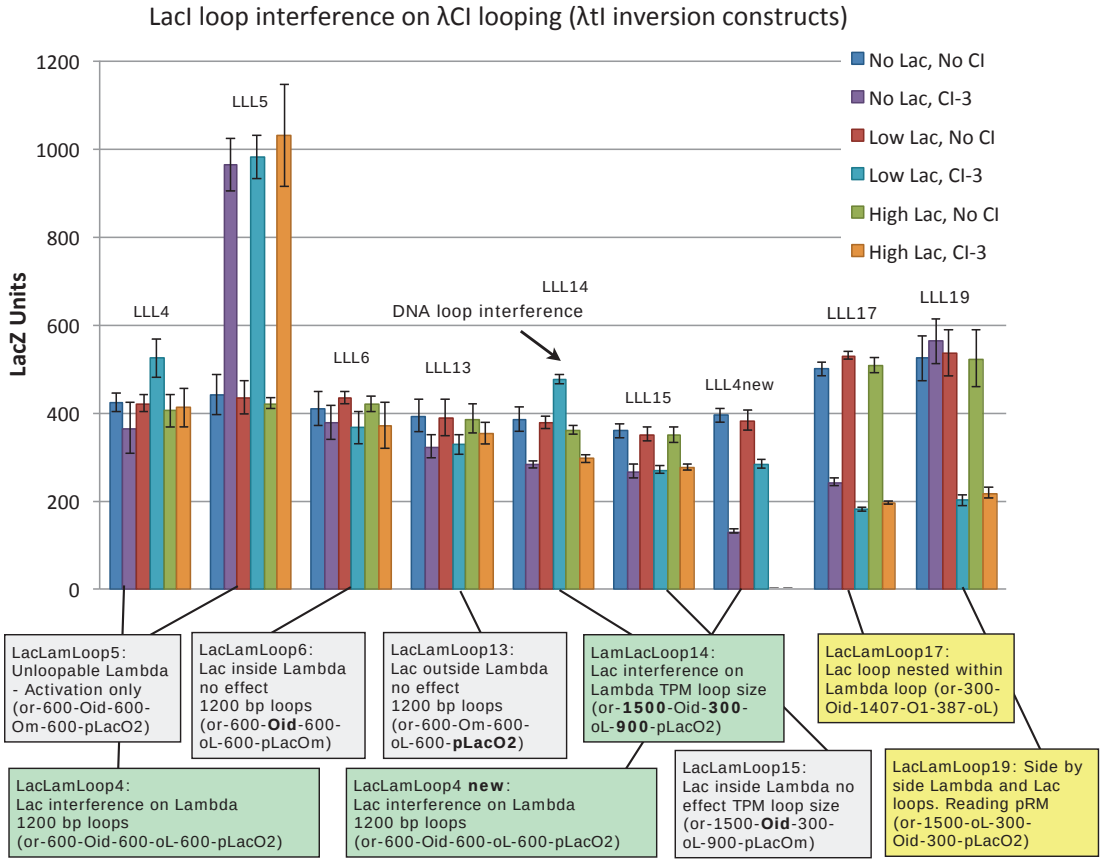


Figure 4.7: LacZ data for original set of constructs reading from P_{RM} as well as the repaired construct LLL4 new, which shows proper repression of P_{RM}. Arrow on LLL14 (original interference strain) shows that despite the inversion, there was still evidence of DNA loop interference in these original strains. Green shading shows loop interference strains and grey shading shows the various operator-minus controls. Yellow shading shows the original nested (LLL17) and side-by-side (LLL19) constructs, which show generally higher units due to day-day variation. LLL17 indeed shows loop assistance in the presence of LacI.

[LacI] strains. In the paper we used a previous estimate of 18 nM for LacI tetramers supplied from a chromosomally-integrated wild-type P_{lacI} promoter for the L1 concentration [Garcia and Phillips, 2011]. The L2 concentration was then estimated from L1 by comparing the intensity of LacI bands from extracts of the two strains on a Western Blot using a commercially available LacI antibody (Rockland). Additionally, we also generated translational fusions between P_{lacI} (or $P_{\mathrm{lacI^q}}$) and the first 15 amino acids of *lacI* with *lacZ*. With these methods we obtained an estimate for the ratio of L2/L1 of $\sim$ 10.9, which is in agreement with previous estimates that $P_{\mathrm{lacI^q}}$ is 10x stronger than P_{lacI} (see Chapter 2 Fig S4) [Müller-Hill, 1975].

In addition to these efforts mentioned in the paper, attempts were initiated to quantitate the *absolute* level of LacI in the cell (in numbers of molecules per cell) through: (1) expressing and purifying His-tagged LacI protein and (2) using this protein as a protein standard on Western Blots as well as (3) an antigen for immunisation of rabbits to generate an in-house anti-LacI polyserum. The following sections outline these efforts. Although His-tagged LacI was successfully expressed, it was insoluble, which prevented removal of the His tag via *in vitro* TEV protease digestion. The presence of the His tag on our purified LacI precluded its use as an accurate protein standard because it causes the protein to migrate along the gel at a higher molecular weight, and may affect antibody recognition. Our anti-LacI polyserum turned out to be slightly less efficient than a commercially available antibody.

4.9.1 Purification of His-tagged LacI protein

Although methods for purification of untagged LacI protein exist [Xu and Matthews, 2009], we sought to simplify the process by tagging LacI with a 6x Histidine tag (His tag), thereby allowing purification by nickel affinity chromatography, for which small culture volumes ($\lesssim$ 1L) and therefore small columns (5 mL) can be used. The presence of the His tag however would preclude using His6-LacI as a protein standard on a Western Blot for absolute protein quantitation and so we included a TEV protease recognition site between the His tag and LacI, such that the His tag could be cleaved off. However TEV protease cleavage requires soluble protein substrate, but unfortunately as will be discussed, our expressed LacI was insoluble and the purified protein was not amenable to refolding by dialysis.

The T7 forward and reverse primers (215/315) were used to amplify a DNA fragment for creating His6-TEV N-terminal fusions from the plasmid pAC28–H6TEV (a kind gift from Anne Chapman-Smith), and was cloned into the *Xba*I/*Blp*I site of pET3a–RNA2–His6 (from Keith Shearwin). The RNA2–His6 fragment was cut out, and its presence facilitated size detection of cut plasmid. Next, primers 903/904 were used to amplify the *lacI* gene (from pUHA1) with correct ends for cloning into the *Nco*I/*Bam*HI site of pET3a–His6TEV. In order to clone *lacI* in frame into the *Nco*I site, the natural sequence of the first two codons of *lacI* (GTGAAT) had to be changed to ccATGGAT (an *Nco*I site), which had the effect of changing the initiation codon to the more common ATG as well as changing the second amino acid from Asparagine (AAT) to Aspartic acid (GAT). We felt this change to be acceptable since comprehensive mutagenesis of *lac*I found the first two amino acids to be tolerant to mutagenesis [Markiewicz et al., 1994].

pET3a–His6TEV–LacI was verified by sequencing and transformed into BL21 and HMS174 (λ(DE3)pLysS) protein expression strains and trial inductions were conducted, however in both cases, LacI was found to be in the insoluble fraction (Fig 4.8). In these expression systems, protein induction (through T7 RNA polymerase induction) is under IPTG-inducible control and initially we feared that IPTG induction of LacI *itself* would lead to a negative feedback on *lacI* expression, however sequestration in inclusion bodies of host-provided LacI with insoluble His6–TEV–LacI would have had the opposite effect and this idea is supported by the presence of LacI in the absence of any IPTG induction (Fig 4.8).

Since His6–TEV–LacI in cellular extracts was insoluble the next strategy was to perform a denaturing extraction in 8 M urea and try to refold the protein by dialysis. Thus a sonicated inclusion body prep from the pellet of a 500 mL induced culture of BL21 pET3a–His6TEV–LacI was purified on a 5 mL nickel affinity column in the presence of 8 M urea (see Section 6.1.6 for protocols). However dialysis of elution fractions into PBS proved to be unsuccessful, with LacI precipitating at 1 M urea.

We obtained 5.5 mL of insoluble His6–TEV–LacI at 0.93 mg/mL (measured after re-solubilisation in urea) (Fig 4.9). Due to the dearth of commercially available LacI antibodies we decided to use this protein as an antigen to generate Rabbit polyserum. We employed the services of the Institute of Medical and Veterinary Sciences (IMVS) to immunise a rabbit with four innoculations of

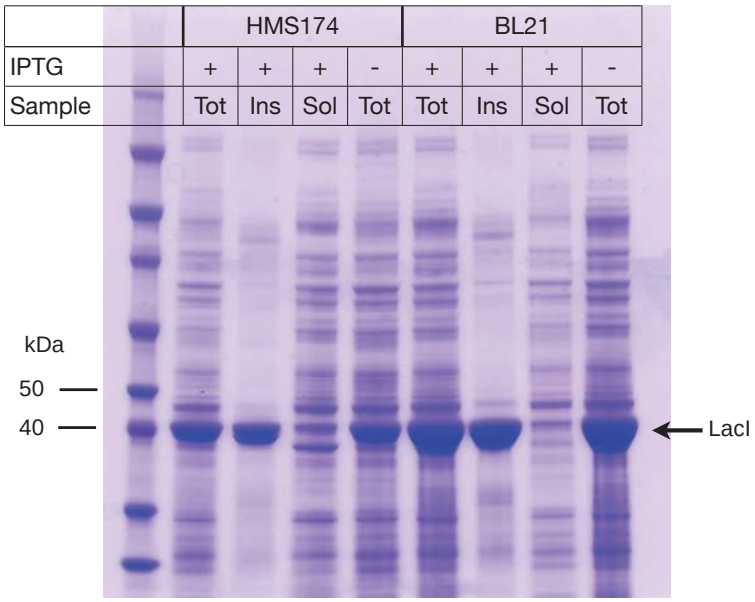


Figure 4.8: SDS PAGE gel of trial LacI inductions. pET3a–His6TEV–LacI was transformed into HMS174 or BL21 (λ(DE3)pLysS) and total extract (Tot), the insoluble pellet fraction (Ins) and the soluble supernatant fraction (Sol) were resolved. LacI is predominantly in the insoluble fraction, and His6TEV–LacI is expressed in the absence of IPTG.

~ 0.5–1 mg purified protein. Once we obtained the rabbit serum, we compared our antibody to a commercially available affinity purified polyclonal LacI antibody (Rockland), and unfortunately found our antibody had less sensitivity (Fig 4.10). Therefore for the relative quantitation of the Low (L1) vs High (L2) LacI levels in the long looping paper [Priest et al., 2014], we used the Rockland antibody (Chapter 2).

The existing protocol in our lab for preparing extracts of *E. coli* for Western Blot called for the resuspension of 1 mL mid-log LB culture in 50 μL lysis reagent (B-PER + benzonase), to which 4x gel loading buffer (GLB) was added and ~ 15-25 μL of the sample loaded onto an SDS-PAGE gel. However due to the poor sensitivity of our LacI antibodies, I experimented with increasing the concentration of *E. coli* extract in the protein sample, eventually lysing 4 mL of culture in the same volume. Note also that M9 Minimal Medium (M9MM) was used as the growth medium in order to reproduce growth conditions of the LacZ assay. Whilst optimising this procedure I found that the more concentrated samples were susceptible to becoming viscous (presumably due to undigested chromosomal DNA), and this was solved by (1) ensuring to freeze-thaw culture pellets (10 min, −80°C) and (2) incubating lysed samples at 4°C for 20

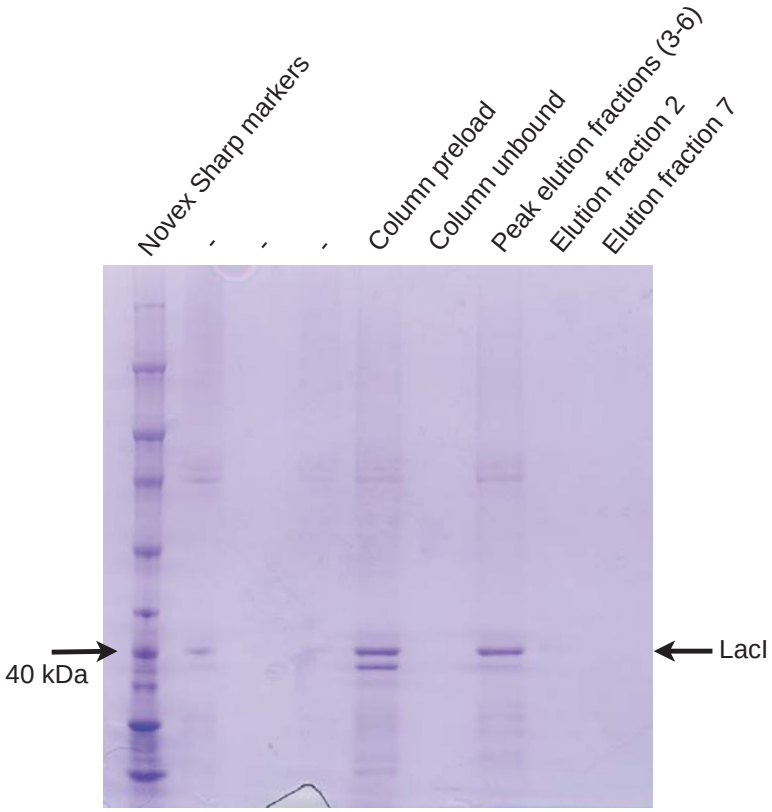


Figure 4.9: SDS PAGE gel showing purified LacI sample. LacI was purified under denaturing conditions on a nickel affinity column. ~ 1 mL elution fractions (at 500 mM imidazole) were taken. Peak fractions (3-6) were combined. Column preload shows sample prior to purification. Column unbound shows column flow through after loading.

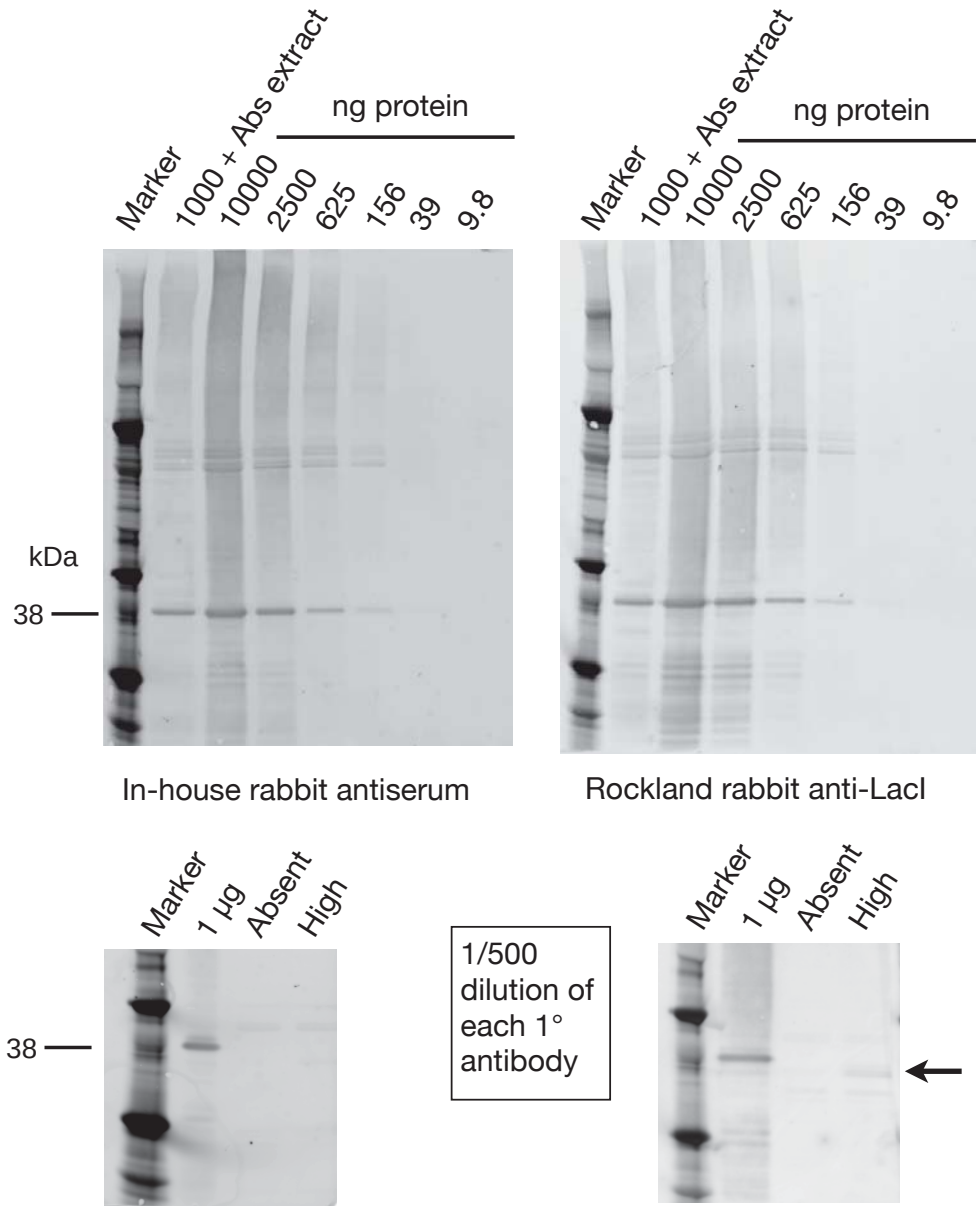


Figure 4.10: Comparison of in-house (left) and commercial (right) rabbit anti LacI antibodies. **Top**: A 1/4 dilution series of purified His6–TEV–LacI (into gel loading buffer) was run in duplicate and transferred onto PVDF membranes, which were probed with a 1/500 dilution of each antibody. The Rockland antibody appears to be slightly more sensitive (see 39 ng band). **Bottom** 1 µg of purified His6–TEV–LacI was run alongside extracts made from the LacI Absent (L0) and High (L2) strains. Detection of endogenous LacI appeared only possible with the Rockland antibody (arrow).

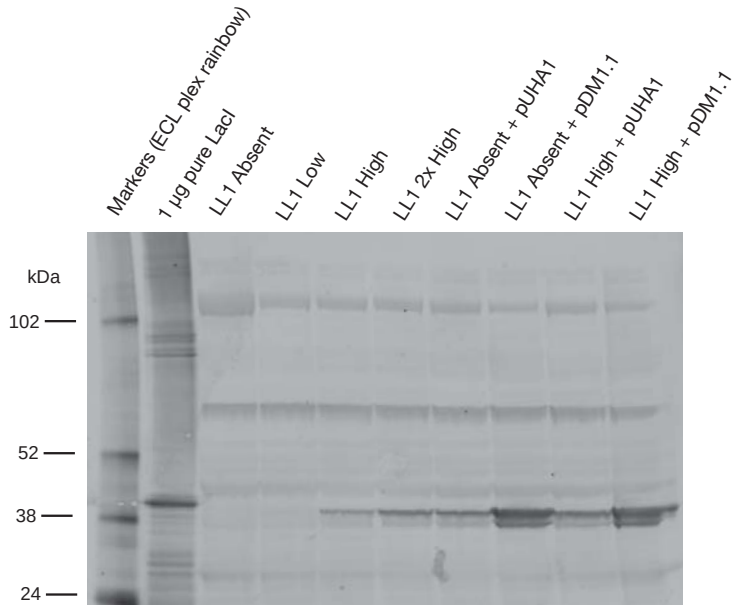


Figure 4.11: Western Blot of cellular extracts of LacLoop1 strains expressing different amounts of LacI using the Rockland primary antibody (1/200 dilution). The 2x High strain has an additional High LacI expression module integrated at another phage attachment site. pUHA1 and pDM1.1 are plasmids expressing LacI from P_{lacI} and the stronger $P_{\mathrm{lacI^q}}$ promoters respectively. Each band equates to $\sim$ 600 µL of an OD_{600} $\approx$ 0.4 M9MM culture. In Lane 2, $\sim$ 1 µg purified His6–TEV–LacI was run for comparison, and the band is slightly larger due to the presence of the tag.

minutes (instead of on ice) which may allow more efficient nucleic acid digestion by benzonase, whilst still cold enough to prevent proteolysis by host proteases. Figure 4.11 shows a Western blot where 20 µL of a 2x extract was run in each lane. Even here, LacI in the Low [LacI] (L1) strain is barely discernible, however the signal was strong enough to obtain a reasonable ratio between the L2 and L1 strains in subsequent Westerns (Chapter 2, Supplementary Figure S4A).

In summary, although attempts to measure the absolute level of LacI in the reporter strains were unsuccessful, we nevertheless obtained a measure of the approximate ratio of LacI levels in the low and high LacI strains, which was a useful constraint in the modelling (Chapters 2 and 3).

Chapter 5

Conclusions and Future Directions

5.1 The length dependence of DNA flexibility

The evolution of life into increasingly complex forms required a vehicle into which heritable genetic variation can be chemically encoded. The problem of information storage was solved by the advent of a polymeric molecule consisting of four different monomeric subunits, deoxyribunocleic acid (DNA). The cell houses DNA as well as the chemical machinery – consisting of other polymers (RNA and protein) – required to replicate and interpret the stored information. The genetic information we see today exists because it has been successful at maintaining and propagating itself through time within an environment rich in chemical building blocks and energy [Dawkins, 1976].

The success of DNA-based life attests to the stability and versatility of the DNA molecule. Fitness is maximised when an organism can maximise its energy availability whilst minimising energy wastage. Whereas it pays to be adaptable – such as having the ability to metabolise a range of sugar types – expressing these functions all at once would be disadvantageous. Hence life moved in the direction of increasing complexity, where genes for different functions can be switched on and off when required [Beckwith, 2011]. Cells therefore required a means to rapidly and reproducibly switch genes on and off in response to stimuli. Expression can be switched off through binding of a monomeric repressor protein over the promoter, however this results in a largely linear scaling of promoter activity to repressor concentration. More switch-like (sig-

moidal) responses are facilitated by cooperative binding of multimeric DNA binding proteins to multiple DNA binding sites. Cooperativity occurs when one binding event alters the free energy change of a subsequent event, with the prototypical example being binding of bacteriophage λ CI at the adjacent operator sites at *OR* [Ptashne, 2004]. However the DNA binding sites need not be adjacent, and it was eventually noticed that both bacteriophage λ and the *lac* operon possess multiple operators and that cooperative binding of the repressors to the operators results in DNA looping. Therefore, DNA loop-dependent repression is a versatile motif employed by evolution to yield robust, switch-like responses across a cellular population [Vilar and Saiz, 2005].

DNA looping is limited by the enthalpic and entropic realities of the DNA molecule itself. Whilst short-range DNA looping is limited by the resistance of DNA to torsional and bending forces, the probability of long-range looping is limited by the entropic cost imposed when two disparate sites are constrained within a small volume. As a result, increasing DNA loop size yields an optimal looping efficiency beyond which looping slowly declines. The DNA tether linking two looping sites facilitates loop formation because binding of a DNA looping protein at one site increases its local concentration at the other site. For LacI, which binds DNA as a pre-formed tetramer, if this concentration (J_{loop}) is greater than the total (cellular) LacI concentration, then DNA loop formation is favoured over the alternate possibility of a separate tetramer binding at each site. Therefore, another evolutionary benefit of DNA looping is that it allows greater repression at a lower cellular concentration of repressor protein, thereby achieving a similar result with fewer cellular resources.

As the distance between two looping sites is increased, J_{loop} decreases, and we (and others) hypothesised that the rate of this decline should depend not on the specifics of the DNA looping proteins, but rather the nature of the intervening DNA tether. We therefore set up an *in vivo* system in which to quantitate DNA looping at longer range (up to 10 kb) by measuring loop-dependent repression of a *lacZ* reporter gene by LacI and λCI [Priest et al., 2014] (Chapter 2). As expected, we found that J_{loop} declines at a similar rate for LacI and λCI and furthermore, we observed a faster decline for J_{loop} *in vitro* (using the technique of tethered particle motion (TPM)) than *in vivo*. We attribute much of this difference to the lack of DNA supercoiling in the TPM setup, however other factors, such as DNA bending proteins present in the cell, may also

contribute to a reduced persistence length of DNA *in vivo*.

Questions remaining include:

1. How does J_{loop} decay over longer range (10-500 kb)?

2. What factors contribute to the apparently reduced persistence length of DNA *in vivo*?

5.1.1 The decay in J_{loop} over longer range

The contribution of Cui Lun (a PhD student in the Shearwin Lab) to the manuscript of Chapter 2 [Priest et al., 2014] was to build and assay the λCI DNA looping reporters, which he extended to 10 kb using the pIT3 series of integrating plasmids. Despite many attempts, I was unable to generate LacI looping constructs larger than 5.6 kb and this may have been due to the peculiarities of the Lac system such as $P_{lac^{UV5}}$ having a higher unrepressed activity than the lambda promoter P_{RM}. Cui Lun's 10 kb spacer pIT3 plasmid is ~ 16.2 kb, which is at the upper size limit for obtaining successful chromosomal integrants. To further extend the spacers, recombineering can be used, which is popular technique for genome editing in *E. coli* [Yu et al., 2000]. To extend the LacI spacers to 10 kb recombineering could be employed to insert an extra spacer cassette into the existing spacer DNA of the 5.6 kb integrants. To extend spacers beyond 10 kb insertions of an upstream (or downstream) operator will be made into *E. coli* genomic regions – such as the borders between genes – where the insertion does not affect cellular viability. In this way loop lengths of ~ 20, 50 and 100 kb could be constructed to observe the decay of J_{loop} at very long range.

As mentioned in Priest et al. [2014] when J_{loop} is low (near the background concentration of LacI) looping will be inefficient. The LacI background concentration in the LacI low (L1) strain is already low (18 nM, ~ 11 tetramers in a 1-fL volume) [Garcia and Phillips, 2011], and therefore at very long range, where J_{loop} is small, achieving efficient Lac looping is limited by our ability to express very low LacI concentrations, lower than J_{loop}. The model fit for Lac looping at 5600 bp suggests *Oid–O2* are looped at least 30% of the time and therefore measuring LacI-mediated looping beyond 10 kb may be difficult. Conversely the λCI system appears to be looped over 50% of the time at 10 kb (Priest et al. [2014] Fig 4B), and therefore λCI can be used to measure J_{loop} over longer distances. Knowing how J_{loop} decays is not the only goal; we hope

to gain insights into eukaryotic enhancer function by generating efficient very-long DNA loops (see below, Section 5.2).

Two series of Lac looping (*Oid–O2*) constructs were made that had different spacer DNA sequences that surprisingly, showed different repression values [Priest et al., 2014]. Whilst this difference may have arisen from trivial sources such as transcription from cryptic promoters affecting the binding of looping proteins and/or rearrangements/deletions in one of the spacer series by recombination, it highlighted the possibility that looping probability may be affected by 'activity' on the intervening spacer DNA. Such activity could include DNA supercoiling introduced by promoters in the spacer or the binding of nucleoid associated proteins (NAPs) that alter DNA structure.

Supercoiling may potentiate long-range DNA looping by reducing the search space of the two looping sites (See Fig 1.3), and may be the prime reason why the *in vivo* J_{loop} values were higher than those measured *in vitro* via TPM. The bacterial nucleoid is maintained in a negatively supercoiled state by DNA gyrase, which facilitates melting of the DNA double helix for processes such as transcription [Dorman, 2006]. Therefore the effect of supercoiling on looping *in vivo* can be tested by treating the pre-existing looping reporter strains with gyrase inhibitors (e.g. novobicin). A transcribing RNA polymerase generates positive supercoiling ahead and negative supercoiling behind, and this supercoiling is relieved by topoisomerases [Gilbert and Allan, 2013; Lavelle, 2014]. The effect of transcription-induced supercoiling on loop formation could be investigated by placing a strong promoter within a DNA loop, yet ensuring to flank it with transcriptional terminators to minimize the effects arising from polymerase transcribing over looping operators, such as dislodgement of DNA looping proteins.

Eukaryotic DNA is wrapped around nucleosomes whereas the *E. coli* nucleoid is bound by NAPs such as HU and Integration Host Factor (IHF), many of which bend the DNA [Dillon and Dorman, 2010]. Whilst NAPs such as IHF have a strong effect on DNA looping at short range ($\lesssim$ 150 bp), it will be interesting to investigate whether NAPs have a significant effect on DNA looping at medium to long–range *in vivo*. Phillips and coworkers [Amit et al., 2011] using the NtrC/*glnA* system (see Section 5.2.3), found that binding of Tet repressor to tandem copies of its operator placed between the NtrC binding sites and the σ54-dependent *glnA* promoter, reduced

enhancer-promoter DNA looping because binding of the Tet repressor makes the DNA more 'rigid' [Amit et al., 2011]. In addition to testing this effect of Tet repressor at longer range, copies of the consensus sequences (or preferred sequences) for NAPs [Dillon and Dorman, 2010] could be placed between *lac* operators to measure their effect on looping. Furthermore, an unbiased screen for effects of the spacer DNA sequence could be performed by inserting a random library of *E. coli* DNA (flanked by terminators) into the spacer. Ian Dodd has suggested a genetic screen where we place an antibiotic resistance gene under loop-dependent repression, and screen for sequences that lose DNA looping, i.e. those that facilitate growth in the presence of the antibiotic.

5.2 The loop domain model at long range in bacteria

A key question in the field of eukaryotic gene regulation is what mechanisms make enhancer-promoter (EP) contacts both efficient and specific? The loop domain model proposes that specific DNA loops formed by other proteins (insulators or architectural proteins) can shape the cis-regulatory landscape, favouring some EP contacts over others. In this study, we have provided the first quantitative, *in vivo* test of this model by measuring the magnitudes of DNA looping interference and assistance and showing how they combine together to create a 'specificity change' (S) imposed by an insulating DNA loop (Chapter 3) [Priest et. al. 2014b]. In measuring these fundamental aspects of DNA behaviour our data provides strong support for the loop domain model, however it cannot yet be concluded that looping interference and assistance are driving forces behind EP contacts in eukaryotic nuclei. A more definitive study of the loop domain model would test the effects of artificially-forced DNA looping on a native eukaryotic locus and Section 5.3 (below) proposes experiments to that aim.

In the meantime however, *E. coli* will continue to serve as a useful model to investigate DNA looping and the loop domain model. The tools developed in this thesis can be directly applied to immediate goals, such as testing new DNA looping proteins (Section 5.3) and questions, such as whether an interfering loop can also have an assisting effect (Section 5.2.1). An eventual aim is to generate an experimental platform in *E. coli* more akin to eukaryotic long-range gene regulation. In pursuit of this aim the salient features of enhancers could be recapitulated in *E. coli* including

very long range (~ 100 kb) DNA looping (Section 5.2.2) and bacterial enhancer-like elements (Section 5.2.3). These tools would allow investigation of the effect of long-range assisting and interfering loops on bacterial enhancer-promoter connections.

5.2.1 Can an interfering loop also have an assisting effect?

For Eukaryotic long range gene regulation, large loops formed by 'insulators' such as CTCF may serve to bring enhancers and promoters into close nuclear proximity [Ong and Corces, 2014]. Increased contact probability afforded by this distance reduction may outweigh any looping interference effects resulting from placement of enhancer and promoter in different DNA loops.

In the manuscript in Chapter 3 an alternating arrangement of operators for the LacI and λCI repressors was used to quantify DNA looping interference *in vivo* [Priest et. al. 2014b]. Interestingly, whilst an equidistant (symmetrical) arrangement of the operators (Manuscript Figure 3, Construct 2) gave ~ 3.7 fold interference, an asymmetrical arrangement (Figure 3, Construct 5) yielded ~ 10 fold interference. In the asymmetrical construct the λCI loop was larger (1800 vs 1200 bp), however the observed λCI-only looping frequency was similar for both constructs (0.89 and 0.85 respectively). Although the error bars for the fitted interference values overlapped, this potential difference highlights the possibility that the positioning of the 'sequestered' operator may affect the level of interference.

Consider the scenario presented in Fig 5.1. Two alternating loops of the same size can be positioned at three extremes with respect to one another; either with a short internal spacer, a long internal spacer or half-way between the two (Fig 5.1 A–C respectively). Importantly, if the internal spacer is long, then formation of one loop brings the operators for the other loop into closer proximity, which may facilitate their interaction. Whilst the interfering loop is still expected to inhibit alternate loop formation via stopping the propagation of supercoiling, it may also assist the other loop by shortening the DNA tether between its operators. Therefore a key question is how the magnitude of this potential assistance effect compares with the interference yielded by the interfering loop. Testing this will be a simple matter of building and assaying constructs similar to those shown in Figure 5.1. If the long internal spacer construct does indeed show less interference (or even assistance), this would suggest that enhancer blocking may not

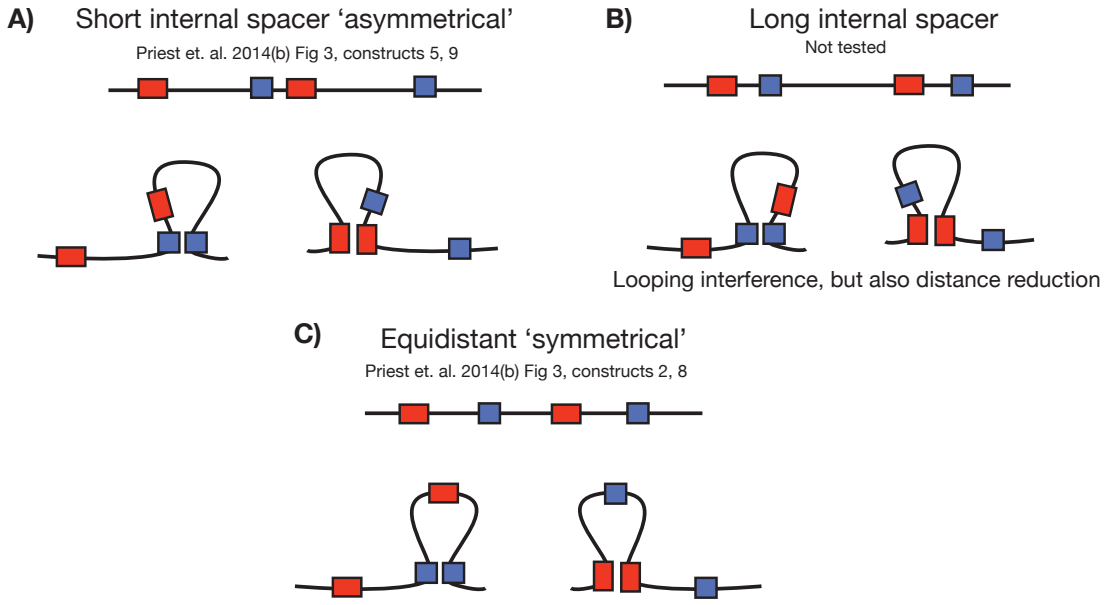


Figure 5.1: Does changing the relative position of loops affect loop interference? Three extremes of positioning between interfering loops are shown. (A) asymmetrical with a short internal spacer, (B) asymmetrical with a long internal spacer and (C) symmetrical arrangement. Note that the sizes of each individual loop doesn't change, only their position relative to one another. If the internal spacer is long, then formation of one loop has the effect of bringing operators for the other loop in closer proximity, which may affect their looping frequency.

only depend on segregation of enhancers and promoters in separate DNA loops, but also the relative location of the interfering loop. The research proposal outlined below (Section 5.3) will provide a means to test this directly at a mammalian developmental locus.

5.2.2 Very long range looping

Looping by LacI or λCI is expected to decay to low levels as loop sizes are extended to 20–50 kb and beyond (Section 5.1.1), however given enhancers and promoters often interact over larger distances (a recent estimate of median EP distance was 124 kb [Jin et al., 2013]), investigating mechanisms that drive efficient very long range looping (10–500 kb) is a high priority. By nesting operators for LacI and λCI we observed an ~ 2–3 fold assistance effect at ~ 1500 bp [Priest et. al. 2014b]. Therefore in a similar strategy as described above (Section 5.1.1), to generate strong, very-long DNA loops, recombineering could be used to insert a module containing operators for both λCI and LacI (*OL* and *Oid*) far from the *OR*.300bp. $P_{lac^{UV5}}$.*O2 lacZ* reporter.

DNA looping assistance is a form of positive cooperativity since formation of one loop imposes constraints on the DNA tether that increases the probability that the second loop will form. Adding a third DNA looping element is predicted to yield more looping assistance, but a question is to what degree does a third DNA loop affect looping assistance? Is there synergy in looping assistance as the number of looping elements is increased? Furthermore, if looping between sets of three (or more) compatible elements is greatly enhanced, then looping specificity could be dictated by a 'combinatorial code' of interactions between compatible clusters of insulators, as was recently suggested [Ghirlando et al., 2012]. Indeed *D. melanogaster* has a number of different insulator types, which are often found clustered in the genome as 'aligned insulators' [Van Bortle et al., 2012].

Consider the scenario presented in Figure 5.2. One could construct a triple loop assistance system by combining LacI and λCI looping with a third looping protein, such as the CI repressor from bacteriophage 186 [Wang et al., 2013]. Pairwise loop assistance in the presence of operators (or looping proteins) for only two of the looping elements could then be compared with the loop assistance when all three looping elements are present. Extending the statistical mechanical loop interaction model to a triple loop system should be feasible, thereby allowing us to extract

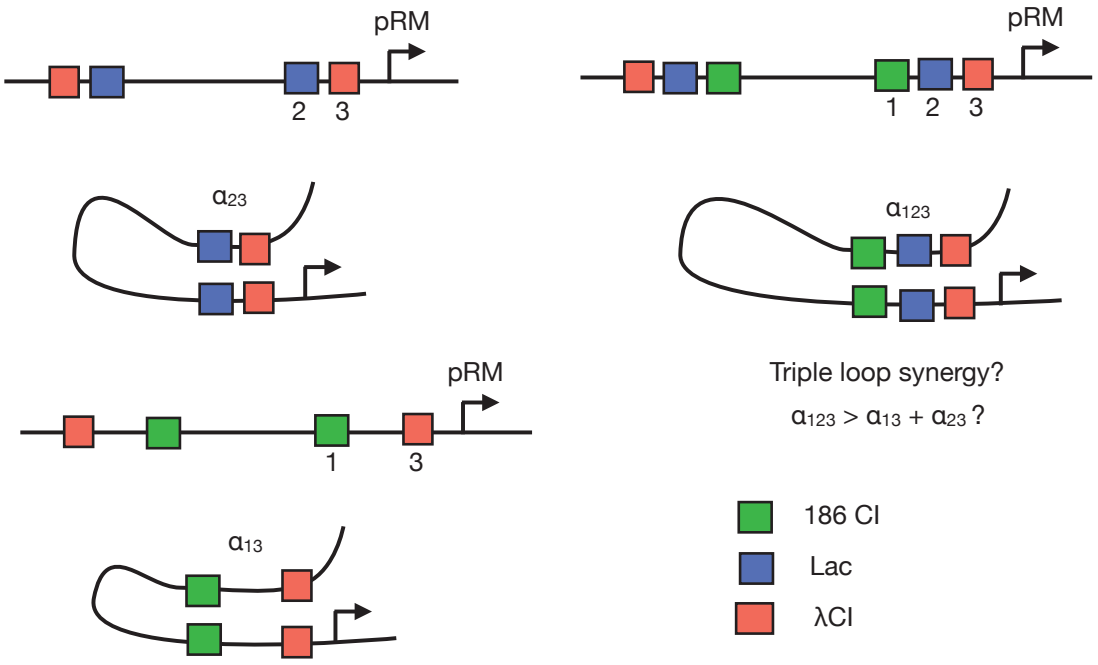


Figure 5.2: Do loop assistance factors combine synergistically? In a triple loop assistance system, pairwise loop assistance factors can be measured (left) and compared to the assistance garnered from the triple loop system (right). Such multi-loop systems may drive specificity of looping in eukaryotes.

the triple loop assistance factor (α_{123}). From this analysis it will be evident whether triple loop assistance is greater than the sum of the pairwise assistance factors.

5.2.3 Putting it all together: Bacterial enhancers

As more mechanistic studies of enhancers are published, a picture is emerging of a myriad of molecular mechanisms by which enhancers can activate and silence gene expression [Shlyueva et al., 2014]. For example, a recent study identified a class of distal regulatory elements, termed anti-pause enhancers (A-PEs) that control gene expression by regulating release of promoter-proximal paused RNA polymerase II [Liu et al., 2013]. A general theme is that enhancers provide binding sites for various transcription factors and that EP connection generates a higher-order complex upon which molecular transactions (such as histone modifications and the recruitment and eviction of co-factors) take place to either inhibit or activate productive elongation by RNA polymerase. Some enhancers (such as A-PEs) may form a silencing pre-complex with

their target promoters with the whole structure serving as a molecular switch for transcription, whereas other enhancers may undergo repeated rounds of promoter contact, each time depositing activating stimuli and/or factors to the promoter. Therefore, not all enhancers are alike; they may show different dependencies on DNA looping, and some enhancers may not require DNA looping at all, instead functioning through 'tracking' models [Zhu et al., 2007] or even '*trans*'-acting non-coding RNAs [Orom and Shiekhattar, 2013]. Nevertheless, biochemical mechanisms employed by enhancers such as recruitment and allostery can be recapitulated in bacteria, where their role in looping interference and assistance can be easily assayed, as will be discussed below.

There are currently two enhancer-like gene regulatory systems in *E. coli*; the σ54-dependent promoters [Bush and Dixon, 2012] and distant elements capable of recruiting RNAP via interactions with the C-terminal domain (CTD) of its α subunit [Cui et al., 2013]. The σ54-dependent promoters are bound by an RNAP complex containing an alternate σ subunit (σ54, rather than the usual σ70), which binds as an inactive, closed complex. Conversion of the closed complex into the transcriptionally productive open complex requires remodelling of σ54 by ATP-dependent bacterial enhancer binding proteins (bEBPs) bound to upstream activator sequences (UASs). In their natural context, UASs are usually located 80–150 bp upstream of their target promoters, however Phillips and colleagues recently constructed a synthetic bacterial enhancer by placing the binding sites for bEBP NtrC up to 315 bp upstream of the σ54-dependent *glnAp2* promoter [Amit et al., 2011].

The CTDs of the two α subunits of RNAP are connected to their N-termini (and the main body of the enzyme) via a flexible linker sequence. The α-CTD helps recruitment of RNAP to promoters by interacting with transcriptional activators, such as catabolite activator protein (CAP) bound upstream of the promoter, or through direct contact with DNA elements (termed UP elements) located just upstream of the promoter. In a λ prophage, *c*I expression is subject to both negative and positive autoregulation via CI repression and activation of its own transcription from P_{RM}. Whilst looping between *OL* and *OR* facilitates repression by allowing tetramer formation between *OL3* and *OR3*, intriguingly, ablation of repression by mutation of *OL3* revealed an activatory effect of the DNA loop [Anderson and Yang, 2008]. Whilst it was shown that some of this activation is mediated by a CI dimer bound at *OL3*, it was recognised that there

is an UP element next to *OL*, which was suggested to be brought into close proximity to P_{RM} via DNA looping to subsequently aid RNAP recruitment [Anderson and Yang, 2008]. A study in the Shearwin lab [Cui et al., 2013] showed this to be the case, and furthermore demonstrated robust promoter enhancement by a module situated 10 kb away from *OR* and consisting of *OL* paired with an UP element. Therefore, DNA loop-mediated juxtaposition of α-CTD binding elements to promoters serves as a second, distinct bacterial model enhancer that functions via a recruitment rather than a allosteric mechanism. The σ54- and UP element-dependent enhancers may display different dependencies on looping since the σ54-dependent enhancer has a catalytic mechanism and may only require transient EP contacts, whereas contact between the α-CTD and CAP or an UP element may require more long-lived looping.

To generate in *E. coli* an enhancer-promoter system more relevant to eukaryotes, the distance in these two enhancer systems could be extended to 20–100 kb and looping assistance could be used to bring together the enhancer and promoter (Fig 5.3A). Once these long-range enhancer systems have been established in *E. coli* one could introduce long-range alternating loops and assay for loop interference to provide a test of the loop domain model at very long range (Fig 5.3B). Furthermore, by changing the location of the insulating loop relative to the EP loop, the suggested proximity reduction effect of an insulating loop (Section 5.2.1) could be tested at very long range.

The experiments outlined here show the utility of *E. coli* as a tractable model organism for examining fundamental biophysical mechanisms that apply to all living organisms. Furthermore, the fundamental nature of this work will no doubt contribute to continued efforts to develop *E. coli* (and other organisms) as platforms for synthetic biology applications (for review see Church et al. [2014]).

5.3 Research Proposal: Testing the loop domain model at the human β-globin locus

The β-globin locus is a paradigm for studying the relationship between chromatin looping and transcription. The human and mouse β-globin loci encode different β-like globin genes that are

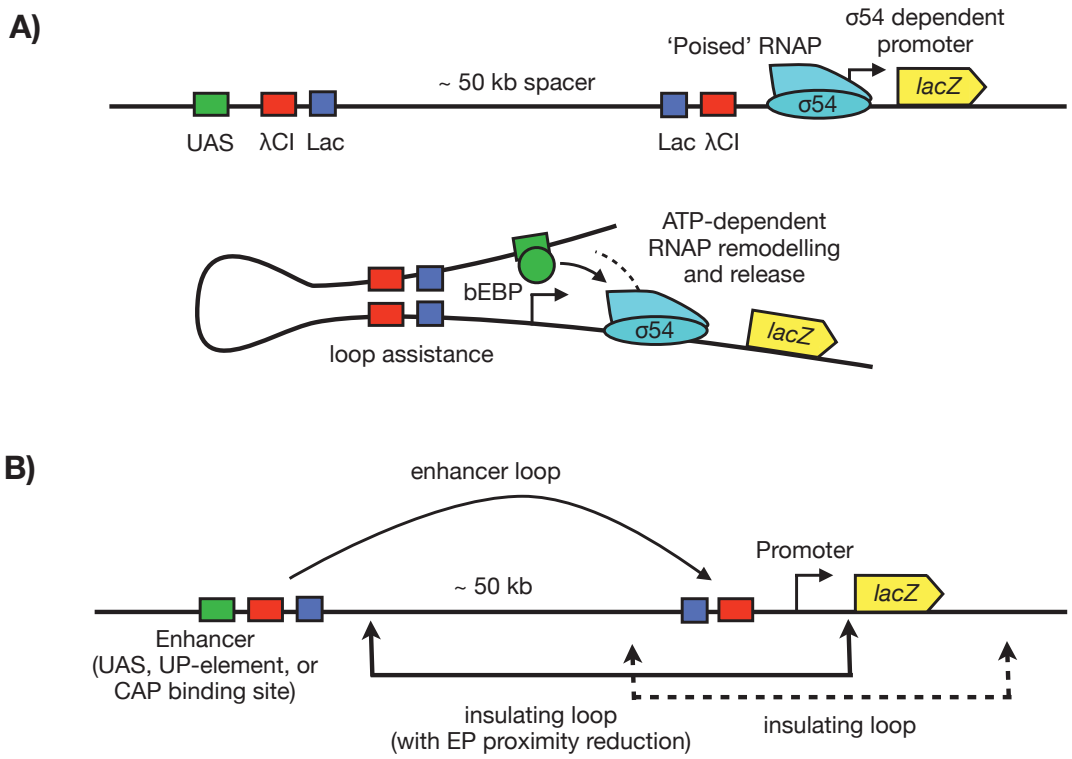


Figure 5.3: Testing the loop domain model at very long range with bacterial enhancers. **A.** A synthetic long-range bacterial enhancer system. A looping assistance module (consisting of a double or triple loop) brings an upstream activator sequence (UAS) in proximity to a σ54-dependent promoter, facilitating transcriptional activation of RNAP by bacterial enhancer binding proteins (bEBPs). **B.** Long-range looping interference. An insulating loop that does not alter EP proximity (dashed line arrows), may yield more interference than the other loop (solid line arrows) that, whilst being an interfering loop, nonetheless brings the enhancer and promoter into closer proximity.

activated during erythropoiesis in a developmental stage-specific manner (Fig 5.4). During erythropoiesis, this locus adopts a looped conformation where two flanking DNAseI hypersensitive sites (HSs) (known as HS5 and 3'HS1 respectively) loop to one another as well as the strong upstream enhancer, the locus control region (LCR) which consists of a further four HSs (HS1–4). This looped structure – termed an 'active chromatin hub' (ACH) [Tolhuis et al., 2002] – is absent in non-expressing, non-erythroid cells, and is a prerequisite for globin gene activation [Palstra et al., 2003]. Studies of β-globin locus architecture via 3C have shown that activation of a particular globin gene is accompanied by DNA looping of the gene to specific LCR HSs within the ACH, with inactive genes generally 'looped out' and not in contact (reviewed in Kim and Dean [2012]). During development, β-globin-like isoform expression shifts from the fetal (γ-globin) genes to adult (β-globin) genes. In patients harbouring abnormal adult sickle β-globin, reactivation of γ-globin expression could be a successful strategy to treat sickle cell disease.

Knockdown studies have revealed contributions of hematopoietic transcription factors and co-factors (such as GATA-1, EKLF, LMO2, NLI/Ldb1 and TAL1) to activation of globin gene expression. For example, among the functions of GATA-1 is recruitment of the histone acetylase CBP/p300. The resulting histone hyperacetylation leads to more open chromatin, which contributes to HS formation by facilitating recruitment of nucleosome remodelling factors. Subsequently, complexes of transcription factors and co-factors bind to LCR HSs and bridge contacts to globin promoters. For example, in the human erythroblast cell line K562, the fetal γ-globin genes are in contact with the LCR and show relatively high expression, whereas the embryonic and adult genes are not expressed [Woon Kim et al., 2011]. The TAL1 protein is a basic helix-loop-helix DNA binding transcription factor essential for erythropoiesis, and data suggests it forms part of the DNA binding component of a complex consisting of GATA-1, E2A, LMO2 and NLI/Ldb1 [Wadman et al., 1997]. Recently, Kim and colleagues [Yun et al., 2014] showed that knockdown of TAL1 in K562 cells resulted in loss of contact between the LCR and γ-globin genes and reduced γ-globin expression, yet did not affect ACH and HS formation or chromatin modifications (such as H3K27ac and H3K4me2). Therefore in K562 cells TAL1 is essential for looping between γ-globin genes and the LCR, but its absence seems to have little effect on the epigenetic structure of the locus and may therefore not preclude the ability of the

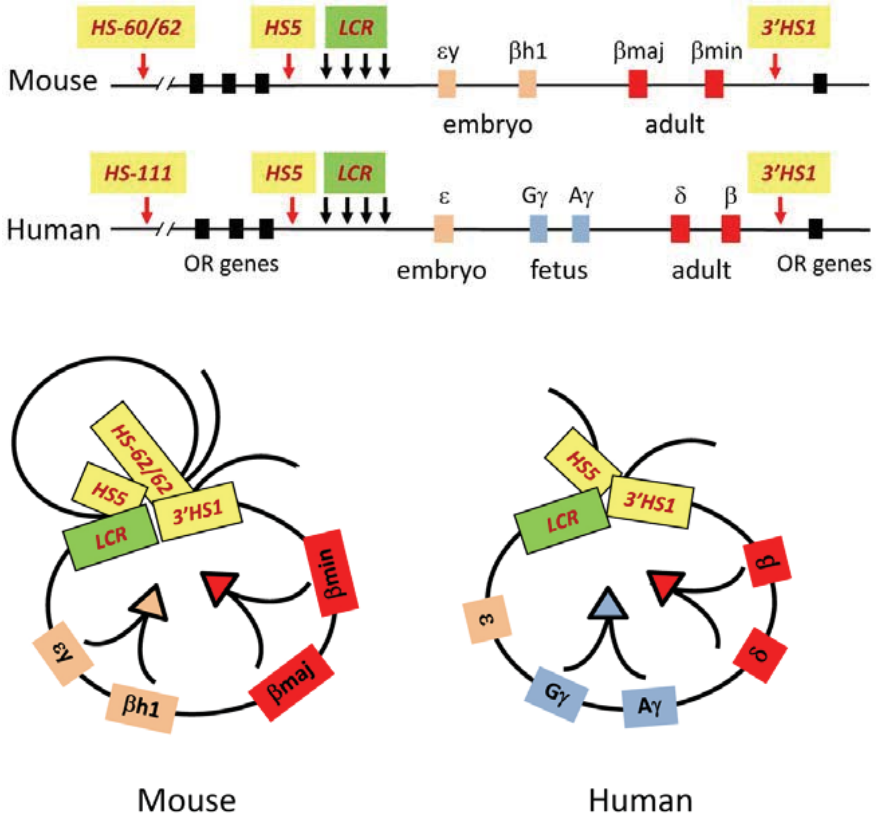


Figure 5.4: Figures 1 and 2 from Kim and Dean [2012] showing the linear and proposed looped arrangement of the mouse and human β-globin loci. Chromatin loop formation between the flanking hypersensitive regions and LCR forms and active chromatin hub (ACH), and globin gene expression is dependent upon contact with the LCR. During development β-like globin expression in differentiating erythrocytes switches from the embryonic genes (near the LCR) to the fetal (in humans) and finally the adult genes.

LCR to enhance globin gene expression.

The K562 cell line therefore provides an ideal model system in which to study effects of chromatin architecture on gene expression since γ-globin expression is directly related to its contact with the LCR and this contact can be modulated through TAL1 knockdown. Previously at the β-globin locus, DNA loop-mediated enhancer blocking [Hou et al., 2008], and transcription factor-mediated DNA loop assistance [Deng et al., 2012] have been suggested (see Section 1.2), however in both cases, the factors used for enhancer blocking or LCR tethering (CTCF and Ldb1 respectively), are naturally present at the locus and therefore their experimental manipulation could potentially contribute to unexpected, pleiotropic effects. A better test of loop interference/assistance at the β-globin locus would employ an ectopic DNA looping protein where any observed effects are likely to result from induced changes in chromatin architecture.

By generating strong, targeted DNA loops at the β-globin locus, two hypotheses could be tested in the K562 cell system:

1. **DNA looping interference**. Whether a DNA loop formed around the γ-globin genes interferes with the loop formed between these genes and the LCR, hence reducing γ-globin expression (Fig 5.5(1))

2. **DNA looping assistance**. Whether after TAL1 knockdown a DNA loop that tethers γ-globin gene promoters to specific LCR HSs can restore γ-globin expression (Fig 5.5(2)).

To test these hypotheses, a system is required to generate targetable DNA loops in mammalian cell culture. One avenue could be to use genome editing tools (see below) to introduce into the locus operators for a well-characterised DNA looping protein (e.g. λCI), however changes to the DNA sequence may have unexpected effects, and furthermore there is little data on the use of prokaryotic DNA looping proteins in mammalian cell culture [Nolis et al., 2009]. Recent development of the CRISPR/Cas9 system (reviewed in [Sander and Joung, 2014]) provides the means to probe the effects of enforced DNA looping at the β-globin locus via development of orthogonal, DNA sequence-targetable DNA looping proteins. The CRISPR/Cas9 system is a bacterial immunity mechanism that degrades foreign, invasive DNA (e.g. from infecting bacteriophage). Segments of foreign DNA are integrated at specific bacterial loci and then transcribed into an

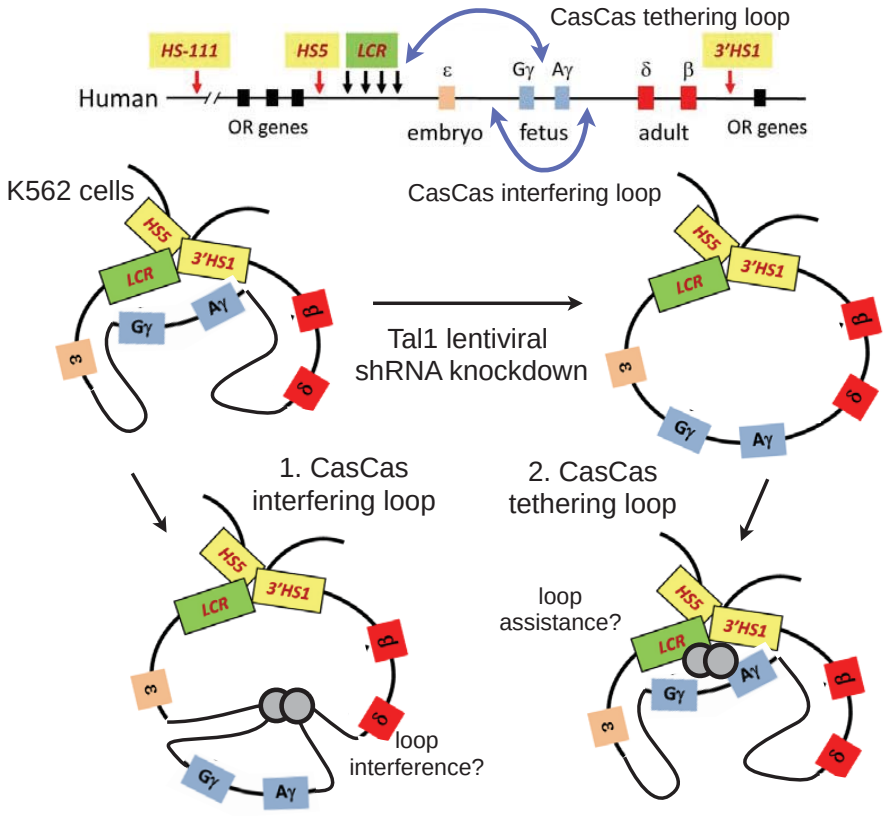


Figure 5.5: Testing the loop domain model at the β-globin locus using CasCas targetable DNA looping proteins. (1) In K562 cells, the ability of an interfering loop to disrupt γ-globin–LCR contact will be tested. (2) When TAL1 is knocked down, γ-globin–LCR contact is lost, yet an assisting loop may restore this contact and restore γ-globin expression. Figure adapted from Kim and Dean [2012].

RNA species that forms a complex with the Cas9 nuclease directing it to digest invading DNA through direct RNA-DNA basepairing. Remarkably, the CRISPR/Cas9 system has been engineered to function as an RNA-guided DNA nuclease in mammalian cells therefore facilitating genome editing applications [Mali et al., 2013]. Furthermore, fusion of various proteins (such as transcriptional co-factors and fluorescent proteins) to a catalytically inactive Cas9 (dCas9) has allowed the CRISPR-Cas9 system to be repurposed for transcriptional regulation [Gilbert et al., 2013; Qi et al., 2013], and chromatin imaging [Chen et al., 2013]. Because since dCas9 longer cuts the DNA, it instead functions as a sequence-targetable DNA binding complex.

Different bacterial species encode different Cas9 proteins and importantly, George Church and colleagues have engineered these to develop a set of orthogonal dCas9 proteins allowing

simultaneous targeting of two or more Cas9 proteins to different DNA sites within the same cell [Esvelt et al., 2013]. Ian Dodd, Keith Shearwin and I therefore propose to develop a synthetic, targetable DNA looping protein by creating a translational fusion between two orthogonal dCas9 proteins (here called 'CasCas'). Co-expression of the small-guide RNAs (sgRNAs) for each dCas9 protein would theoretically allow targeted DNA looping between virtually any two sequences. Generating a functional fusion protein however is a non-trivial exercise, and optimisation would be greatly aided by a simple system to assay DNA looping. Traditionally DNA looping is measured by loop-dependent repression of a promoter situated at one of the looping sites, however our DNA looping interference reporter chassis (Figure 4.2) can circumvent this need because looping can be measured in proxy by looping interference on a well-characterised looping protein. Therefore new DNA looping proteins (such as CasCas) could be optimised by placing their binding sequences either side of the distal *lac* operator and measuring interference on LacI looping. Looping efficiency of new DNA looping proteins could be quantified by calibrating their level of looping interference to that of λCI.

After developing a DNA looping CasCas protein in *E. coli* the system could be transferred into the K562 cell line. Conveniently, systems for expressing Cas9 and sgRNAs in K562 cells for genome editing have been successfully developed [Mali et al., 2013], and would provide a useful guide for optimisation of CasCas function in this cell line. Further optimisation of CasCas in K562 cells would be achieved by assaying its ability to generate a strong 3C signal for targeted DNA loops of various sizes. Once CasCas is optimised, its effect on the β-globin locus could be tested using the techniques of Yun et al. [2014] for TAL1 knockdown, 3C assays, and monitoring globin gene expression in K562 cells. If CasCas can drive a sufficiently strong DNA loop around the γ-globin genes then the loop domain model would predict that contact with the LCR would be lost and γ-globin expression would decrease (hypothesis 1, Fig 5.5). An advantage of CasCas is that DNA loops could be targeted at will simply by transfecting different sgRNAs (or their expression cassettes), and therefore it would be simple to test the effect of various sized insulating loops sequestering regions as small as just one γ-globin promoter, to the entire $\sim$ 50 kb genomic region harbouring the γ-globin genes without having to generate a new DNA looping protein each time. For hypothesis 2, one could assay the effect of tethering

different parts of the γ-globin promoter(s) to different regions of the LCR. For instance, it may be observed that certain loop conformations – although separating the enhancers into different DNA loops – may in fact yield looping assistance because the relevant interacting regions are brought into closer nuclear proximity (see Section 5.2.1).

The experiments outlined in this research proposal will provide a definitive test of the loop domain model in mammalian cells, where long-range gene regulation is widespread. Nevertheless, such experiments benefit greatly from the solid practical and theoretical foundation outlined in this thesis. Specifically, the thermodynamic and statistical mechanical models we have developed are equally applicable to measuring looping fractions (Chapter 2) and the specificity change factor (S) (Chapter 3) in eukaryotic systems. Practically, the DNA looping reporter chassis provides a means to test and develop novel DNA looping proteins (such as CasCas), as well as providing the experimental framework to perform similar quantitative measurements of looping assistance and interference in eukaryotic models such as *D. melanogaster*.

Chapter 6

Materials and Methods

6.1 Methods

6.1.1 Bacterial Procedures

Storage of Bacterial Stocks

Bacterial colonies were maintained in the short term on the appropriate agar plates at 4°C. For long term storage, 500 μL culture was added to 500 μL 80% glycerol in sterile, screw cap 1.5 mL microcentrifuge tubes and tubes were stored at −80°C.

Growth of Bacterial Strains

All cultures were routinely grown at 37°C, unless otherwise specified. Stationary phase cultures were prepared by innoculating broth from a single colony on a fresh streak plate and incubating overnight with aeration (i.e. on a rotating drum or a shaking table) at 37°C. Log-phase cultures were prepared by diluting a fresh stationary phase culture or adding a colony of bacteria into sterile broth and incubating with aeration at 37°C, until the required cell density was reached. Cell density was measured by observing the optical density at 600 nm of the culture (OD_{600}) using an Ultrospec 10 cell density meter (Amersham Biosciences). Alternatively, the A_{620} of cultures was measured in microtitre plates in a plate reader, and converted back to A_{600} through and empirically determined relationship.

Preparation and transformation of TSS competent cells

Chemically competent *E. coli* cells are competent enough to take up plasmid DNA and often suitable for transforming with assembled (or ligated) pIT plasmids for chromosomal integration. Chemically competent cells were prepared using the Transformation and Storage Solution (TSS) method [Chung et al., 1989]. TSS is LB broth with 10% (wt/vol) PEG (MW 3350 or 8000), 5% (vol/vol) DMSO and 20–50 mM Mg^{2+} at a final pH of 6.5 filter sterilised through a 0.45 micron filter and stored at −20°C until needed. TSS competent cells are prepared by resuspending chilled early log phase cultures ($OD_{600} \approx 0.4$) in 1/10 culture volume TSS and stored at −80°C as 1 mL aliquots.

When cells are required for transformation, aliquots of TSS cells are thawed on ice and then 100 μL of cells are added to ~ 1–5 μL DNA in an eppendorf tube, which is incubated on ice for 15 min. In a modification to the original protocol of [Chung et al., 1989], cells are then heat-shocked at 42°C for 1 min immediately after which ~ 1 mL LB is added and cells are allowed to recover for ~ 1 hour at 37°C. Recovered cells are then pelleted (4 min, 6000 g), resuspended in 100 μL of the remaining supernatant and spread onto freshly-prepared agar plates.

Preparation and transformation of electro-competent cells

For applications where a higher competency was required, such as transforming large ($\gtrsim$ 9 kb) assembled pIT plasmids, electro-competent *E. coli* cells were prepared. A stationary phase culture was diluted 1/250 in the appropriate broth and incubated with shaking until early-mid log phase ($OD_{600} \approx$ 0.4–0.6) and then chilled on an ice-water slurry. Cells are then washed twice in ice cold H_2O and twice in 10 % glycerol, reducing the resuspension volume as necessary to fit the desired centrifuge, and then resuspended in 1/100–1/200 culture volume ice cold 10% glycerol, and divided into ~ 130 μL aliquots for storage at −80°C. Numerous washes are required to remove the salt from the growth media, which would otherwise cause arcing in the electroporation apparatus, and since cells are maintained in H_2O (hypoosmotic), care must be taken to keep cells cold.

DNA for electroporation must be free of salt and was cleaned up with a suitable kit (PCR cleanup or Gel extraction). The following are prepared on ice: (1) thawed electro-competent

cells, (2) 1–3 μL DNA in an eppendorf tube and (3) an electroporation cuvette. Using chilled pipette tips ~ 40 μL of cells are mixed with the DNA and added to the cuvette, which is then placed in the electroporation apparatus and the current applied (BioRad Micropulser, EC1 setting). Immediately, cells are flushed from the cuvette with ~ 1 mL LB back into the eppendorf tube that had the DNA, and allowed to recover (37°C, 1 hour), after which they are plated as for TSS cells.

6.1.2 DNA manipulation

DNA purification kits

All Kits were used according to the manufacturer's directions. Kits used included:

- NucleoBond Xtra Midi Plus, Macherey-Nagel
- QIAprep Spin Plasmid Miniprep Kit, Qiagen
- Zymoclean Gel DNA Recovery Kit, Zymo Research
- DNA Clean & Concentrator-5 Kit, Zymo Research

Preparation of plasmid DNA

Routine preparation of plasmid DNA was done using the QIAprep Spin Plasmid Miniprep Kit (Qiagen), according to the manufacturers instructions. DNA was eluted using 40 μL H_2O and stored at −20°C.

Large-scale plasmid purification

When larger amounts of plasmid DNA were required, the NucleoBond Xtra Midi Plus (Macherey-Nagel) plasmid midiprep kit was used, according to the manufacturer's instructions. Plasmid DNA eluates were concentrated with the NucleoBond Finalizer and the final elution volume was 400 μL.

Analysis of DNA size, purity and concentration

DNA samples were assessed for size, purity and concentration using agarose gel electrophoresis with a Mini Sub Cell GT (BioRad) electrophoresis device at 110V. Agarose was mixed with 1X TAE and melted in a microwave oven and then cast into minigels ranging from 1% – 2.5% agarose. Prior to loading into the wells, DNA samples were mixed with EZ-Vision 6x DNA Dye as Loading Buffer (Amresco). Gels were visualized using a UV transilluminator coupled to a CCD camera (ChemiDoc, Biorad). Comparison of DNA bands with 2-Log DNA molecular markers (New England Biolabs) allowed estimation of DNA size and concentration. Accurate measurement of DNA concentration was performed using a NanoDrop 2000 spectrophotometer (Thermo Scientific).

Restriction enzyme digestion

Restriction digests of DNA were performed using restriction enzymes from New England Biolabs according to the manufacturer's directions. Reaction volumes ranged from 5–50 μL, reaction times ranged from 1 hour to overnight and reaction conditions and incubation temperatures were as specified by the manufacturer.

Gel extraction of DNA fragments

Digested DNA fragments or plasmids required for cloning were run on a large-well (~ 50 μL) agarose gel at 90V. The gel was then immersed in SYBR Safe DNA gel stain (Invitrogen) for 20 minutes and DNA bands were visualized using a Safe Imager bright light transilluminator (Invitrogen). Bands were excised from gels using a scalpel blade, and DNA was purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research) according to the manufacturers instructions, with a 6–15 μL elution volume. 1 μL of the gel extraction was checked for purity and concentration on an agarose gel or NanoDrop spectrophotometer.

Treating digestions with Antarctic Phosphatase (Sapping)

Linearized vectors with compatible ends were treated with Antarctic Phosphatase (New England Biolabs), which catalyzes the removal of 5' phosphate groups from DNA thereby preventing self

religation. Treatment with Antarctic Phosphatase thus reduced the background of colonies on transformation plates resulting from religated, singly-cut vector without the desired insert. 5 units of Antarctic Phosphatase were added to restriction digests made up to 1x Antarctic Phosphatase buffer (New England Biolabs) and reactions were incubated for 15 minutes at 37° C and then heat inactivated at 65°C for 5 minutes or cleaned up with the appropriate kit.

Ligating restricted DNA fragments into restricted vectors

Ligations were performed in 10 μL reaction volumes, containing 0.5–2 units T4 DNA ligase, 1x ligase buffer (New England Biolabs) and an approximate 3:1 molar ratio of insert to vector DNA, with the concentration of vector DNA in the range 10–50 ng. A control reaction containing vector but lacking insert DNA was prepared identically to determine the number of background colonies resulting from the vector DNA alone. Following incubation for 10–15 minutes at room temperature, 5 μL of the ligation reaction was transformed into TSS competent cells.

Gibson Assembly of DNA fragments

During my PhD our lab adopted a new technique for building DNA constructs, named Gibson Assembly after the researcher who devised it [Gibson, 2011]. Whilst traditional DNA cloning relies on the religation of complementary restricted DNA ends, Gibson Assembly only requires a homologous overlap (optimum 40 bp) between neighbouring DNA fragments, and these homologous ends can be easily introduced in the tails of PCR primers. With traditional DNA cloning, one is limited to changing one fragment (or DNA module) at a time in the target vector, however through optimisation of the Gibson Assembly technique, we were able to simultaneously assemble up to seven (the most we have tried to date) DNA fragments of various sizes (~ 150 bp to 3 kb) in a single linear DNA assembly reaction. Although in this reaction, the seven-fragment-assembled DNA was of relatively low abundance amongst various other spurious products, we found that we could select for it by PCR amplifying with the two most distal primers and then use this linear assembly product subsequently in a two-fragment assembly with a suitably-prepared vector backbone.

Gibson assembly relies on the step-wise action of three common DNA enzymes (T5 exonu-

clease, Phusion DNA polymerase and then Taq DNA ligase), however the reaction conditions have been optimised such that all three phases can be combined into one isothermal step (50 °C, 1 hour). The reaction occurs thus: T5 exonuclease first chews back one strand of linear DNAs from the 5' end, generating complementary overhangs between assembly fragments, which naturally anneal to one another. However T5 exonuclease chews back further than the ~ 40 bp homology between assembly fragments therefore resulting in single-stranded gaps, which are then filled in by the 5' to 3' DNA polymerising activity of Phusion DNA polymerase. Finally, the single-stranded nick, corresponding to the junction between the extent of T5 digestion and Phusion fill in, is sealed by Taq DNA ligase.

Gibson isothermal assembly reactions consisting of 15 μL master mix (made in-house, see below) and 5 μL of DNA assembly fragments, were carried out in thin-walled, 0.2 mL PCR tubes (50 °C, 1 hour). Occasionally 5 μL of the assembly reaction was checked on an agarose gel for the presence of a linear assembly fragment of the desired size or of a slow-running band presumably corresponding to a relaxed circular vector assembly product. Purified vector assembly reactions were then transformed into a cloning host or directly integrated into the chromosome of desired target strains (for pIT plasmids), a technique termed clonetegration (see Section 4).

1.2 mL batches of Gibson assembly master mix (80 reactions) were made with the following ingredients: 320 μL 5x IOS buffer (see below), 0.64 μL 10 U/μL T5 exonuclease (Epicentre), 20 μL 2 U/μL Phusion high-fidelity DNA polymerase (Finnzymes), 160 μL 40 U/μL Taq DNA ligase (NEB), H_2O to 1.2 mL. 15 μL aliquots in thin-walled PCR tubes were stored at −20°C ready for use.

2 mL batches of 5x IOS buffer consisted of: 1 mL 1 M Tris-HCl pH 7.5, 50 μL 2 M $MgCl_2$, 200 μL 10 mM Deoxynucleotide Solution Mix (NEB), 100 μL 1 M DTT, 0.5 g PEG-8000 (Sigma-Aldrich), 200 μL 50 mM NAD (NEB), H_2O to 2 mL and stored at −20°C.

6.1.3 Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction was used for screening of clones (from colonies), amplification of a template for sequencing or amplification of DNA for cloning and Gibson Assembly. KAPA 2G

Robust Polymerase (KAPA biosystems) was used for screening of clones by colony PCR and preparation of templates for DNA sequencing, whilst the high-fidelity KAPA HiFi polymerase was used for amplification of DNA for cloning purposes. Reaction set-up and temperature cycling were performed according to the manufacturer's directions `http://www.kapabiosystems.com/`, using thin-walled, 0.2 mL PCR tubes and a Bio-Rad DNA Engine thermal cycler (Bio-Rad).

Screening of clones by colony PCR

PCR primers were selected to amplify across the cloned region and clones were identified either by the size and presence of the product(s) or, if the newly inserted DNA differed from the original by only a few base pairs, by diagnostic digest of the PCR product. Reaction volumes of 10 μL were prepared containing 1x 2G Buffer B (KAPA Biosystems), 200 μM each dNTP, 1 ng/μL each primer and 0.2–0.4 units of KAPA2G Robust DNA polymerase (KAPA biosystems). A sterile wire was used to touch freshly streaked transformant colonies into the PCR reaction.

Preparation of templates for DNA sequencing

Templates for Sanger DNA sequencing were prepared from bacterial colonies (see above) or from plasmid DNA, in which case ~ 1 ng of plasmid was used as template.

Preparation of DNA for cloning and assembly

DNAs from various sources, such as plasmid and genomic DNA preparations or bacterial colonies were amplified at high fidelity with primers who's tails contained either restriction enzyme sites (for traditional cloning) or 30–40 bp of homology with desired target DNAs (for Gibson Assembly). It was found that a 40 μL reaction volume followed by gel extraction usually provided a clean DNA fragment suitable for downstream applications.

6.1.4 Big Dye Sequencing reactions

DNA sequencing was performed using BigDye Terminator Version 3.1 (Applied Biosystems) according to the directions provided by the Australian Genomics Research Facility (AGRF)

sequencing service.http://www.agrf.org.au/index.php?id=72. Briefly, PCR products were purified using a DNA Clean & Concentrator-5 Kit (Zymo Research). Sequencing reactions contained 1 μL BigDye v3.1 Ready Mix, 1.5 μL 5x BigDye BDT dilution buffer, 5–75 ng template and 20 ng primer in a total volume of 10 μL. Sequencing reactions were incubated in thin-walled 0.2 mL PCR tubes at 96°C for 20 seconds, then 30 cycles of 96°C for 10 seconds, 47°C for 20 seconds and 57°C for 1 minute. Sequencing reactions were then precipitated by adding 40 μL 75% isopropanol and transferred to a 1.5 mL Eppendorf tube. Sequencing products were collected by centrifugation at 16100 g at room temperature for 10 minutes. The pellet was then washed with 200 μL 75% isopropanol, centrifuged for a further 5 minutes with the tubes in the same orientation, and the supernatant was removed. Samples were then dried by heating with the tube lid open. The resultant chromatograph and sequence files were viewed using ApE Plasmid Editor.

6.1.5 Microtitre plate LacZ assays

Original assay protocol

LacZ activity in a culture was determined by a kinetic assay in microtitre plates described by Dodd et al. [2001]. For this assay, 100 μL LB media, plus antibiotics was added to 96 well flat-bottomed microtitre plates (Corning Costar 3599) and innoculated with either 1 μL of an overnight culture or a colony picked with a sterile wire from fresh streaks on selective agar plates. The plate was sealed by stickytaping on a modified lid with a rubber insert (which prevents evaporation) and incubated overnight at 37°C without shaking. The following day, the plate was shaken for 15 seconds using a Multiskan Ascent plate reader (Labsystems) to resuspend the cultures. Cultures were then diluted 1/5 into the same media and their optical density was assayed using the plate reader and a 620 nm filter. OD_{620} values were converted to equivalent OD_{600} values by the method of Dodd et al. [2001]. Cultures were then normalized to $OD_{620} \sim 0.15$ by adding 50 μL of the 1/5 dilution to variable amounts of the same media. Finally, to achieve a starting $OD_{620} \sim 0.003$, 2 μL of this dilution was added to 98 μL of the same media and the plate was incubated at 37°C with shaking in a plate shaker-incubator (Grant-bio PHMP-4). Once the cultures had grown to $OD_{620} \sim 0.1$–0.2, 20 μL was added to wells containing 220

μL TZ8+ lysis and assay buffer (220 μL TZ8+ contains 150 μL TZ8 (100 mM Tris HCl, pH 8.0, 1 mM $MgSO_4$, 10 mM KCl), 40 μL of ONPG (o-nitrophenyl-β-D-galactoside 4 mg/mL in TZ8), 1.9 μL of 2-mercapoethanol, and 0.95 μL of polymyxin B (20 mg/mL; Sigma)). The plate was shaken in the plate reader pre-warmed to 28°C, and then readings of A_{414} were taken every 30 seconds for 30 minutes. The final A_{620} of the culture and A_{414} vs time was used to to calculate the LacZ units as:

$$\text{LacZ Units} = \frac{200000 \times \text{slope}}{\text{OD}_{600} \times \text{culture volume (in } \mu\text{L)}}$$

Revised assay protocol with minimal media

Stationary phase cultures of strains were grown as above but in M9 Minimal Medium (1x M9 salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% glycerol, 0.01 mM $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$, 3 μL of which was subcultured into 97 μL fresh MM in the corresponding well of a new plate, which was sealed and cultures grown until early–mid log phase (aiming for an Optical Density (OD), A_{620} = 0.1–0.15). Generally, the overnight plate grew quite evenly such that the OD normalisation step (mentioned above) wasn't required, and most often the final A_{620} of cultures was ~ 0.1–0.12. 20 μL of this culture was added to 220 μL TZ8+ buffer (with MM in place of LB), in another plate pre-warmed to ~ 30°C and the plate was read as above.

6.1.6 Protein purification and measurement

Small-scale culture of expression strain to test IPTG induction

A 4 mL overnight culture of BL21 pET3a–His6–TEV–*lac*I was diluted 1/200 into 10 mL LB + Cm_{30} and grown at 37°C to OD_{600} = 0.6. The culture was then split into two, 1 mM IPTG was added to one and the cultures were incubated for a further 2 hours. Cultures were then chilled on ice (~ 10 minutes) and a final OD_{600} reading was taken. 1 mL of each culture was taken into an Eppendorf tube and pelleted (5', 13 krpm, 4°C), after which the supernatant was removed and a freeze/thaw step was performed (10', −80°C) to aid cell lysis. Pellets were then resuspended in 100 μL B-Per + Benzonase (25 U/μL) and incubated (15', 4°C) to allow digestion of chromosomal DNA. 40 μL of this resuspension was then taken into a screw-cap sample storage tube (the total protein sample). A further 50 μL was transferred to a sample

storage tube and was pelleted (10', 13 krpm) after which 40 μL of supernatant was taken (the soluble fraction). The remaining supernatant was removed and the pellet was resuspended in 50 μL B-Per (the insoluble fraction). All fractions were then mixed with 50 μL 2x SDS loading buffer and heated (10', 70°C) and stored at −20°C for later use in SDS-PAGE.

Large-scale culture for protein purification

A 4 mL overnight culture of BL21 pET3a–His6–TEV–*lac*I was diluted 1/100 into 500 mL LB + Amp_{80} and grown at 37°C to OD_{600} = 0.6. It was found that the addition of IPTG was unnecessary for induction of LacI, and thus the culture was simply grown for a further 2 hours and then chilled on ice for 30 minutes. The culture was then divided in two, cells were pelleted (15', 6000 g, 4°C), the supernatant was poured off and pellets were stored at −20°C for later use.

Preparation of insoluble proteins using a cell homogeniser

A pellet from a large-scale culture was resuspended in 20 mL ice cold TBS + 0.1% Triton and passed, at 17000 PSI, three times through a pre-chilled Microfluidics Homogeniser fitted with a H10Z interaction chamber. The collected homogenate was then pelleted (20', 30000 g), the supernatant was discarded and the entire process was repeated with the final pellet being stored at −20°C for later use.

Nickel column purification of LacI protein

N-terminally His-tagged LacI protein was purified via Nickel affinity chromatography using a 5 mL HiTrap Chelating HP column (GE Healthcare). Column flow-through was regularly checked for the presence of protein by adding 10 μL flow through to 200 μL Bradford reagent. The column, which had been stored (20% Ethanol, 4°C), was charged with Nickel ions using a 5 ml syringe (3 column volumes (CVs) H_2O, 3 CVs 100 mM $NiSO_4$, 3 CVs H_2O, flow rate ~ 5 mL/min). The column was then equilibrated in Dissolve Buffer (Buffer A + 7 M Urea + 20 mM imidazole) (3 CVs). The pellet from the cell homogenisation step was resuspended in 9 mL Dissolve Buffer, sonicated (3 × 30 seconds) and applied slowly to the equilibrated column (flow rate ~ 1 mL/min with a 10 minute incubation every 5 mL). Unbound protein was washed from

the column with Dissolve Buffer (5 CVs, flow rate ~ 5 mL/min). The column was then washed with Wash Buffer (Buffer A + 7 M urea + 100 mM imidazole) (~ 10 CVs). Immobilised His6-LacI was then eluted in 10 mL Elution Buffer (Buffer A + 7 M urea + 500 mM imidazole) in 1 mL fractions. Elution fractions were tested for the presence of protein using Bradford reagent and it was found that His6-LacI eluted in a tight elution peak (fractions 3-6). The column was then stripped of Nickel ions (3 CVs 100 mM EDTA) and stored (3 CVs H_2O, 3 CVs 20% Ethanol, 4°C).

Dialysis of eluted protein

Peak elution fractions from the Nickel column were combined and sealed in dialysis tubing and the tubing was placed in 500 mL PBS + 4M urea at 4°C from 4 hours to overnight. Dialysis was continued in PBS + 2M urea, PBS + 1 M urea and finally twice in PBS alone and it was found that LacI precipitated after equilibration to 1 M urea. The dialysis tubing was then carefully pierced and protein was removed using a pasteur pipette and stored at −20 or −80°C.

6.1.7 Preparation of cellular extracts for Western Blotting

A 3 mL overnight culture of a LacI-expressing strain was diluted 1/50 into 6 mL M9MM and grown with shaking at 37°C to OD_{600} = 0.4–0.6. Cultures were then chilled on ice (~ 10 minutes) and a final OD_{600} reading was taken. 1 mL (standard concentration) or 2–4 mL (higher concentration) of each culture was taken (in 1 mL amounts) into a screw-cap sample storage tube and pelleted (5', 13 krpm, 4°C), after which the supernatant was thoroughly removed and a freeze/thaw step was performed (10', −80°C) to aid cell lysis. Pellets were then resuspended in 50 μL B-Per + Benzonase (25 U/μL) and incubated (15', 4°C) to allow digestion of chromosomal DNA. Then 17 μL 4x LDS loading buffer (Invitrogen NuPAGE) was added and tubes were heated (10', 70°C) after which samples were ready to be analysed or were stored at −20°C.

Polyacrylamide gel electrophoresis of proteins and transfer to PVDF membrane

Protein samples (~ 20 μL) were loaded onto 12-well 4-12% Bis Tris gels (Invitrogen) and run at 200 V with MOPS running buffer (Novex). Finished gels were transferred onto 0.2 μm PVDF

membrane using an iBlot transfer apparatus (Version 1, Invitrogen) following the manufacturer's instructions.

Antibody detection of LacI on Western Blots

Freshly transferred membranes were blocked in 5% BSA (30'–Overnight, 4°C, shaking) and washed twice in TBS-T (a wash is adding 20 mL TBS-T, swirling the membranes and pouring off). Membranes were then incubated with a 1/200 dilution of polyclonal rabbit anti-LacI antibody (Rockland) preadsorbed against an extract of the parental strain (E4643) (60', RT, shaking) and then washed 4x in TBS-T with the final wash having a 2 minute shake. Next, membranes were incubated in a 1/4000 dilution of Cy5-labelled goat-anti-rabbit 2° antibody (GE Healthcare) (30', RT, shaking, in the dark) followed by 4x wash in TBS-T and 2x wash in TBS. Membranes were dabbed dry on blotting paper and then dried fully by placing them on blotting paper in the dark at 37°C for ~ 30 min. Dried membranes were imaged on the Cy5 channel of a GE Typhoon imager (Amersham Biosciences) (PMT 475 V). Western images were analyzed using ImageJ software.

6.2 Reference and Tables

6.2.1 Bacterial Strains

Table 6.1 details all the *E. coli* strains used and manipulated in this study. If the glycerol number appears as the name, then the strain was made or altered in-house. Further description of genotypes can be found at `http://openwetware.org/wiki/E._coli_genotypes`.

Table 6.1 ***E. coli* Strains used in this study**

Name	Glycerol	Genotype	Comment
MG1655	E4640	*F⁻ λ⁻ ilvG⁻ rfb-50 rph-1*	A widely used 'wild-type' *E. coli* K-12 strain (BW30270 (CGSC 7925))

Table 6.1 (Continued on next page)

Table 6.1 (Continued from last page)

Name	Glycerol	Genotype	Comment
E4643	E4643	MG1655 Δ*lac*IZYA	Recombineering was used to delete the entire Lac operon from MG1655. This is the base strain into which LacI and λCI expression cassettes and LacZ reporter constructs were integrated.
EC100D pir$^+$	E4644	F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80d(*lac*ZΔM15) Δ*lacX74* *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ$^-$ *rps*L *str*^{R} *nup*G pir$^+$(DHFR)	EC100D pir$^+$. Strain for propagating integration plasmids with pir-dependent origins. Low copy number. Obtained from Epicentre.
EC100D pir-116	E4645	F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80d(*lac*ZΔM15) Δ*lacX74* *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ$^-$ *rps*L *str*^{R} *nup*G pir-116(DHFR)	EC100D pir-116. Strain for propagating integration plasmids with pir-dependent origins. High copy number. Mutations that remove strong promoters and/or for genes with some toxic effect will likely be selected for.
DH5α	E4241	F$^-$ *end*A1 *gln*V44 *thi*–1 *rec*A1 *rel*A1 *gyr*A96 *deo*R *nup*G Φ80d(*lac*ZΔM15) Δ(*lacZYA–argF*)U169, *hsd*R17($r_K^-m_K^+$), λ$^-$	Routine cloning strain, e.g. for plasmids with ColE1-dependent origins.

Table 6.1 (Continued on next page)

Table 6.1 (Continued from last page)

Name	Glycerol	Genotype	Comment
DH5αZ1	E2869	DH5α (P_{lacI^q}–*lacI* P_{N25}–*tet*^{R} *spec*^{R})$_\lambda$	DH5α containing a LacI expression cassette. Used for cloning plasmids harbouring genes under the control of the $P_{lac^{UV5}}$ promoter.
BL21 λ(DE3) pLysS	E4287	F$^-$ *omp*T *gal dcm lon* *hsd*S$_B$(r$_B^-$m$_B^+$) λ(DE3) pLysS(*chlor*^{R})	Protein expression strain
HMS174 λ(DE3) pLysS	E4284	F$^-$ *rec*A1 *hsd*R(r$_{K12}^-$m$_{K12}^+$) (*rif*^{R}) λ(DE3) + pLysS	Protein expression strain
DP120	DP120	E4643 (pIT3–SH–P_{lacI^q}–*lacI*)$_{HK022}$	High LacI expressing strain into which Lac looping constructs were integrated.
DP116	DP116	E4643 (pIT3–SH–P_{lacI}–*lacI*)$_{HK022}$	Low LacI expressing strain into which Lac looping constructs were integrated.
DP115	DP115	E4643 (pIT3–SH–empty)$_{HK022}$	Absent LacI expressing strain into which Lac looping constructs were integrated.
DP129	DP129	DP120 + pλINT	LacI expressing strain with λ helper plasmid [Haldimann and Wanner, 2001] ready for integration of reporter constructs.

Table 6.1 (Continued on next page)

Table 6.1 (Continued from last page)

Name	Glycerol	Genotype	Comment
DP127	DP127	DP116 + pλINT	LacI expressing strain with λ helper plasmid ready for integration of reporter constructs.
DP126	DP126	DP115 + pλINT	LacI expressing strain with λ helper plasmid ready for integration of reporter constructs.
DP503	DP503	E4643 (pIT3–SH–empty)$_{HK022}$ (pIT3–TO–empty)$_{186}$	LacI absent λCI absent strain for reporter integration
DP505	DP505	E4643 (pIT3–SH–empty)$_{HK022}$ (pIT3–TO–P_{RM}*cI–oL3–4*)$_{186}$	LacI low λCI absent strain for reporter integration
DP507 or 511	DP507 or 511	E4643 (pIT3–SH–P_{lacI}–*lacI*)$_{HK022}$ (pIT3–TO–empty)$_{186}$	LacI high λCI absent stain for reporter integration
DP437	DP437	E4643 (pIT3–SH–P_{lacI}–*lacI*)$_{HK022}$ (pIT3–TO–P_{RM}*cI–oL3–4*)$_{186}$	LacI absent λCI 3WLU strain for reporter integration
DP439	DP439	E4643 (pIT3–SH–P_{lacI^q}–*lacI*)$_{HK022}$ (pIT3–TO–empty)$_{186}$	LacI low λCI 3WLU strain for reporter integration
DP441	DP441	E4643 (pIT3–SH–P_{lacI^q}–*lacI*)$_{HK022}$ (pIT3–TO–P_{RM}*cI–oL3–4*)$_{186}$	LacI high λCI 3WLU strain for reporter integration
DP508	DP508	DP503 + pλINT helper (λCIts controlled)	LacI and λCI expressing strains with λattB integration helper plasmid.

Table 6.1 (Continued on next page)

Table 6.1 (Continued from last page)

Name	Glycerol	Genotype	Comment
DP509	DP509	DP505 + pλINT helper (λCIts controlled)	LacI and λCI expressing strains with λattB integration helper plasmid.
DP510	DP510	DP507 + pλINT helper (λCIts controlled)	LacI and λCI expressing strains with λattB integration helper plasmid.
DP446	DP446	DP437 + pλINT helper (186CIts λattB)	LacI and λCI expressing strains with λ integrase (under the control of 186CIts) integration helper plasmid (Constructed by Ian Dodd).
DP447	DP447	DP439 + pλINT helper (186CIts λattB)	LacI and λCI expressing strains with λ integrase (under the control of 186CIts) integration helper plasmid.
DP448	DP448	DP441 + pλINT helper (186CIts λattB)	LacI and λCI expressing strains with λ integrase (under the control of 186CIts) integration helper plasmid.

6.2.2 Primers

Table 6.2 details all the oligonucleotides used in this study. Lowercase type generally indicates a primer tail containing restriction enzyme sites or overlap for Gibson Assembly.

Table 6.2 **Primers used in this study**

Primer Name	Sequence	Use
158 RNAse site A	cccaagcttCCTAACTAACTAGCGATCCCGA	For screening and sequencing clones. Primes within the DNA looping chassis. Also for amplifying the *lacZ* gene.
185 pZC31.3	CGTGCCGATCAACGTCTCATT	Amplifying the *lacZ* gene for sequencing.
215 T7 promoter	TAATACGACTCACTATAGG	Primes at the T7 promoter. For secreening and sequencing clones.
259 lacZ R2	CCCTAACGCCTGGGTCG	For screening and sequencing clones. Primes within *lac*Z.
315 T7 term	GCTAGTTATTGCTCAGCGGTGG	For screening and sequencing clones. Primes at the T7 terminator.
326 LacZ-bio	CCAGCGCCCGTTGCACCACAG	For screening and sequencing clones. Primes within *lac*Z.
329 long 57	TAAACTGCCAGGAATTGGGGATC	For screening and sequencing clones.
354 cos bottom	CATACCCCGCCACCATCACG	tW PCR to test new LLLs
407 tL3-r	AGGATGCGTCATCGCCATTA	For screening and sequencing clones. Primes within the DNA looping chassis.
440 USP+1 pure	GTAAAACGACGGCCAGTG	For screening and sequencing clones. Primes within *lac*Z.
462 lam attP 2	ATGACAGAGGCAGGGAGTGG	For screening integrants at the phage λ*att* attachment site.
466 Wanner L P1	GGCATCACGGCAATATAC	For screening integrants at the phage λ*att* attachment site.

Table 6.2 (Continued on next page)

Table 6.2 (Continued from last page)

Primer Name	Sequence	Use
467 CRIM P2	ACTTAACGGCTGACATGG	For screening integrants at the phage attachment sites.
468 CRIM P3	ACGAGTATCGAGATGGCA	For screening integrants at the phage attachment sites.
469 Wanner L P4	TCTGGTCTGGTAGCAATG	For screening integrants at the phage λ*att* attachment site.
585 attHK022 B1	GGAATCAATGCCTGAGTG	For screening integrants at the phage HK022 attachment site.
586 attHK022 B2	GGCATCAACAGCACATTC	For screening integrants at the phage HK022 attachment site.
610 186attB left	CTCATTCGAAACCACCCACCG	Screening integrants at 186 attachment site.
611 186attB right	GATCATCATGTTTATTGCGTGG	Screening integrants at 186 attachment site.
613 pBSHSL 2484 KasI	tccagaggcgccGGGGGGTTCGTGCA CACAG	For screening and sequencing clones. Primes within *lac*I.
614 pBSlamCI 3377 NheI	tggtaagctagcGGGGCGAAAACTCT CAAGGATCTTAC	For screening and sequencing clones. Primes next to *amp*^{R}.
617 plac -122 SalI	attacgtcgacTGGCACGACAGGTTT CC	For screening and sequencing clones. Primes within pIT–HF.
625 phi80attP-1	AAAGAAACAGAGAAGGGCAC	Phi80 integrant screening.
626 phi80attP-2	GTTCGCAGAGTGTTATGGTT	Phi80 integrant screening.
627 phi80attB-2	TGGCCTTAACAAAGACATA	Phi80 integrant screening.
789 CmR upstream	CTCAAAAAATACGCCCGGTAGTG	For screening and sequencing clones. Primes next to *chlor*^{R}.

Table 6.2 (Continued on next page)

Table 6.2 (Continued from last page)

Primer Name	Sequence	Use
813 BamHI tonBterm up	gagttggatccGGTTAATTAACGGCAC CACC	Screening pIT5 clones, used to integrate Oid or Om at the Phi80 att site.
826 LacCheck-TopInLacI	TCTACCATCGACACCACCAC	For screening and sequencing clones. Primes within *lac*I.
837 pIT3 Insert Scr	GCCTGTCAGTTTAGGTTAGGCG	For screening and sequencing clones. Primes within pIT–HF.
839 pBla down-stream	ggggggggcccggtaccttaccggtGAAGCA TTTATCAGGGTTATTG	For screening and sequencing clones. Primes within DNA looping chassis.
852 LoopChas SX rev	GCGCTAATGCTCTGTTACAGG	For screening and sequencing clones. Primes within DNA looping chassis.
853 LoopChas SN for	TTTGCGACCAGTTCAAGACG	For screening and sequencing clones. Primes within DNA looping chassis.
857 M13-rev	CAGGAAACAGCTATGACCATG	For screening and sequencing clones.
859 ftsk 2360bp Rev	GATTATCGCGGAATTTGGCGTTA TCC	For amplifying *fts*K from the *E. coli* chromosome in order to lengthen Spacer 3 in the DNA looping chassis.
862 pIT HF nheI screen	AGTCACTATGAATCAACTACTTAG ATGG	For screening and sequencing clones. Primes about DNA looping chassis.
863 pIT HF sphI screen	GCCACTCTTGCGAATGACC	For screening and sequencing clones. Primes about DNA looping chassis.

Table 6.2 (Continued on next page)

Table 6.2 (Continued from last page)

Primer Name	Sequence	Use
864 LacSpaceSbf	gccctgcaggGAAACTGCGCACCGCTATC	For amplifying *val*S from the *E. coli* chromosome in order to lengthen Spacer 1 in the DNA looping chassis. Also used for screening and sequencing clones within the DNA looping chassis.
871 P1(p21 test)	ATCGCCTGTATGAACCTG	Screening ϕ21 att site integrants
872 P4(p21 test)	TAGAACTACCACCTGACC	Screening ϕ21 att site integrants
885 LacSpaceAsc600L	atgcatggcgcgccGATCATCGGCAGGGCGTG	Screening long lac loop constructs, specifically the 10 kb construct.
903 LacI N NcoI	atatatccatggAACCAGTAACGTTATACGATGTCG	For screening integrants at the phage ϕ21 attachment site.
904 LacI C BamHI	atatatggatcctcaCTGCCCGCTTTCCAG	For screening integrants at the phage ϕ21 attachment site.
967 LacIfuse for	ttaattggtaccTCAGTAGCTGAACAGGAGGG	Forward primer for creating translational fusion of LacI to LacZ. Acc65I end.
968 LacIfuse rev	ttaattggatccgcGACACCGGCATACTCTGC	Reverse primer for creating translational fusion of LacI to LacZ. BamHI end.
969 FtsK up to 300	aaaatgtcgacAATTACTCTCCGGGGCCG	For amplifying *fts*K from the *E. coli* chromosome in order to lengthen Spacer 3 in the DNA looping chassis to 300 bp.

Table 6.2 (Continued on next page)

Table 6.2 (Continued from last page)

Primer Name	Sequence	Use
970 Rne anchor Nsi	ttaattatgcatCGACTACCGCTTCTTCGGC	For amplifying *rne* from the *E. coli* chromosome in order to lengthen Spacer 2 in the DNA looping chassis.
971 Rne up to 300	ttaattggccggccAACGTCAGGCGCAACAAG	For amplifying *rne* from the *E. coli* chromosome in order to lengthen Spacer 2 in the DNA looping chassis to 300 bp.
977 pUC Ori-BlaNheI	atatatgctagcGCGGTAATACGGTTATCCACAG	For amplifying the Origin and amp^R from pUC57 in order to join to the DNA looping chassis to create pUC–HF.
978 pUC Ori-BlaBsu36I	atatatcctgaggCACTTTTCGGGGAAATGTGC	For amplifying the Origin and amp^R from pUC57 in order to join to the DNA looping chassis to create pUC–HF.
983 GBA LC Nhe	cagagaagcacaaagcctcgcaatccagtgcaaaGCTAGCTTCTTCGTCTGTTTCTACTG	Screening clones.
986 GBA into LC	CTCGAGAGTGCGACAGGTTTG	Screening clones. In particular LLL16.
1022 Mod1R-Sph	cgggatcgctagttagttaggatcgcccaagcttgcatgcGGTACCACCCGGGATTCTGG	Gibson Assembly.
1023 Mod1L	CCTGCAGGTGACTAACTGATATAGTG	Gibson Assembly.
1024 Mod2R	GGCGCGCCGACACCG	Gibson Assembly.
1025 Mod2L	GGCCGGCCCTCCCATTTC	Gibson Assembly.

Table 6.2 (Continued on next page)

Table 6.2 (Continued from last page)

Primer Name	Sequence	Use
1026 Mod3R	ATGCATTCAGCTTCTGCAAAAAG G	Gibson Assembly.
1027 Mod4L-Bam	tcccgtagtatccccactcacttagtcaggtacc GGATCCGTTTCTTTTTTGTGCTCA TAC	Gibson Assembly.
1028 Space1R	GCTCTGTTACAGGTCACTATATC AGTTAGTCACCTGCAGGgaaactgc gcaccgctatc	Gibson Assembly.
1029 Space2L	GACGTGGTAATGCCTTTTTGCAG AAGCTGAATGCATcgactaccgcttcttc ggc	Gibson Assembly.
1030 Space1L600	CAGCGTTTTGAGTTCGCTAAATC CGGTGTCGGCGCGCCgatcatcggc agggcgtg	Gibson Assembly.
1031 Space2R600	TCAACGTTACACCGCCGAAATGG GAGGGCCGGCCaaaccaaaccgaccg agcaa	Gibson Assembly.
1041 Space1L900	CAGCGTTTTGAGTTCGCTAAATC CGGTGTCGGCGCGCCCTCAACC GCTTCAACCGC	Gibson Assembly.
1042 Space1L1800	CAGCGTTTTGAGTTCGCTAAATC CGGTGTCGGCGCGCCATCGAAG CGGAAGCTGTCC	Gibson Assembly.
1043 Space2R900	TCAACGTTACACCGCCGAAATGG GAGGGCCGGCCtgcgcgtgcgtaaagg g	Gibson Assembly.

Table 6.2 (Continued on next page)

Table 6.2 (Continued from last page)

Primer Name	Sequence	Use
1044	TCAACGTTACACCGCCGAAATGG	Gibson Assembly.
Space2R1400	GAGGGCCGGCCtaacactaacctcgaa gctgcc	
1045	CAGCAATGTATCAGCCGCTGTTT	Gibson Assembly.
Space3R600	TCGGTCGACCAACCGGTTGCGC CAC	
1046 Mod3L	GTCGACCGAAAACAGCGGC	Gibson Assembly.
1047 Mod4R	GGGACCTTAAACTGCCAGGAATT G	Gibson Assembly.
1048 Space3L	CACTCATTTAGATCCCCAATTCCT GGCAGTTTAAGGTCCCgtgacaagtt agaaatgcgcgc	Gibson Assembly.
1068	CAGCGTTTTGAGTTCGCTAAATC	Gibson Assembly.
Space1L400	CGGTGTCGGCGCGCCCAGTATC GCCCAGCAGGG	
1069	CAGCGTTTTGAGTTCGCTAAATC	Gibson Assembly.
Space1L500	CGGTGTCGGCGCGCCTCGTCGC CAACGATCGG	
1070 Space4L	CACTCATTTAGATCCCCAATTCCT GGCAGTTTAAGGTCCCTTTGCCG TCTGGTTGATGGC	Gibson Assembly.
1071	GCTCTGTTACAGGTCACTATATC	Gibson Assembly.
Space4R5.6kb	AGTTAGTCACCTGCAGGAGAAGC GAAACCGGAACGTC	
1072	GCTCTGTTACAGGTCACTATATC	Gibson Assembly.
Space4R10kb	AGTTAGTCACCTGCAGGCGGGAT CGCTACCCAGATGG	

Table 6.2 (Continued on next page)

Table 6.2 (Continued from last page)

Primer Name	Sequence	Use
1079	agtcccgtagtatccccactcacttagtcaGGT	Gibson Assembly.
Mod1Rflip	ACCACCCGGGATTCTGG	
1080 Mod4Lflip	cgggatcgctagttagttaggatcgcccaagctt	Gibson Assembly.
	gcatgcGGTACCGGATCCGTTTCTT	
	TTTTGTGCTCATAC	
1112	acgcccgccataaactgccaggaattggggat	Gibson Assembly.
Mod3LpIT5	cggaattcGTCGACCGAAAACAGCG	
	GC	
1113	ctgtcagtttaggttaggcgccatgcatctcgagc	Gibson Assembly.
Mod3RpIT5	tgcagATGCATTCAGCTTCTGCAAA	
	AAGG	
1114	CCTGCAGGTGACTAACTGAacc	Gibson Assembly.
Mod1Lnew		
1115	ttattccgggttgtcgagggttcagttagtcacctg	Gibson Assembly.
SpaceFtsK-	caggGTGACAAGTTAGAAATGCGC	
R	GC	
1116	cagcgttttgagttcgctaaatccggtgtcggcg	Gibson Assembly.
SpaceFtsK-	cgccATCACTCGCTTTGAATTGAAC	
L300	CTG	
1117	cagcgttttgagttcgctaaatccggtgtcggcgc	Gibson Assembly.
SpaceFtsK-	gccCAATATCAGCAGCCGCAACAA	
L600	C	
1118 tW-	GGAAGCATAAAGTGTAAAGCCTG	Gibson Assembly.
XhoEnd	GGGTGCCTACTCGAGcgtctgagacg	
	tgatggtgg	
1119 tW-	CGCATGTACCAAGTGTACAGCG	Gibson Assembly.
AscEnd		

Table 6.2 (Continued on next page)

Table 6.2 (Continued from last page)

Primer Name	Sequence	Use
1129	CAGCAATGTATCAGCCGCTGTTT	Gibson Assembly.
Space3R1500sal	TCGGTCGACGTACCCGTCGCGAT GATACC	
1130	CACTCATTTAGATCCCCAATTCCT	Gibson Assembly.
Space3L1500ppu	GGCAGTTTAAGGTCCCTACTGAT TGCGCTGGGCTTC	
1131	TCAACGTTACACCGCCGAAATGG	Gibson Assembly.
Space2R300	GAGGGCCGGCCAacgtcaggcgcaac aag	
1152	ttattccgggttgtcgagggttcagttagtcacctg	Gibson Assembly.
SpaceFtsK-R2	caggGTTTGCGCAAACTGACGAGC	
1153	cagcgttttgagttcgctaaatccggtgtcggcgc	Gibson Assembly.
SpaceFtsK2-L900	gccAACAGCAGCCTGTTGTGGAG	
1154	cagcgttttgagttcgctaaatccggtgtcggcgc	Gibson Assembly.
SpaceFtsK2-L1800	gccCAGCGGGCCTGACGTTG	
1202 rev lacz lib	GATGGGCGCATCGTAACCG	For screening long lac loop clones.
1225 Mod4R-Mod2	aaccgcgaatcgaaaggttcgatgtggGGC GCGCCGGGACCTTAAACTGCCA GGAATTG	Gibson Assembly.
1226 Mod4L-Mod2	tcatcggcagcttcgaggttagtgttaGGCCG GCCGGATCCGTTTCTTTTTTGTG CTCATAC	Gibson Assembly.

Table 6.2 (Continued on next page)

Table 6.2 (Continued from last page)

Primer Name	Sequence	Use
1227 Space3R-300	CAGCAATGTATCAGCCGCTGTTTGCCG	Gibson Assembly.
1228 Space1L-300	CAGCGTTTTGAGTTCGCTAAATCCGGTGTCGGCGCGCCccacatcgaacctttcgattcg	Gibson Assembly.
1304 Space3R-458	cagcaatgtatcagccgctgttttcggtcgacCAGCCGTCCGTTGCATAAAC	Gibson Assembly.
1306 tW PCR Sbf end	cgacttccaggtcagagatagcggtgcgcagtttcCCTGCAGGTGACTAACTGAAC	Gibson Assembly.
1307 tW XhoEnd17	gatatgataaagtgtaaagcctggggtgcctactcgagCGTCTGAGACGTGATGGTGG	Gibson Assembly.

6.2.3 Reporter construct details

This section contains tables detailing the construction of the DNA looping reporter constructs. Table 6.3 details the LacI looping constructs, table 6.4 the LacI and λCI constructs, and Table 6.5 details the Gibson assembly fragments.

Each construct is named LL (LacLoop) or LLL (LacLamLoop). M1–4 stands for Modules 1–4, and S1–3 stands for Spacers 1–3. Thus the modules and spacers are indicated as described in Figure 4.2. Under each module(s) is indicated the source of the DNA, be it a restriction fragment (where the plasmid is named) or a Gibson assembly fragment (where a fragment number is given, see 6.5 for description of fragments). For the glycerol stock details, DP glycerol numbers are given. The construct may be contained in a pUC57 vector (pUC) or pIT integrating plasmid (pIT) and can be integrated into the LacI Absent (Abs), Low (Low) or High (High), or (x,y), where x is LacI and y is λCI. NA indicates not available.

Table 6.3 **Assembly Details for LacI looping constructs**

	Construct details							Glycerol stock details				
Description	**M4**	**S3**	**M3**	**S2**	**M2**	**S1**	**M1**	**pUC**	**pIT**	Abs	Low	High
LL1 *Oid–O3* 300 bp spacing	O^-	222	O^-	142	*Oid*	300	$P_{lac^{UV5}}$ *O3*	111	138	160	158	154
	pUC57–DS1–KpnI digest with Acc65I and religate											
LL2 $P_{lac^{UV5}}$ *O3* alone	O^-	222	O^-	142	O^-	300	$P_{lac^{UV5}}$ *O3*	124	139	169	167	163
	LacLoop1				pUC–DS2		LacLoop1					
LL3 No Lac operators	O^-	222	O^-	142	O^-	300	$P_{lac^{UV5}}$ O^-	131	142	202	198	200
			LacLoop2				pUC–DS2					
LL4 Proximal Lac minus, upstream *Oid* has no effect	O^-	222	O^-	142	*Oid*	300	$P_{lac^{UV5}}$ O^-	132	144	208	204	206
			LacLoop1				pUC–DS2					
LL5 *O1–O1* 300 bp spacing	O^-	222	O^-	142	*O1*	300	$P_{lac^{UV5}}$ *O1*	134	146	214	210	212
	LacLoop6				pUC–DS4		LacLoop6					
LL6 $P_{lac^{UV5}}$ *O1* alone	O^-	222	O^-	142	O^-	300	$P_{lac^{UV5}}$ *O1*	133	148	219	215	217
			LacLoop2				pUC–DS4					

Table 6.3 (Continued on next page)

Table 6.3 (Continued from previous page)

Description	M4	S3	M3	S2	M2	S1	M1	pUC	pIT	Abs	Low	High
LL7 Upstream *O1* has no effect	O^-	222 LacLoop3	O^-	142	*O1* pUC–DS4	300	$P_{lac^{UV5}}$ O^- LacLoop3	136	150	225	221	223
LL8 *Oid–O2*, 300 bp spacing	O^-	222	O^- LacLoop1	142	*Oid*	300	$P_{lac^{UV5}}$ *O2* pUC–DS3	171	194	182	180	177
LL9 $P_{lac^{UV5}}$ *O2* alone	O^-	222	O^- LacLoop2	142	O^-	300	$P_{lac^{UV5}}$ *O2* pUC–DS3	173	196	192	190	186
LL10 $P^-_{lac^{UV5}}$ *O1*	O^-	222	O^- pUC57–LL2	142	O^-	300	$P^-_{lac^{UV5}}$ *O1* pUC–KES5	284	291	299	300	301
LL11 $P^-_{lac^{UV5}}$ O^-	O^-	222	O^- pUC57–LL2	142	O^-	300	$P^-_{lac^{UV5}}$ O^- pUC–KES6	285	293	303	304	305
LL12 $P_{lac^{UV5}}$ *Oid* alone	O^-	222	O^- LacLoop2	142	O^-	300	$P_{lac^{UV5}}$ *Oid* pUC57–KES4	287	295	307	308	309

Table 6.3 (Continued on next page)

Table 6.3 (Continued from previous page)

Description	M4	S3	M3	S2	M2	S1	M1	pUC	pIT	Abs	Low	High
LL13 *Oid–Oid* 300 bp spacing	O^-	222	O^-	142	*Oid*	300	$P_{lac^{UV5}}$ *Oid*	289	297	311	312	313
	LacLoop4						pUC57–KES4					
pUC–HF–LL14 Spacer 3 to 300 bp	O^-	300	O^-	142	O^-	300	$P^-_{lac^{UV5}}$ *O1*	354	NA	NA	NA	NA
	LL10	pUC–DS2			pUC–HF–LL10							
pUC–HF–LL15 Spacer 2 to 300 bp	O^-	300	O^-	300	O^-	300	$P^-_{lac^{UV5}}$ *O1*	356	NA	NA	NA	NA
	pUC–HF–LL14			*rne*	pUC–HF–LL14							
pUC–HF–LL16 *Oid–O1*, 300 bp for TPM	O^-	300	O^-	300	*Oid*	300	$P^-_{lac^{UV5}}$ *O1*	358	NA	NA	NA	NA
	pUC–HF–LL15				pUC–DS1	pUC–HF–LL15						
pUC–HF–LL17 *Oid–O1*, 600 bp for TPM	O^-	300	*Oid*	300	O^-	300	$P^-_{lac^{UV5}}$ *O1*	360	NA	NA	NA	NA
	pUC–HF–LL15		pUC–DS5		pUC–HF–LL15							
pUC–HF–LL18 *Oid–O1*, 900 bp for TPM	*Oid*	300	O^-	300	O^-	300	$P^-_{lac^{UV5}}$ *O1*	362	NA	NA	NA	NA
	pUC–DS4				pUC–HF–LL15							

Table 6.3 (Continued on next page)

Table 6.3 (Continued from previous page)

Description	M4	S3	M3	S2	M2	S1	M1	pUC	pIT	Abs	Low	High
LL19 $P_{lac^{UV5}}$ *O3* distal minus	O^-	300	O^-	300	O^-	300	$P_{lac^{UV5}}$ *O3* LacLoop2	NA	NA	373	374	375
			pUC–HF–LL15									
LL20 *Oid–O3*, 300 bp	O^-	300	O^-	300	*Oid*	300	$P_{lac^{UV5}}$ *O3* LacLoop2	NA	NA	376	377	378
			pUC–HF–LL16									
LL21 *Oid–O3*, 600 bp	O^-	300	*Oid*	300	O^-	300	$P_{lac^{UV5}}$ *O3* LacLoop2	NA	NA	379	380	381
			pUC–HF–LL17									
LL22 *Oid–O3*, 900 bp	*Oid*	300	O^-	300	O^-	300	$P_{lac^{UV5}}$ *O3* LacLoop2	NA	NA	382	383	384
			pUC–HF–LL18									
LL23 $P_{lac^{UV5}}$ *O2* distal minus	O^-	300	O^-	300	O^-	300	$P_{lac^{UV5}}$ *O2* 1	NA	NA	397	398	399
			pUC–HF–LL15									
LL24 *Oid–O2* 300 bp	O^-	300	O^-	300	*Oid*	300	$P_{lac^{UV5}}$ *O2* 1	NA	NA	400	401	402
			pUC–HF–LL16									

Table 6.3 (Continued on next page)

Table 6.3 (Continued from previous page)

Description	M4	S3	M3	S2	M2	S1	M1	pUC	pIT	Abs	Low	High
LL25 *Oid–O2* 600 bp	O^-	300	*Oid*	300	O^-	300	$P_{lac^{UV5}}$ *O2*	NA	NA	403	404	405
			pUC–HF–LL17				1					
LL26 *Oid–O2* 900 bp	*Oid*	300	O^-	300	O^-	300	$P_{lac^{UV5}}$ *O2*	NA	NA	406	407	408
			pUC–HF–LL18				1					
LL27 *Oid–O2*, 1200 bp	O^-	300	*Oid*	600	O^-	600	$P_{lac^{UV5}}$ *O2*	NA	NA	421	422	423
		5		4	3	2	1					
LL28 *Oid–O2*, 1800 bp	O^-	300	*Oid*	900	O^-	900	$P_{lac^{UV5}}$ *O2*	NA	NA	485	486	487
		5		7	3	6	1					
LL29 *Oid–O2*, 3200 bp	O^-	300	*Oid*	1400	O^-	1800	$P_{lac^{UV5}}$ *O2*	NA	NA	488	489	490
		5		9	3	8	1					
LL30 $P_{lac^{UV5}}$ *O2* distal minus	O^-	300	O^-	1400	O^-	1800	$P_{lac^{UV5}}$ *O2*	NA	NA	491	492	493
		15		9	3	8	1					

Table 6.3 (Continued on next page)

Table 6.3 (Continued from previous page)

Description	M4	S3	M3	S2	M2	S1	M1	pUC	pIT	Abs	Low	High
LL31 *Oid–O2*, 400 bp	O^-	300	O^-	300	*Oid*	400	P_{lacUV5} *O2*	NA	NA	532	533	534
			16			17	1					
LL32 *Oid–O2*, 500 bp	O^-	300	O^-	300	*Oid*	500	P_{lacUV5} *O2*	NA	NA	535	536	537
			16			18	1					
LL33 *Oid–O2*, 5600 bp	*Oid*	5600 (‘spacer 4’)					P_{lacUV5} *O2*	Box12				
LL35 *Oid–O1*, 1200 bp for TPM	O^-	300	*Oid*	600	O^-	600	P^-_{lacUV5} *O1*	588	NA	NA	NA	NA
			25				24					
LL36 *Oid–O1*, 1800 bp for TPM	O^-	300	*Oid*	900	O^-	900	P^-_{lacUV5} *O1*	590	NA	NA	NA	NA
			26				24					
LL37 *Oid–O1*, 3200 bp for TPM	O^-	300	*Oid*	1400	O^-	1800	P^-_{lacUV5} *O1*	592	NA	NA	NA	NA
		5		9	3	8	24					

Table 6.3 (Continued on next page)

Table 6.3 (Continued from previous page)

Description	M4	S3	M3	S2	M2	S1	M1	pUC	pIT	Abs	Low	High
LL38 $P^{-}_{lac^{UV5}}$ *O1* distal minus for TPM	*O*⁻	300	*O*⁻	1400	*O*⁻	1800	$P^{-}_{lac^{UV5}}$ *O1*	593	NA	NA	NA	NA
	15		9		3	8	24					
LL39 $P_{lac^{UV5}}$ *O2* only, Series 2	*O*⁻	222	*O*⁻	142	*O*⁻	600	$P_{lac^{UV5}}$ *O2*	NA	NA	628	629	630
	Chassis PCR from pUC57–LL9					30	27					
LL40 *Oid–O2*, 242 bp, Series 2	*O*⁻	222	*O*⁻	142	*Oid*	242	$P_{lac^{UV5}}$ *O2*	NA	NA	631	632	633
	Chassis PCR from pUC57–LL8new											
LL41 *Oid–O2*, 300 bp, Series 2	*O*⁻	222	*O*⁻	142	*Oid*	300	$P_{lac^{UV5}}$ *O2*	NA	NA	634	635	636
	Chassis PCR from pUC57–LL8new					29	27					
LL42 *Oid–O2*, 600 bp, Series 2	*O*⁻	222	*O*⁻	142	*Oid*	600	$P_{lac^{UV5}}$ *O2*	NA	NA	637	638	639
	Chassis PCR from pUC57–LL8new					30	27					
LL43 *Oid–O2*, 900 bp, Series 2	*O*⁻	222	*O*⁻	142	*Oid*	900	$P_{lac^{UV5}}$ *O2*	NA	NA	663	664	665
	Chassis PCR from pUC57–LL8new					34	27					

Table 6.3 (Continued on next page)

Table 6.3 (Continued from previous page)

Description	**M4**	**S3**	**M3**	**S2**	**M2**	**S1**	**M1**	**pUC**	**pIT**	Abs	Low	High
LL44 *Oid–O2*, 1800 bp,	O^-	222	O^-	142	*Oid*	1800	$P_{lac^{UV5}}$ *O2*	NA	NA	666	667	668
Series 2	Chassis PCR from pUC57–LL8new					35	27					

Table 6.3: Details for LacI looping strains

Table 6.4 **Assembly Details for LacI and λCI looping constructs**

	Construct details							**Glycerol stock details, Base strain (LacI, λCI)**							
Description	**M4**	**S3**	**M3**	**S2**	**M2**	**S1**	**M1**	**pUC**	**pIT**	(−, −)	(Lo, −)	(Hi, −)	(−, +)	(Lo, +)	(Hi, +)
LLL1 λCI loop interference on LacI	*oR* P_{RM} 10	600	*Oid* 11	600 4	*oL* 12	600 2	$P_{lac^{UV5}}$ *O2* 1	NA	NA	449	450	451	452	453	454
LLL2 Unloopable λCI outside LacI loop	*oR* P_{RM} 10	600	*Oid* 11	600 4	O^- 3	600 2	$P_{lac^{UV5}}$ *O2* 1	NA	NA	455	456	457	458	459	460
LLL3 Unloopable λCI within LacI loop	O^- 14	600 13	*Oid* 11	600 4	*oL* 12	600 2	$P_{lac^{UV5}}$ *O2* 1	NA	NA	461	462	463	464	465	466
LLL4 LacIloop interference on λCI	*oR* P_{RM} Flip PCR from LLL1 linear assembly	600	*Oid*	600	*oL*	600	$P_{lac^{UV5}}$ *O2*	NA	833	550	551	552	553	554	555
loop **LLL5** λCI distal minus	*oR* P_{RM} Flip PCR from LLL2 linear assembly reaction	600	*Oid*	600	O^-	600	$P_{lac^{UV5}}$ *O2*	NA	835	556	557	558	559	560	561
LLL6 Unloopable LacI within λCI loop	*oR* P_{RM} 23	600	*Oid* 11	600 4	*oL* 12	600 2	$P_{lac^{UV5}}$ O^- 22	NA	837	562	563	564	565	566	567

Table 6.4 (Continued on next page)

Table 6.4 (Continued from previous page)

Description	M4	S3	M3	S2	M2	S1	M1	pUC	pIT	(−, −)	(Lo, −)	(Hi, −)	(−, +)	(Lo, +)	(Hi, +)
LLL7 TPM: Interference construct	*oR* P_{RM} 31	1500 32	*Oid* 11	300 33	*oL* 12	900 pUC–HF–LL36	P^{-}_{lacUV5} *O1*	652	NA	NA	NA	NA	NA	NA	NA
LLL8 TPM: Control	*oR* P_{RM} 31	1500 32	*Oid* 11	300 33	O^{-} 3	900 pUC–HF–LL36	P^{-}_{lacUV5} *O1*	654	NA	NA	NA	NA	NA	NA	NA
LLL9 TPM: Control	O^{-} 14	1500 32	*Oid* 11	300 33	*oL* 12	900 pUC–HF–LL36	P^{-}_{lacUV5} *O1*	656	NA	NA	NA	NA	NA	NA	NA
LLL10 λCI interference on LacI loop	*oR* P_{RM} Chassis PCR from LLL7	1500	*Oid*	300	*oL*	900 1	P_{lacUV5} *O2*	NA	694	700	701	702	703	704	705
LLL12 Unloopable λCI within LacI loop	O^{-} Chassis PCR from LLL9	1500	*Oid*	300	*oL*	900 1	P_{lacUV5} *O2*	NA	698	706	707	708	709	710	711
LLL13 Unloopable LacI outside λCI loop	*oR* P_{RM} 23	600 36	O^{-} PCR from LLL1 (1079/1029)	600	*oL*	600	P_{lacUV5} *O2*	NA	725	730	731	732	733	734	735

Table 6.4 (Continued on next page)

Table 6.4 (Continued from previous page)

Description	M4	S3	M3	S2	M2	S1	M1	pUC	pIT	(−, −)	(Lo, −)	(Hi, −)	(−, +)	(Lo, +)	(Hi, +)
LLL14 LacI interference on λCI loop	*oR* P_{RM}	1500	*Oid*	300	*oL*	900	P_{lacUV5} *O2*	NA	727	736	737	738	739	740	741
	PCR from pIT–HF–LLL10 (1079/1080)														
LLL15 Unloopable LacI within λCI loop	*oR* P_{RM}	1500	*Oid*	300	*oL*	900	P_{lacUV5} O^-	NA	729	742	743	744	745	746	747
	PCR from pIT–HF–LLL10 (1080/1028)						22								
LLL16d pre–construct prior to LLL16	*Oid*	300	*oL*	1400	O^-	300	P_{lacUV5} *O2*	NA	771	NA	NA	NA	NA	NA	NA
	37	38	43	9	3	40	1								
LLL16 λCI loop within a LacI loop	*Oid*	300	*oL*	1400	*oR**	300	P_{lacUV5} *O2*	NA	813	815	816	817	818	819	820
	37	38	43	9	39	40	1								
LLL17 Lac loop within λCI loop	*oR* P_{RM}	300	*Oid*	1400	*O1*	300	*oL*	828	839	773	774	775	776	777	778
	42	38	11	9	46	40	44								
LLL18 Side–by–side λCI and LacI loops	*oR* P_{RM}	1500	*oL*	300	*Oid*	300	P_{lacUV5} *O2*	NA	809	779	780	781	782	783	784
	41	32	43	33	45	40	1								

Table 6.4 (Continued on next page)

Table 6.4 (Continued from previous page)

Description	M4	S3	M3	S2	M2	S1	M1	pUC	pIT	(−, −)	(Lo, −)	(Hi, −)	(−, +)	(Lo, +)	(Hi, +)
LLL19 Side–by–side λCI and LacI loops	*oR* P_{RM}	1500	*oL*	300	*Oid*	300	$P_{lac^{UV5}}$ *O2*	NA	811	785	786	787	788	789	790
	Flip PCR from LLL18 linear assembly reaction														
LLL4new LacIloop interference on λCI loop	*oR* P_{RM} tW 47	600	*Oid*	600	*oL*	600	$P_{lac^{UV5}}$ *O2*	NA	841	855	856	NA	857	858	NA
	pIT–HF–LLL4 digest														
LLL5new $P_{lac^{UV5}}$ *O2*	*oR* P_{RM} tW 47	600	*Oid*	600	O^-	600	$P_{lac^{UV5}}$ *O2*	NA	843	859	860	NA	861	862	NA
	pIT–HF–LLL5 digest														
LLL6new Unloopable LacI within λCI loop	*oR* P_{RM} tW 47	600	*Oid*	600	*oL*	600	$P_{lac^{UV5}}$ O^-	NA	845	863	864	NA	865	866	NA
	pIT–HF–LLL6 digest														
LLL13new Unloopable LacI outside λCI loop	*oR* P_{RM} tW 47	600	O^-	600	*oL*	600	$P_{lac^{UV5}}$ *O2*	NA	847	867	868	NA	869	870	NA
	pIT–HF–LLL13 digest														
LLL14new LacI interference on λCI loop	*oR* P_{RM} tW 47	1500	*Oid*	300	*oL*	900	$P_{lac^{UV5}}$ *O2*	NA	849	871	872	NA	873	874	NA
	pIT–HF–LLL14 digest														

Table 6.4 (Continued on next page)

Table 6.4 (Continued from previous page)

Description	M4	S3	M3	S2	M2	S1	M1	pUC	pIT	(−, −)	(Lo, −)	(Hi, −)	(−, +)	(Lo, +)	(Hi, +)
LLL17new Lac loop	*oR* P_{RM} tW	300	*Oid*	1400	*O1*	300	*oL*	NA	851	875	876	NA	877	878	NA
within λCI loop	48	pIT–HF–LLL17 digest													
LLL19new	*oR* P_{RM} tW	1500	*oL*	300	*Oid*	300	$P_{lac^{UV5}}$ *O2*	NA	853	879	880	NA	881	882	NA
Side–by–side new	47	pIT–HF–LLL19 digest													

Table 6.4: Details for LacI and λCI looping strains

Table 6.5 **Gibson Assembly Fragments**

Frag ID	DNA Contents (5'–3')	Left Primer	Right Primer	Template DNA
GBA1	Mod1($P_{lac^{UV5}}$ *O2*)	1023	1022	pUC–DS3
GBA2	Sp1(600)	1030	1028	*E. coli* Genomic DNA prep
GBA3	Mod2(O^-)	1025	1024	pUC–DS2 (or pUC–HFLL17)
GBA4	Sp2(600)	1029	1031	*E. coli* Genomic DNA prep
GBA5	Mod3(*Oid*)–Sp3(300)–Mod4(O^-)	1027	1026	pUCHF–LL17
GBA6	Sp(900)	1041	1028	*E. coli* Genomic DNA prep
GBA7	Sp2(900)	1029	1043	*E. coli* Genomic DNA prep
GBA8	Sp(1800)	1042	1028	*E. coli* Genomic DNA prep
GBA9	Sp2(1400)	1029	1044	*E. coli* Genomic DNA prep
GBA10	Mod4(*oR* P_{RM})–Sp3(600)	1027	1045	pUC–DS2
GBA11	Mod3(*Oid*)	1046	1026	pUC–DS5
GBA12	Mod2(*oL*)	1025	1024	pUC–DS3
GBA13	Sp3(600)	1048	1045	pUC–DS2
GBA14	Mod4(O^-)	1027	1047	pUC–DS1
GBA15	Mod3(O^-)–Sp3(300)–Mod4(O^-)	1027	1026	pUCHF–LL15
GBA16	Mod4(O^-)–Sp3(300)–Mod3(O^-)–Sp2(300)–Mod2(*Oid*)	1027	1024	pUCHF–LL16
GBA17	Sp1(400)	1068	1028	*E. coli* Genomic DNA prep

Table 6.5 Continued on next page

Table 6.5 (Continued from last page)

Frag ID	DNA Contents (5'–3')	Left Primer	Right Primer	Template DNA
GBA18	Sp1(500)	1069	1028	*E. coli* Genomic DNA prep
GBA19	Space4 5–6kb	1070	1071	Cui's 10kb pIT plasmid
GBA20	Space4 10 kb	1070	1072	Cui's 10kb pIT plasmid
GBA21	Mod4(*Oid*)	1027	1047	pUC–DS4
GBA22	Mod1(P_{lacUV5} O^-) flipped	1023	1079	pUC–DS2
GBA23	Mod4(*oR* P_{RM})–Sp3(600) flipped	1080	1045	pUC–DS2
GBA24	Mod1(P_{lacUV5} – *O1*)	1023	1022	pUC–HF–LL15
GBA25	Mod4(O^-)–Sp3(300)–Mod3(*Oid*)–Sp2(600)–Mod1(O^-)–Sp1(600)	1027	1028	LL27 Linear Gibson PCR
GBA26	Mod4(O^-)–Sp3(300)–Mod3(*Oid*)–Sp2(900)–Mod1(O^-)–Sp1(900)	1027	1028	LL28 Linear Gibson PCR
GBA27	Mod1(P_{lacUV5} *O2* new)	1114	1022	pUC57–LL8 new
GBA28	empty			
GBA29	Sp1new(300)	1116	1115	pUC–DS2
GBA30	Sp1new(600)	1117	1115	pUC–DS2
GBA31	Mod4(*oR* P_{RM})	1027	1047	pUC–DS2
GBA32	Sp3(1500)	1130	1129	pIT–HF–Cui 224
GBA33	Sp2(300)	1029	1131	pUC–HF–LL15
GBA34	Sp1new(900)	1153	1152	*E. coli* Genomic DNA prep
GBA35	Sp1new(1800)	1154	1152	*E. coli* Genomic DNA prep
GBA36	Mod3(O^-)	1046	1026	pUC–DS1

Table 6.5 Continued on next page

Table 6.5 (Continued from last page)

Frag ID	DNA Contents (5'–3')	Left Primer	Right Primer	Template DNA
GBA37	Mod4(*Oid*)	1027	1047	pUC–DS4
GBA38	Sp3(300)	1048	1227	pUC–DS2
GBA39	Mod2(*oR* P_{RM})	1225	1226	pUC–DS2
GBA40	Sp1(300)	1228	1028	pUC–HF–LL35
GBA41	Mod4(*oR* P_{RM})	1027	1047	pUC–DS2
GBA42	Mod4(*oR* P_{RM}) flipped	1080	1047	pUC–DS2
GBA43	Mod3(*oL*)	1046	1026	pUC–DS2
GBA44	Mod1(*oL*) flipped	1023	1079	pUC–DS5
GBA45	Mod2(*Oid*)	1025	1024	pUC–DS1
GBA46	Mod2(*O1*)	1025	1024	pUC–DS4
GBA47	tW fragment for LLLs	1306	1118	pUC57–LLL8new
GBA48	tW fragment for LLL17 only	1306	1307	pUC57–LLL8new

Chemicals

All chemicals were of analytical grade or the highest purity possible

Table 6.6 **Chemicals used in this study**

Name	Abrv.	Company	Description
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside	X-gal	Sigma Chemical Co.	Stock solutions at 30 mg/mL in dimethyl formamide were kept at -20°C.
Acetic acid		B.D.H. Labs., Australia.	

Table 6.6 (Continued on next page)

Table 6.6 (Continued from last page)

Name	**Abrv.**	**Company**	**Description**
Agarose		Sigma Chemical Co.	
Ammonium acetate		B.D.H. Labs., Australia.	
Ampicillin	Amp	Sigma Chemical Co.	Stock solutions of sodium salt 25-100 mg/mL in H_2O) were millipore filtered and stored at -20°C.
β-Mercaptoethanol	βME	Sigma Chemical Co.	
Bovine serum albumin	BSA	Sigma Chemical Co.	Kept as a 10 mg/mL solution in H_2O at -20°C.
Bromophenol blue		B.D.H. Labs., Australia.	
Calcium chloride	$CaCl_2$	Sigma Chemical Co.	
Chloramphenicol	Chlor	Sigma Chemical Co.	Stock solutions (30 mg/mL in ethanol) were stored at -20°C.
Deoxyribonucleoside triphosphates	dNTPs	Sigma Chemical Co.	Stock solutions at 20 mM (in 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) were kept at -20°C.

Table 6.6 (Continued on next page)

Table 6.6 (Continued from last page)

Name	**Abrv.**	**Company**	**Description**
Ethanol (95%) (RNase-free)		Crown Scientific.	
Ethylenediaminetetraacetic acid (Disodium Salt)	EDTA	Sigma Chemical Co.	
Glucose		Ajax.	
Glycerol		B.D.H. Labs., Australia.	
Glycogen		Boehringer Mannheim.	
Hydrochloric acid	HCl	B.D.H. Labs., Australia.	
Imidazole Hydrochloride		Aldrich Chemical Co.	Stock solutions (4M in H_2O) were prepared and stored at room temperature.
Isopropanol		May and Baker Ltd.	
Isopropyl-β-D-thiogalactopyranoside	IPTG	Sigma Chemical Co.	Stock solutions (1M in H_2O) were millipore filtered and stored at -20°C.
Kanamycin	Kan	Sigma Chemical Co.	Stock solutions (50 mg/mL in H_2O) were millipore filtered and stored at -20°C.
Magnesium chloride	$MgCl_2$	Ajax.	

Table 6.6 (Continued on next page)

Table 6.6 (Continued from last page)

Name	**Abrv.**	**Company**	**Description**
Magnesium sulfate	$MgSO_4$	Ajax.	
Nickel sulfate	$NiSO_4$	Sigma	
O-nitrophenyl-β-D-galactopyranoside	ONPG	Diagnostic Chemicals Ltd.	Used as a freshly made 4 mg/mL solution in TZ8 buffer (for microtitre plate LacZ assays).
Polymyxin-B sulfate		Sigma Chemical Co.	Stored as 20 mg/mL solution in H_20 at -20°C.
Potassium acetate	KAc	B.D.H. Labs., Australia.	
Potassium chloride	KCl	B.D.H. Labs., Australia.	
Sodium acetate	NaAc	B.D.H. Labs., Australia.	
Sodium chloride	NaCl	B.D.H. Labs., Australia.	
Sodium citrate	Na3 citrate	B.D.H. Labs., Australia.	
Sodium dihydrogen phosphate	NaH_2PO_4	May and Baker Ltd.	
Sodium dodecyl sulphate	SDS	Sigma Chemical Co.	
Sodium hydroxide	NaOH	Ajax.	
Spectinomycin	Spec	Sigma Chemical Co.	Stock solutions (50 mg/mL in H_2O) were millipore filtered and stored at -20°C.

Table 6.6 (Continued on next page)

Table 6.6 (Continued from last page)

Name	**Abrv.**	**Company**	**Description**
Tetracycline	Tet	Upjohn Pty Ltd.	Stock solutions (10 mg/mL in ethanol) were stored at -20°C.
Tris acetate		B.D.H. Labs., Australia.	
Urea		Merck.	Stock solutions (8M in H_2O) were prepared and stored at room temperature.
Xylene cyanol		Sigma Chemical Co.	

6.2.4 Media and Buffers

- LB (1% Bacto-tryptone, 1% NaCl, 0.5% yeast extract, pH 7.0)
- TZ8 (100 mM Tris HCl, pH 8.0, 1 mM $MgSO_4$, 10 mM KCl)

Chapter 7

References

Ameres, S. L., Drueppel, L., Pfleiderer, K., et al. Inducible DNA-loop formation blocks transcriptional activation by an SV40 enhancer. *EMBO J*, 24(2):358–67, Jan 2005. doi: 10.1038/sj.emboj.7600531.

Amit, R., Garcia, H. G., Phillips, R., and Fraser, S. E. Building enhancers from the ground up: a synthetic biology approach. *Cell*, 146(1):105–18, Jul 2011. doi: 10.1016/j.cell.2011.06.024.

Anderson, L. M. and Yang, H. DNA looping can enhance lysogenic CI transcription in phage lambda. *Proc Natl Acad Sci U S A*, 105(15):5827–32, Apr 2008. doi: 10.1073/pnas.0705570105.

Andersson, R., Gebhard, C., Miguel-Escalada, I., et al. An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493):455–61, Mar 2014. doi: 10.1038/nature12787.

Andrey, G., Montavon, T., Mascrez, B., et al. A switch between topological domains underlies

HoxD genes collinearity in mouse limbs. *Science*, 340(6137):1234167, Jun 2013. doi: 10.1126/science.1234167.

Arnold, C. D., Gerlach, D., Stelzer, C., et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*, 339(6123):1074–7, Mar 2013. doi: 10.1126/science.1232542.

Banerji, J., Rusconi, S., and Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27(2 Pt 1):299–308, Dec 1981.

Becker, N. A., Kahn, J. D., and Maher, L. J., 3rd. Bacterial repression loops require enhanced DNA flexibility. *J Mol Biol*, 349(4):716–30, Jun 2005a. doi: 10.1016/j.jmb.2005.04.035.

Becker, N. A., Kahn, J. D., and Maher, L. J., 3rd. Bacterial repression loops require enhanced DNA flexibility. *J Mol Biol*, 349(4):716–30, Jun 2005b. doi: 10.1016/j.jmb.2005.04.035.

Becker, N. A., Kahn, J. D., and Maher, L. J., 3rd. Effects of nucleoid proteins on DNA repression loop formation in Escherichia coli. *Nucleic Acids Res*, 35(12):3988–4000, 2007. doi: 10.1093/nar/gkm419.

Beckwith, J. The operon as paradigm: normal science and the beginning of biological complexity. *J Mol Biol*, 409(1):7–13, May 2011. doi: 10.1016/j.jmb.2011.02.027.

Bell, A. C. and Felsenfeld, G. Stopped at the border: boundaries and insulators. *Curr Opin Genet Dev*, 9(2):191–8, Apr 1999. doi: 10.1016/S0959-437X(99)80029-X.

Bellen, H. J., O'Kane, C. J., Wilson, C., et al. P-element-mediated enhancer detection: a versatile method to study development in Drosophila. *Genes Dev*, 3(9):1288–300, Sep 1989.

Blanton, J., Gaszner, M., and Schedl, P. Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev*, 17(5):664–75, Mar 2003. doi: 10.1101/gad.1052003.

Bondarenko, V. A., Jiang, Y. I., and Studitsky, V. M. Rationally designed insulator-like elements can block enhancer action in vitro. *EMBO J*, 22(18):4728–37, Sep 2003. doi: 10.1093/emboj/cdg468.

Bourgeois, S. and Riggs, A. D. The lac repressor-operator interaction. IV. Assay and purification of operator DNA. *Biochem Biophys Res Commun*, 38(2):348–54, Jan 1970.

Bulger, M. and Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell*, 144(3):327–39, Feb 2011. doi: 10.1016/j.cell.2011.01.024.

Bush, M. and Dixon, R. The role of bacterial enhancer binding proteins as specialized activators of σ54-dependent transcription. *Microbiol Mol Biol Rev*, 76(3):497–529, Sep 2012. doi: 10.1128/MMBR.00006-12.

Byrd, K. and Corces, V. G. Visualization of chromatin domains created by the gypsy insulator of Drosophila. *J Cell Biol*, 162(4):565–74, Aug 2003. doi: 10.1083/jcb.200305013.

Cai, H. and Levine, M. Modulation of enhancer-promoter interactions by insulators in the Drosophila embryo. *Nature*, 376(6540):533–6, Aug 1995. doi: 10.1038/376533a0.

Cai, H. N. and Shen, P. Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. *Science*, 291(5503):493–5, Jan 2001. doi: 10.1126/science.291.5503.493.

Callen, B. P., Shearwin, K. E., and Egan, J. B. Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol Cell*, 14(5):647–656, 2004. ISSN 1097-2765 (Print). doi: 10.1016/j.molcel.2004.05.010.

Capelson, M. and Corces, V. G. Boundary elements and nuclear organization. *Biol Cell*, 96(8): 617–29, Oct 2004. doi: 10.1016/j.biolcel.2004.06.004.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y.-f., and Fraser, P. Long-range chromatin regulatory interactions in vivo. *Nat Genet*, 32(4):623–6, Dec 2002. doi: 10.1038/ng1051.

Chen, B., Gilbert, L. A., Cimini, B. A., et al. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell*, 155(7):1479–91, Dec 2013. doi: 10.1016/j.cell.2013.12.001.

Chetverina, D., Aoki, T., Erokhin, M., Georgiev, P., and Schedl, P. Making connections: Insulators organize eukaryotic chromosomes into independent cis-regulatory networks. *Bioessays*, 36(2):163–72, Feb 2014. doi: 10.1002/bies.201300125.

Chung, C. T., Niemela, S. L., and Miller, R. H. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A*, 86(7):2172–5, Apr 1989.

Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A., and Weiss, R. Realizing the potential of synthetic biology. *Nat Rev Mol Cell Biol*, 15(4):289–94, Apr 2014. doi: 10.1038/nrm3767.

Cisse, I. I., Izeddin, I., Causse, S. Z., et al. Real-time dynamics of RNA polymerase II clustering in live human cells. *Science*, 341(6146):664–7, Aug 2013. doi: 10.1126/science.1239053.

Comet, I., Schuettengruber, B., Sexton, T., and Cavalli, G. A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. *Proc Natl Acad Sci U S A*, 108(6):2294–9, Feb 2011. doi: 10.1073/pnas.1002059108.

Corces, V. G. Chromatin insulators. Keeping enhancers under control. *Nature*, 376(6540): 462–3, Aug 1995. doi: 10.1038/376462a0.

Cremer, T. and Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, 2(4):292–301, Apr 2001. doi: 10.1038/35066075.

Cremer, T. and Cremer, M. Chromosome territories. *Cold Spring Harb Perspect Biol*, 2(3): a003889, Mar 2010. doi: 10.1101/cshperspect.a003889.

Cui, L., Murchland, I., Shearwin, K. E., and Dodd, I. B. Enhancer-like long-range transcriptional activation by λ CI-mediated DNA looping. *Proc Natl Acad Sci U S A*, 110(8): 2922–7, Feb 2013. doi: 10.1073/pnas.1221322110.

Culard, F. and Maurizot, J. C. Lac repressor - lac operator interaction. Circular dichroism study. *Nucleic Acids Res*, 9(19):5175–84, Oct 1981.

Datsenko, K. A. and Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia

coli K-12 using PCR products. *Proc Natl Acad Sci U S A*, 97(12):6640–5, Jun 2000. doi: 10.1073/pnas.120163297.

Dawkins, R. *The Selfish Gene*. Oxford University Press, Oxford, UK, 1976.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. Capturing chromosome conformation. *Science*, 295(5558):1306–11, Feb 2002. doi: 10.1126/science.1067799.

Deng, W., Lee, J., Wang, H., et al. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, 149(6):1233–44, Jun 2012. doi: 10.1016/j.cell.2012.03.051.

Dillon, S. C. and Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol*, 8(3):185–95, Mar 2010. doi: 10.1038/nrmicro2261.

Dixon, J. R., Selvaraj, S., Yue, F., et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, Apr 2012. doi: 10.1038/nature11082.

Dodd, I. B., Perkins, A. J., Tsemitsidis, D., and Egan, J. B. Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. *Genes Dev*, 15(22):3013–3022, 2001. ISSN 0890-9369 (Print). doi: 10.1101/gad.937301.

Dodd, I. B., Shearwin, K. E., Perkins, A. J., et al. Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev*, 18(3):344–354, 2004. ISSN 0890-9369 (Print). doi: 10.1101/gad.1167904.

Dorman, C. J. DNA supercoiling and bacterial gene expression. *Sci Prog*, 89(Pt 3-4):151–66, 2006.

Dunlap, D., Zurla, C., Manzo, C., and Finzi, L. Probing DNA topology using tethered particle motion. *Methods Mol Biol*, 783:295–313, 2011. doi: 10.1007/978-1-61779-282-3_16.

Dunn, T. M., Hahn, S., Ogden, S., and Schleif, R. F. An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc Natl Acad Sci U S A*, 81(16): 5017–20, Aug 1984.

ENCODE Project Consortium, Bernstein, B. E., Birney, E., et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74, Sep 2012. doi: 10.1038/nature11247.

Englesberg, E., Irr, J., Power, J., and Lee, N. Positive control of enzyme synthesis by gene C in the L-arabinose system. *J Bacteriol*, 90(4):946–57, Oct 1965.

Esvelt, K. M., Mali, P., Braff, J. L., et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods*, 10(11):1116–21, Nov 2013. doi: 10.1038/nmeth.2681.

Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M. S., and Mhlanga, M. M. Chromosomal contact permits transcription between coregulated genes. *Cell*, 155(3):606–20, Oct 2013. doi: 10.1016/j.cell.2013.09.051.

Garcia, H. G. and Phillips, R. Quantitative dissection of the simple repression input-output function. *Proc Natl Acad Sci U S A*, 108(29):12173–8, Jul 2011. doi: 10.1073/pnas.1015616108.

Geyer, P. K. and Clark, I. Protecting against promiscuity: the regulatory role of insulators. *Cell Mol Life Sci*, 59(12):2112–27, Dec 2002.

Geyer, P. K. and Corces, V. G. DNA position-specific repression of transcription by a Drosophila zinc finger protein. *Genes Dev*, 6(10):1865–73, Oct 1992.

Geyer, P. K., Green, M. M., and Corces, V. G. Reversion of a gypsy-induced mutation at the yellow (y) locus of Drosophila melanogaster is associated with the insertion of a newly defined transposable element. *Proc Natl Acad Sci U S A*, 85(11):3938–42, Jun 1988.

Ghirlando, R., Giles, K., Gowher, H., et al. Chromatin domains, insulators, and the regulation of gene expression. *Biochim Biophys Acta*, Feb 2012. doi: 10.1016/j.bbagrm.2012.01.016.

Gibson, D. G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol*, 498: 349–61, 2011. doi: 10.1016/B978-0-12-385120-8.00015-2.

Gilbert, L. A., Larson, M. H., Morsut, L., et al. CRISPR-mediated modular RNA-guided

regulation of transcription in eukaryotes. *Cell*, 154(2):442–51, Jul 2013. doi: 10.1016/j.cell.2013.06.044.

Gilbert, N. and Allan, J. Supercoiling in DNA and chromatin. *Curr Opin Genet Dev*, 25C: 15–21, Dec 2013. doi: 10.1016/j.gde.2013.10.013.

Gilbert, W. and Maxam, A. The nucleotide sequence of the lac operator. *Proc Natl Acad Sci U S A*, 70(12):3581–4, Dec 1973.

Gilbert, W. and Müller-Hill, B. Isolation of the lac repressor. *Proc Natl Acad Sci U S A*, 56(6): 1891–8, Dec 1966.

Gohl, D., Aoki, T., Blanton, J., et al. Mechanism of chromosomal boundary action: roadblock, sink, or loop? *Genetics*, 187(3):731–48, Mar 2011. doi: 10.1534/genetics.110.123752.

Griffith, J., Hochschild, A., and Ptashne, M. DNA loops induced by cooperative binding of lambda repressor. *Nature*, 322(6081):750–2, 1986. doi: 10.1038/322750a0.

Gyurkovics, H., Gausz, J., Kummer, J., and Karch, F. A new homeotic mutation in the Drosophila bithorax complex removes a boundary separating two domains of regulation. *EMBO J*, 9(8):2579–85, Aug 1990.

Haldimann, A. and Wanner, B. L. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol*, 183(21): 6384–93, Nov 2001. doi: 10.1128/JB.183.21.6384-6393.2001.

Han, L., Garcia, H. G., Blumberg, S., et al. Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS One*, 4(5):e5621, 2009. doi: 10.1371/journal.pone.0005621.

Hochschild, A. and Ptashne, M. Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. *Cell*, 44(5):681–7, Mar 1986.

Holdridge, C. and Dorsett, D. Repression of hsp70 heat shock gene transcription by the suppressor of hairy-wing protein of Drosophila melanogaster. *Mol Cell Biol*, 11(4): 1894–900, Apr 1991.

Holwerda, S. J. B. and de Laat, W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc Lond B Biol Sci*, 368(1620):20120369, 2013. doi: 10.1098/rstb.2012.0369.

Hou, C., Zhao, H., Tanimoto, K., and Dean, A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci U S A*, 105(51):20398–403, Dec 2008. doi: 10.1073/pnas.0808506106.

Hou, C., Li, L., Qin, Z. S., and Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell*, 48(3): 471–84, Nov 2012. doi: 10.1016/j.molcel.2012.08.031.

Hsieh, W. T., Whitson, P. A., Matthews, K. S., and Wells, R. D. Influence of sequence and distance between two operators on interaction with the lac repressor. *J Biol Chem*, 262(30): 14583–91, Oct 1987.

Irimia, M., Tena, J. J., Alexis, M. S., et al. Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res*, 22(12):2356–67, Dec 2012. doi: 10.1101/gr.139725.112.

Irimia, M., Maeso, I., Roy, S. W., and Fraser, H. B. Ancient cis-regulatory constraints and the evolution of genome architecture. *Trends Genet*, 29(9):521–8, Sep 2013. doi: 10.1016/j.tig.2013.05.008.

Jacob, F. and Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, 3:318–56, Jun 1961.

Jin, F., Li, Y., Dixon, J. R., et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*, Oct 2013. doi: 10.1038/nature12644.

Johnson, S., Lindén, M., and Phillips, R. Sequence dependence of transcription factor-mediated DNA looping. *Nucleic Acids Res*, 40(16):7728–38, Sep 2012. doi: 10.1093/nar/gks473.

Kaern, M., Elston, T. C., Blake, W. J., and Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet*, 6(6):451–64, Jun 2005. doi: 10.1038/nrg1615.

Kellum, R. and Schedl, P. A position-effect assay for boundaries of higher order chromosomal domains. *Cell*, 64(5):941–50, Mar 1991.

Kieffer-Kwon, K.-R., Tang, Z., Mathe, E., et al. Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell*, 155(7):1507–20, Dec 2013. doi: 10.1016/j.cell.2013.11.039.

Kikuta, H., Laplante, M., Navratilova, P., et al. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res*, 17(5): 545–55, May 2007. doi: 10.1101/gr.6086307.

Kim, A. and Dean, A. Chromatin loop formation in the β-globin locus and its role in globin gene transcription. *Mol Cells*, 34(1):1–5, Jul 2012. doi: 10.1007/s10059-012-0048-8.

Krivega, I. and Dean, A. Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev*, 22(2):79–85, Apr 2012. doi: 10.1016/j.gde.2011.11.001.

Kuhn, E. J. and Geyer, P. K. Genomic insulators: connecting properties to mechanism. *Curr Opin Cell Biol*, 15(3):259–65, Jun 2003.

Kuhn, E. J., Viering, M. M., Rhodes, K. M., and Geyer, P. K. A test of insulator interactions in Drosophila. *EMBO J*, 22(10):2463–71, May 2003. doi: 10.1093/emboj/cdg241.

Kyrchanova, O., Maksimenko, O., Stakhov, V., et al. Effective blocking of the white enhancer requires cooperation between two main mechanisms suggested for the insulator function. *PLoS Genet*, 9(7):e1003606, Jul 2013. doi: 10.1371/journal.pgen.1003606.

Lavelle, C. Pack, unpack, bend, twist, pull, push: the physical side of gene expression. *Curr Opin Genet Dev*, 25C:74–84, Feb 2014. doi: 10.1016/j.gde.2014.01.001.

Law, S. M., Bellomy, G. R., Schlax, P. J., and Record, M. T., Jr. In vivo thermodynamic analysis of repression with and without looping in lac constructs. Estimates of free and local lac repressor concentrations and of physical properties of a region of supercoiled plasmid DNA in vivo. *J Mol Biol*, 230(1):161–73, Mar 1993. doi: 10.1006/jmbi.1993.1133.

Le, T. B. K., Imakaev, M. V., Mirny, L. A., and Laub, M. T. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science*, Oct 2013. doi: 10.1126/science.1242059.

Lettice, L. A., Horikoshi, T., Heaney, S. J. H., et al. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A*, 99(11):7548–53, May 2002. doi: 10.1073/pnas.112212199.

Lewis, E. B. The phenomenon of position effect. *Adv Genet*, 3:73–115, 1950.

Lewis, M. The lac repressor. *C R Biol*, 328(6):521–48, Jun 2005. doi: 10.1016/j.crvi.2005.04.004.

Li, W., Notani, D., Ma, Q., et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, 498(7455):516–20, Jun 2013. doi: 10.1038/nature12210.

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950): 289–93, Oct 2009. doi: 10.1126/science.1181369.

Linn, T. and St Pierre, R. Improved vector system for constructing transcriptional fusions that ensures independent translation of lacZ. *J Bacteriol*, 172(2):1077–84, Feb 1990.

Liu, W., Ma, Q., Wong, K., et al. Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. *Cell*, 155(7):1581–95, Dec 2013. doi: 10.1016/j.cell.2013.10.056.

Lutz, R. and Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res*, 25 (6):1203–1210, 1997. ISSN 0305-1048 (Print).

Maeda, R. K. and Karch, F. Gene expression in time and space: additive vs hierarchical organization of cis-regulatory regions. *Curr Opin Genet Dev*, 21(2):187–93, Apr 2011. doi: 10.1016/j.gde.2011.01.021.

Mali, P., Yang, L., Esvelt, K. M., et al. RNA-guided human genome engineering via Cas9. *Science*, 339(6121):823–6, Feb 2013. doi: 10.1126/science.1232033.

Maniatis, T., Goodbourn, S., and Fischer, J. A. Regulation of inducible and tissue-specific gene expression. *Science*, 236(4806):1237–45, Jun 1987.

Marinić, M., Aktas, T., Ruf, S., and Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev Cell*, 24(5):530–42, Mar 2013. doi: 10.1016/j.devcel.2013.01.025.

Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., and Miller, J. H. Genetic studies of the lac repressor. XIV. Analysis of 4000 altered Escherichia coli lac repressors reveals essential and non-essential residues, as well as "spacers" which do not require a specific sequence. *J Mol Biol*, 240(5):421–33, Jul 1994. doi: 10.1006/jmbi.1994.1458.

Matthews, K. S. DNA looping. *Microbiol Rev*, 56(1):123–36, Mar 1992.

Melnikova, L., Juge, F., Gruzdeva, N., et al. Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila. *Proc Natl Acad Sci U S A*, 101(41):14806–11, Oct 2004. doi: 10.1073/pnas.0403959101.

Miller, J. H. *A Short Course in Bacterial Genetics*. Cold Spring Harbor Laboratory Press, 1992.

Miyanari, Y., Ziegler-Birling, C., and Torres-Padilla, M.-E. Live visualization of chromatin dynamics with fluorescent TALEs. *Nat Struct Mol Biol*, Oct 2013. doi: 10.1038/nsmb.2680.

Mossing, M. C. and Record, M. T., Jr. Upstream operators enhance repression of the lac promoter. *Science*, 233(4766):889–92, Aug 1986.

Müller, J., Oehler, S., and Müller-Hill, B. Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J Mol Biol*, 257(1):21–9, Mar 1996. doi: 10.1006/jmbi.1996.0143.

Müller, M. M., Gerster, T., and Schaffner, W. Enhancer sequences and the regulation of gene transcription. *Eur J Biochem*, 176(3):485–95, Oct 1988.

Müller-Hill, B. Lac repressor and lac operator. *Prog Biophys Mol Biol*, 30(2-3):227–52, 1975.

Muravyova, E., Golovnin, A., Gracheva, E., et al. Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science*, 291(5503):495–8, Jan 2001. doi: 10.1126/science.291.5503.495.

Murrell, A., Heeson, S., and Reik, W. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat Genet*, 36(8):889–93, Aug 2004. doi: 10.1038/ng1402.

Nagano, T., Lubling, Y., Stevens, T. J., et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469):59–64, Oct 2013. doi: 10.1038/nature12593.

Naughton, C., Avlonitis, N., Corless, S., et al. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat Struct Mol Biol*, 20(3):387–95, Mar 2013. doi: 10.1038/nsmb.2509.

Nolis, I. K., McKay, D. J., Mantouvalou, E., et al. Transcription factors mediate long-range enhancer-promoter interactions. *Proc Natl Acad Sci U S A*, 106(48):20222–7, Dec 2009. doi: 10.1073/pnas.0902454106.

Noordermeer, D. and Duboule, D. Chromatin looping and organization at developmentally regulated gene loci. *Wiley Interdiscip Rev Dev Biol*, 2(5):615–30, Sep 2013. doi: 10.1002/wdev.103.

Noordermeer, D., de Wit, E., Klous, P., et al. Variegated gene expression caused by cell-specific long-range DNA interactions. *Nat Cell Biol*, 13(8):944–51, Aug 2011. doi: 10.1038/ncb2278.

Nora, E. P., Lajoie, B. R., Schulz, E. G., et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398):381–5, May 2012. doi: 10.1038/nature11049.

Nord, A. S., Blow, M. J., Attanasio, C., et al. Rapid and Pervasive Changes in Genome-wide Enhancer Usage during Mammalian Development. *Cell*, 155(7):1521–31, Dec 2013. doi: 10.1016/j.cell.2013.11.033.

Oehler, S., Eismann, E. R., Krämer, H., and Müller-Hill, B. The three operators of the lac operon cooperate in repression. *EMBO J*, 9(4):973–9, Apr 1990.

Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B., and Müller-Hill, B. Quality and position of the three lac operators of E. coli define efficiency of repression. *EMBO J*, 13 (14):3348–55, Jul 1994.

Ong, C.-T. and Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*, 15(4):234–46, Apr 2014. doi: 10.1038/nrg3663.

Orom, U. A. and Shiekhattar, R. Long Noncoding RNAs Usher In a New Era in the Biology of Enhancers. *Cell*, 154(6):1190–3, Sep 2013. doi: 10.1016/j.cell.2013.08.028.

Paliy, O. and Gunasekera, T. S. Growth of E. coli BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Appl Microbiol Biotechnol*, 73(5):1169–72, Jan 2007. doi: 10.1007/s00253-006-0554-8.

Palstra, R.-J., Tolhuis, B., Splinter, E., et al. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet*, 35(2):190–4, Oct 2003. doi: 10.1038/ng1244.

Papantonis, A. and Cook, P. R. Transcription Factories: Genome Organization and Gene Regulation. *Chem Rev*, Apr 2013. doi: 10.1021/cr300513p.

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., and Bejerano, G. Enhancers: five essential questions. *Nat Rev Genet*, 14(4):288–95, Apr 2013. doi: 10.1038/nrg3458.

Peters, J. P., 3rd and Maher, L. J. DNA curvature and flexibility in vitro and in vivo. *Q Rev Biophys*, 43(1):23–63, Feb 2010. doi: 10.1017/S0033583510000077.

Phillips, J. E. and Corces, V. G. CTCF: master weaver of the genome. *Cell*, 137(7):1194–211, Jun 2009. doi: 10.1016/j.cell.2009.06.001.

Phillips-Cremins, J. E. and Corces, V. G. Chromatin insulators: linking genome organization to cellular function. *Mol Cell*, 50(4):461–74, May 2013. doi: 10.1016/j.molcel.2013.04.018.

Priest, D. G., Cui, L., Kumar, S., et al. Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and Lambda repressors. *Proc Natl Acad Sci U S A*, 111(1):349–54, Jan 2014. doi: 10.1073/pnas.1317817111.

Ptashne, M. *A genetic switch: phage lambda revisited*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., dritte edition, 2004. ISBN 0879697172 9780879697174 0879697164 9780879697167. URL `http://www.worldcat.org/search?qt=worldcat_org_all&q=9780879697167`.

Qi, L. S., Larson, M. H., Gilbert, L. A., et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–83, Feb 2013. doi: 10.1016/j.cell.2013.02.022.

Reznikoff, W. S., Winter, R. B., and Hurley, C. K. The location of the repressor binding sites in the lac operon. *Proc Natl Acad Sci U S A*, 71(6):2314–8, Jun 1974.

Rippe, K. Making contacts on a nucleic acid polymer. *Trends Biochem Sci*, 26(12):733–40, Dec 2001.

Roseman, R. R., Pirrotta, V., and Geyer, P. K. The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects. *EMBO J*, 12(2): 435–42, Feb 1993.

Ruf, S., Symmons, O., Uslu, V. V., et al. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat Genet*, 43(4):379–86, Apr 2011. doi: 10.1038/ng.790.

Sabri, S., Steen, J. A., Bongers, M., Nielsen, L. K., and Vickers, C. E. Knock-in/Knock-out (KIKO) vectors for rapid integration of large DNA sequences, including whole metabolic pathways, onto the Escherichia coli chromosome at well-characterised loci. *Microb Cell Fact*, 12:60, 2013. doi: 10.1186/1475-2859-12-60.

Sander, J. D. and Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*, Mar 2014. doi: 10.1038/nbt.2842.

Sanyal, A., Lajoie, B. R., Jain, G., and Dekker, J. The long-range interaction landscape of gene promoters. *Nature*, 489(7414):109–13, Sep 2012. doi: 10.1038/nature11279.

Savitskaya, E., Melnikova, L., Kostuchenko, M., et al. Study of long-distance functional interactions between Su(Hw) insulators that can regulate enhancer-promoter communication in Drosophila melanogaster. *Mol Cell Biol*, 26(3):754–61, Feb 2006. doi: 10.1128/MCB.26.3.754-761.2006.

Schleif, R. DNA looping. *Annu Rev Biochem*, 61:199–223, 1992. doi: 10.1146/annurev.bi.61.070192.001215.

Schoborg, T., Rickels, R., Barrios, J., and Labrador, M. Chromatin insulator bodies are nuclear structures that form in response to osmotic stress and cell death. *J Cell Biol*, 202(2):261–76, Jul 2013. doi: 10.1083/jcb.201304181.

Sellitti, M. A., Pavco, P. A., and Steege, D. A. lac repressor blocks in vivo transcription of lac control region DNA. *Proc Natl Acad Sci U S A*, 84(10):3199–203, May 1987.

Sexton, T., Yaffe, E., Kenigsberg, E., et al. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148(3):458–72, Feb 2012. doi: 10.1016/j.cell.2012.01.010.

Shen, Y., Yue, F., McCleary, D. F., et al. A map of the cis-regulatory sequences in the mouse genome. *Nature*, 488(7409):116–20, Aug 2012. doi: 10.1038/nature11243.

Shimada, J. and Yamakawa, H. RING-CLOSURE PROBABILITIES FOR TWISTED WORMLIKE CHAINS - APPLICATION TO DNA. *MACROMOLECULES*, 17(4):689–698, 1984. ISSN 0024-9297. doi: {10.1021/ma00134a028}.

Shlyueva, D., Stampfel, G., and Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*, 15(4):272–86, Apr 2014. doi: 10.1038/nrg3682.

Shore, D. and Baldwin, R. L. Energetics of DNA twisting. I. Relation between twist and cyclization probability. *J Mol Biol*, 170(4):957–81, Nov 1983.

Shore, D., Langowski, J., and Baldwin, R. L. DNA flexibility studied by covalent closure of short fragments into circles. *Proc Natl Acad Sci U S A*, 78(8):4833–7, Aug 1981.

Shuman, H. A. and Silhavy, T. J. The art and design of genetic screens: Escherichia coli. *Nat Rev Genet*, 4(6):419–31, Jun 2003. doi: 10.1038/nrg1087.

St-Pierre, F., Cui, L., Priest, D. G., et al. One-step cloning and chromosomal integration of DNA. *ACS Synth Biol*, 2(9):537–41, Sep 2013. doi: 10.1021/sb400021j.

Stergachis, A. B., Neph, S., Reynolds, A., et al. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell*, 154(4):888–903, Aug 2013. doi: 10.1016/j.cell.2013.07.020.

Sutherland, H. and Bickmore, W. A. Transcription factories: gene expression in unions? *Nat Rev Genet*, 10(7):457–66, Jul 2009. doi: 10.1038/nrg2592.

Symmons, O., Uslu, V. V., Tsujimura, T., et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res*, Feb 2014. doi: 10.1101/gr.163519.113.

Thurman, R. E., Rynes, E., Humbert, R., et al. The accessible chromatin landscape of the human genome. *Nature*, 489(7414):75–82, Sep 2012. doi: 10.1038/nature11232.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., and de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*, 10(6):1453–65, Dec 2002.

Van Bortle, K., Ramos, E., Takenaka, N., et al. Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res*, 22(11):2176–87, Nov 2012. doi: 10.1101/gr.136788.111.

Vilar, J. M. G. and Saiz, L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev*, 15 (2):136–44, Apr 2005. doi: 10.1016/j.gde.2005.02.005.

Wadman, I. A., Osada, H., Grütz, G. G., et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47,

GATA-1 and Ldb1/NLI proteins. *EMBO J*, 16(11):3145–57, Jun 1997. doi: 10.1093/emboj/16.11.3145.

Wang, H., Dodd, I. B., Dunlap, D. D., Shearwin, K. E., and Finzi, L. Single molecule analysis of DNA wrapping and looping by a circular 14mer wheel of the bacteriophage 186 CI repressor. *Nucleic Acids Res*, 41(11):5746–56, Jun 2013. doi: 10.1093/nar/gkt298.

Wang, H. H., Isaacs, F. J., Carr, P. A., et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460(7257):894–8, Aug 2009. doi: 10.1038/nature08187.

Wendt, K. S., Yoshida, K., Itoh, T., et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, 451(7180):796–801, Feb 2008. doi: 10.1038/nature06634.

Wood, A. M., Van Bortle, K., Ramos, E., et al. Regulation of chromatin organization and inducible gene expression by a Drosophila insulator. *Mol Cell*, 44(1):29–38, Oct 2011. doi: 10.1016/j.molcel.2011.07.035.

Woon Kim, Y., Kim, S., Geun Kim, C., and Kim, A. The distinctive roles of erythroid specific activator GATA-1 and NF-E2 in transcription of the human fetal γ-globin genes. *Nucleic Acids Res*, 39(16):6944–55, Sep 2011. doi: 10.1093/nar/gkr253.

Xu, J. and Matthews, K. S. Flexibility in the inducer binding region is crucial for allostery in the Escherichia coli lactose repressor. *Biochemistry*, 48(22):4988–98, Jun 2009. doi: 10.1021/bi9002343.

Yaniv, M. The 50th anniversary of the publication of the operon theory in the Journal of Molecular Biology: past, present and future. *J Mol Biol*, 409(1):1–6, May 2011. doi: 10.1016/j.jmb.2011.03.041.

Yu, D., Ellis, H. M., Lee, E. C., et al. An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A*, 97(11):5978–83, May 2000. doi: 10.1073/pnas.100127597.