Probing Allostery Through DNA

Sangjin Kim,[1]*† Erik Broströmer,[1]* Dong Xing,[2]* Jianshi Jin,[2,3]* Shasha Chong,[1] Hao Ge,[2,4] Siyuan Wang,[1] Chan Gu,[5] Lijiang Yang,[5] Yi Qin Gao,[5] Xiao-dong Su,[2]‡ Yujie Sun,[2]‡ X. Sunney Xie[1,2]‡

Allostery is well documented for proteins but less recognized for DNA-protein interactions. Here, we report that specific binding of a protein on DNA is substantially stabilized or destabilized by another protein bound nearby. The ternary complex's free energy oscillates as a function of the separation between the two proteins with a periodicity of ~10 base pairs, the helical pitch of B-form DNA, and a decay length of ~15 base pairs. The binding affinity of a protein near a DNA hairpin is similarly dependent on their separation, which—together with molecular dynamics simulations—suggests that deformation of the double-helical structure is the origin of DNA allostery. The physiological relevance of this phenomenon is illustrated by its effect on gene expression in live bacteria and on a transcription factor's affinity near nucleosomes.

Upon binding of a ligand, a macromolecule often undergoes conformational changes that modify the binding affinity of a second ligand at a distant site. This phenomenon, known as "allostery," is responsible for dynamic regulation of biological functions. Although extensive studies have been done on allostery in proteins or enzymes (*1*, *2*), less is known for that through DNA, which is normally considered as a mere template providing binding sites. In fact, multiple proteins, such as transcription factors and RNA or DNA polymerases, bind close to each other on genomic DNA to carry out their cellular functions in concert. Such allostery through DNA has been implicated in previous studies (*3*–*10*) but has not been quantitatively characterized or mechanistically understood.

We performed a single-molecule study of allostery through DNA by measuring the dissociation rate constant (k_{off}) of a DNA-bound protein affected by the binding of another protein nearby. In the assay, DNA duplexes (dsDNA), tethered on the passivated surface of a flow cell, contained two specific protein binding sites separated by a linker sequence of L base pairs (bp) (Fig. 1A and figs. S1 and S2) (*11*). One of the proteins was fluorescently labeled, and many individual protein-DNA complexes were monitored in a large field of view with a total internal reflection fluorescence microscope. Once the labeled protein molecules were bound to DNA, the second protein at a certain concentration was flowed in. Stochastic dissociation times of hundreds of labeled protein molecules were then recorded, the average of which yields the k_{off} (fig. S3) (*12*).

We first present a protein pair that does not substantially bend DNA, namely a Cy3B-labeled DNA binding domain of glucocorticoid receptor (GRDBD), a eukaryotic transcription factor, together with BamHI, a type II endonuclease (Fig. 1A) (*13*, *14*). To prevent the endonuclease activity of BamHI, we used buffer containing Ca^{2+} instead of Mg^{2+}. At a saturating concentration of BamHI, the k_{off} of GRDBD was found to oscillate as a function of L with significant amplitude spanning a factor of 4 and a periodicity of 10 bp, which intriguingly coincides with the helical pitch of B-form DNA (Fig. 1B, red).

When we reversed the DNA sequence of the nonpalindromic GRDBD binding site (GRE) with respect to that of BamHI, the k_{off} of GRDBD oscillated with a phase shift of 4 bp, nearly 180° relative to that of the forward GRE (Fig. 1B). On the other hand, the binding sequence of BamHI is palindromic; therefore, its reversion is not expected to cause any phase shift.

Similar oscillatory modulation in k_{off} was observed with other protein pairs, such as lac repressor (LacR) and EcoRV or LacR and T7 RNA polymerase (T7 RNAp) (figs. S5 and S6). These proteins differ in size, shape, surface

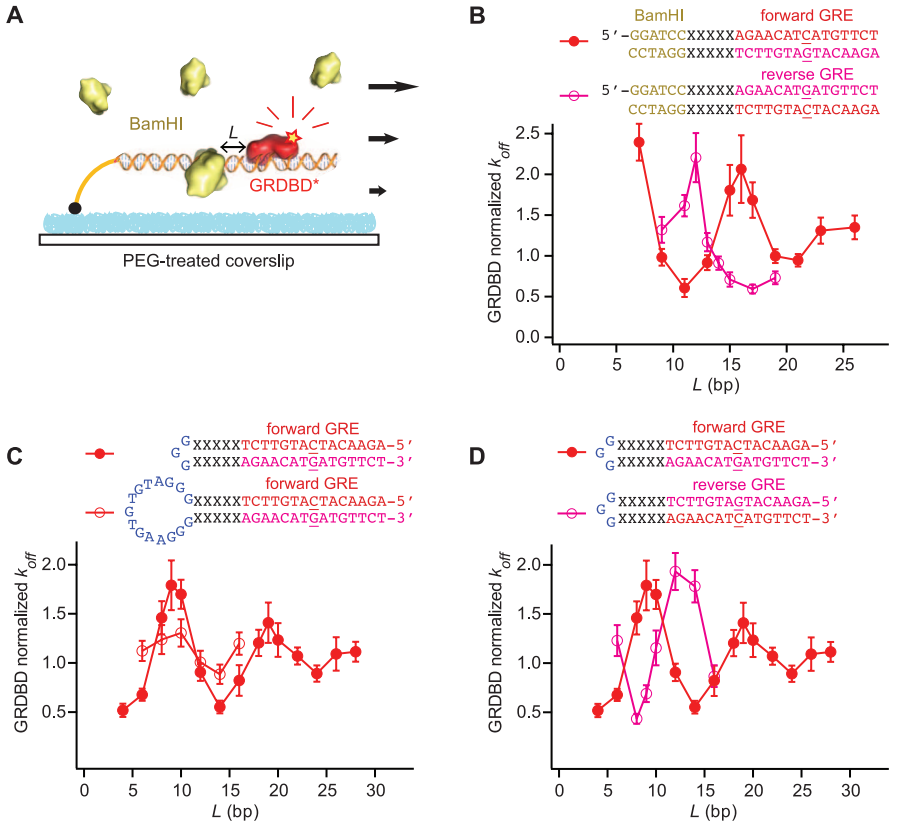

Fig. 1. Allostery through DNA affecting k_{off} of GRDBD near BamHI or near a hairpin loop. (**A**) Schematic for the single-molecule assay in a flow cell. The structural model is for L = 11 with GRDBD from Protein Data Bank (PDB) ID 1R4R (*13*) and BamHI from PDB ID 2BAM (*32*). (**B**) Oscillation in the k_{off} of GRDBD for the forward (red solid circles) and reverse (magenta open circles) GRE sequences, normalized to that measured in the absence of BamHI (±SEM). DNA sequences are shown with the linker DNA (L = 5). The central base of GRE, which makes the sequence nonpalindromic, is underlined. (**C**) Protein binding affinity affected by a nearby DNA hairpin loop, 3 bp and 15 bp (±SEM). (**D**) Effect of 3-bp loop on the forward and reverse GRE (±SEM). The DNA sequence is shown for L = 5. k_{off} is normalized to that measured on DNA without a hairpin loop.

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. [2]Biodynamic Optical Imaging Center (BIOPIC), School of Life Sciences, Peking University, Beijing 100871, China. [3]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China. [4]Beijing International Center for Mathematical Research, Peking University, Beijing 100871, China. [5]Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China.

*These authors contributed equally to this work.
†Present address: Department of Molecular Cellular and Developmental Biology, Yale University, New Haven, CT 06511, USA.
‡To whom correspondence should be addressed. E-mail: xdsu@pku.edu.cn (X.-D.S.); sun_yujie@pku.edu.cn (Y.S.); xie@chemistry.harvard.edu (X.S.X.)

charge distribution, and DNA binding affinity (*15*–*18*). In fact, the oscillation was independent of ionic strength (fig. S7), suggesting that the electrostatic interaction between the two proteins is not the origin of the allosteric phenomenon. However, the presence of a nick, mismatched bases, or GC-rich sequences in the linker region attenuated the oscillation (figs. S8 to S10), implying that the allostery is largely dependent on the mechanical properties of the linker DNA.

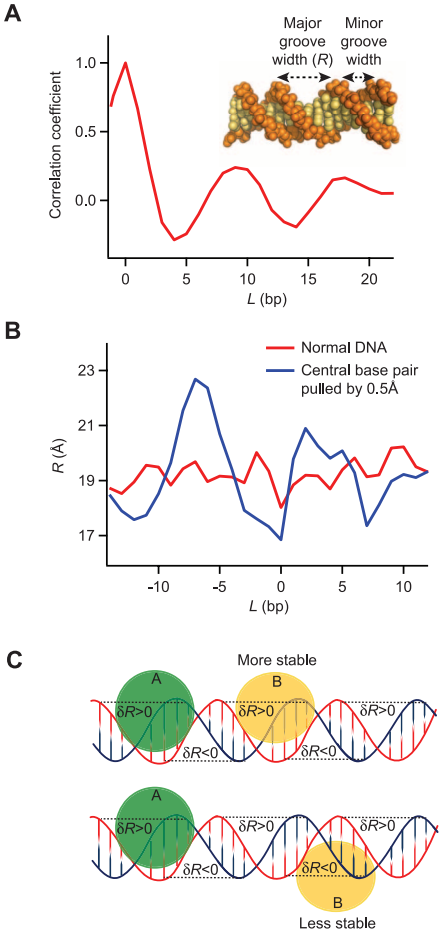

Fig. 2. Allostery through DNA induced by distortion of the major groove. (**A**) MD simulation at room temperature reveals the spatial correlation between the major groove widths (inset, defined as the distance between C3 atoms of the *i*th and *i* + 7th nucleotide sugar-rings) at two positions as a function of their separation *L*, averaged over time *t*, $[\langle \delta R(i;t) \delta R(i+L;t) \rangle_t, i = 5]$. The correlation oscillates with a periodicity of ~10 bp and is attributed to thermally excited low-frequency vibrational modes of dsDNA. (**B**) Upon breaking the symmetry by pulling apart a base pair in the middle of the dsDNA (defined as $L = 0$) by 0.5Å (*12*), the time-averaged *R* (blue) deviates from that of a free DNA (red) and oscillates as a function of the distance (*L*) from the perturbed base pair with a periodicity of ~10 bp. (**C**) Oscillation of *R*(*L*) causes the variation of the allosteric coupling between two DNA-binding proteins A and B. If protein B widens *R*, it would energetically favor binding at positions where *R* is already widened ($\delta R > 0$) by protein A (top), but disfavor where *R* is narrowed ($\delta R < 0$) (bottom).

To prove this hypothesis, we replaced the BamHI binding site with a DNA hairpin loop (Fig. 1C), which allows examination of the effect of DNA distortion alone. When the length of linker DNA between the hairpin loop and GRE (*L*) was varied, we observed a similar oscillation in the k_{off} of GRDBD (Fig. 1C). A larger hairpin loop decreased the amplitude of the oscillation, likely because of a smaller distortion induced by the larger hairpin (Fig. 1C). Again when GRE was reversed, the oscillation showed a 4-bp phase shift (Fig. 1D).

The oscillation dampens out with a characteristic decay length of ~15 bp (Fig. 1 and fig. S12) (*12*), which is much shorter than either the bending persistence length (~150 bp) (*19*) or the twisting persistence length (~300 bp) of DNA (*20*). On the other hand, recognizing that proteins primarily interact with the DNA major groove (*21*, *22*), we hypothesized that allostery through DNA results mainly from distortion of the major groove.

We carried out molecular dynamics (MD) simulations first on free dsDNA in aqueous solutions at room temperature (*12*). We evaluated the spatial correlation between the major groove widths (*R*) (Fig. 2A, inset) at two positions (base pairs *i* and *i* + *L*) as a function of their separation, averaged over time *t*. We observed that the correlation coefficient has a clear oscillation with a periodicity of ~10 bp and dampens within a few helical turns (Fig. 2A). A similar yet slightly weaker oscillation was also observed for the correlation of the minor groove widths. We attribute the oscillation in Fig. 2A to thermally excited low-frequency vibrational modes of dsDNA, which are dictated by the double-helical structure of DNA.

Such spatial correlation as well as the time-averaged *R* (Fig. 2B, red curve) are translationally invariant across a free DNA unless the symmetry is broken, as in the case of hairpin formation or protein binding. We simulated such an effect by applying a harmonic potential to pull a base pair apart in the middle of the strand. Under this condition, we observed that the time-averaged *R* (Fig. 2B, blue curve) deviates from that of a free DNA (Fig. 2B, red curve) and oscillates as a function of the distance (*L*) from the perturbed base pair with a periodicity of ~10 bp. In contrast, no such oscillation was observed for the inter-helical distance.

Such deviation from the free DNA, $\delta R(L)$, is expected to cause variation in the binding of the second DNA-binding protein at a distance *L* bases away. For example, in Fig. 2C, if protein B widens *R*, its binding would be energetically favored at positions where *R* is already widened ($\delta R > 0$) by the hairpin or protein A (Fig. 2C, top) but disfavored where *R* is narrowed ($\delta R < 0$) (Fig. 2C, bottom). Consequently, reversing a non-palindromic binding sequence would invert the binding preference of the protein, explaining the phase shift in Fig. 1. This model is also well supported by the observation that the k_{off} of LacR oscillates with an opposite phase in the presence of BamHI or EcoRV (fig. S6), which is consistent with the fact that BamHI widens whereas EcoRV narrows the major groove (*14*, *15*).

Next, to investigate the effect of DNA allostery on transcription regulation, we studied modulation of RNA polymerase binding affinity when a protein binds near the promoter both in vitro and in vivo. The protein pair we chose is LacR and T7 RNAp, both of which, unlike GRDBD and BamHI, bend DNA (*17*, *23*, *24*) but nevertheless exhibit a similar allosteric effect.

In the in vitro assay, we measured the binding affinity of unlabeled T7 RNAp on its promoter by titrating k_{off} of labeled LacR on *lac* operator O_1 (*lacO*1) with T7 RNAp. k_{off} exhibited hyperbolic T7 RNAp concentration dependence (Michaelis-Menten–like kinetics) (Fig. 3, A and B, and fig. S4), as can be rigorously derived from the kinetic scheme for LacR (protein A) and T7 RNAp (protein B) in Fig. 3C (*12*):

$$k_{off} = \frac{(k_{3\to1} + k_{3\to2})(k_{1\to0} - k_{3\to2})}{[B] \cdot k_{1\to3} + (k_{3\to1} + k_{3\to2})} + k_{3\to2} \quad (1)$$

where $k_{i\to j}$ is the rate constant from state *i* to *j* and [B] is the concentration of T7 RNAp. The plateau value in the titration curve is $k_{3\to2}$. We observed that $k_{3\to2}$ oscillates as a function of *L* with the periodicity and amplitude similar to those of GRDBD and BamHI (Fig. 3D, top).

According to Eq. 1, the dissociation constant of B on the A-bound DNA, $K^{A}_{d,B}(L)$, can be measured from the value of [B] at which k_{off} reaches half of the plateau value in the titration curves (Fig. 3, A and B) (*12*). We found that $K^{A}_{d,B}(L)$ oscillates as a function of *L* (Fig. 3D, middle) in phase with $k_{3\to2}$ [that is, $K^{B}_{d,A}(L) \cdot k_{2\to3}$] (Fig. 3D, top). Therefore, the cooperativity in DNA binding, if present, exhibits either simultaneous stabilization or destabilization between the two proteins. This is a consequence of the fact that free energy is a path-independent thermodynamic state function (Fig. 3C) (*12*):

$$\begin{aligned} \Delta G^{\circ}_{0\to3}(L) &= \Delta G^{\circ}_{1\to3}(L) - \Delta G^{\circ}_{1\to0} \\ &= k_B T \ln[K^{A}_{d,B}(L) \cdot K_{d,A}] \\ &= \Delta G^{\circ}_{2\to3}(L) - \Delta G^{\circ}_{2\to0} \\ &= k_B T \ln[K^{B}_{d,A}(L) \cdot K_{d,B}] \end{aligned} \quad (2)$$

where $K_{d,A}$ and $K_{d,B}$ are the dissociation constant of a protein in the absence of the other, k_B is the Boltzmann constant, and *T* is temperature.

Based on the second line of Eq. 2, the free energy of the ternary complex, $\Delta G^{\circ}_{0\to3}(L)$, was found to oscillate with a periodicity of ~10 bp and an amplitude of ~2 $k_B T$ (Fig. 3D, bottom).

In general, for a ternary complex formed with DNA and proteins A and B, the free energy of the overall system is $\Delta G^{\circ}_{0\to3}(L) = \Delta G^{\circ}_{A} + \Delta G^{\circ}_{B} + \Delta\Delta G^{\circ}_{AB}(L)$, where ΔG°_{A} and ΔG°_{B} are the binding free energies of the two individual proteins on DNA, respectively. $\Delta\Delta G^{\circ}_{AB}(L)$, small relative to ΔG°_{A} or ΔG°_{B}, is the energetic coupling involving in the linker DNA, given by the sum of two terms—the variation of protein A binding caused by protein B and the variation of protein B binding caused by protein A:

$$\Delta\Delta G^{\circ}_{AB}(L) \propto \delta R^{A}_{A}\delta R^{B}_{A}(L) + \delta R^{A}_{B}(L)\delta R^{B}_{B} \quad (3)$$

In each δR, or the distortion of the major groove widths, the subscripts indicate where the distortion occurs (binding site of protein A or B), and the superscripts indicate the protein that causes the DNA distortion (*12*). According to our proposed mechanism, $\delta R^{B}_{A}(L)$ and $\delta R^{A}_{B}(L)$ propagate periodically (Fig. 2B), yielding a damped oscillation in $\Delta\Delta G^{\circ}_{AB}(L)$. This explains the oscillations of the coupling energy for LacR and T7 RNAp (Fig. 3D, bottom, and fig. S14A) and for GRDBD and BamHI (fig. S14B).

The allosteric coupling between LacR and T7 RNAp is likely to affect transcription in vivo

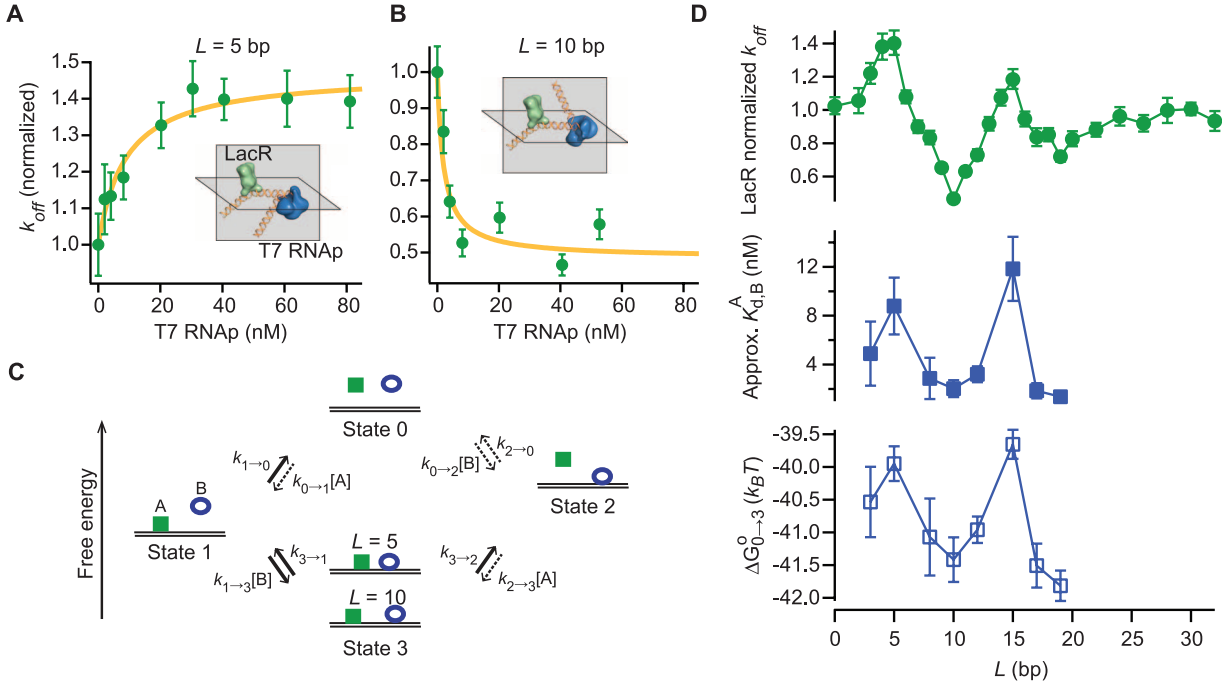

Fig. 3. Allostery through DNA between LacR and T7 RNAp in vitro. (**A** and **B**) Titration curves, where k_{off} values were normalized to those measured in the absence of T7 RNAp on the given template (±SEM). The hyperbolic fit (yellow) is based on Eq. 1. Structural models illustrate the ternary complex of LacR [PDB ID 2PE5 (*33*)] and T7 RNAp [PDB ID 3E2E (*18*)]. (**C**) Kinetic model for the binding of proteins A (LacR) and B (T7 RNAp). Our experiments start with state 1 and proceed to the dissociation of LacR to state 0 or state 2 (via state 3), as shown with solid arrows. Dashed arrows indicate reactions that are not considered in our derivation of Eq. 1. (**D**) The maximum k_{off} of LacR ($k_{3\to2}$), K_d of T7 RNAp in the presence of LacR ($K^{A}_{d,B}$), and the free energy of the ternary complex ($\Delta G^{\circ}_{0\to3}$), as function of *L*, oscillating with a periodicity of 10 bp. Error bars reflect SEM for $k_{3\to2}$ and 1 SD of the χ^2 fit for $K^{A}_{d,B}$ and $\Delta G^{\circ}_{0\to3}$ (*12*).

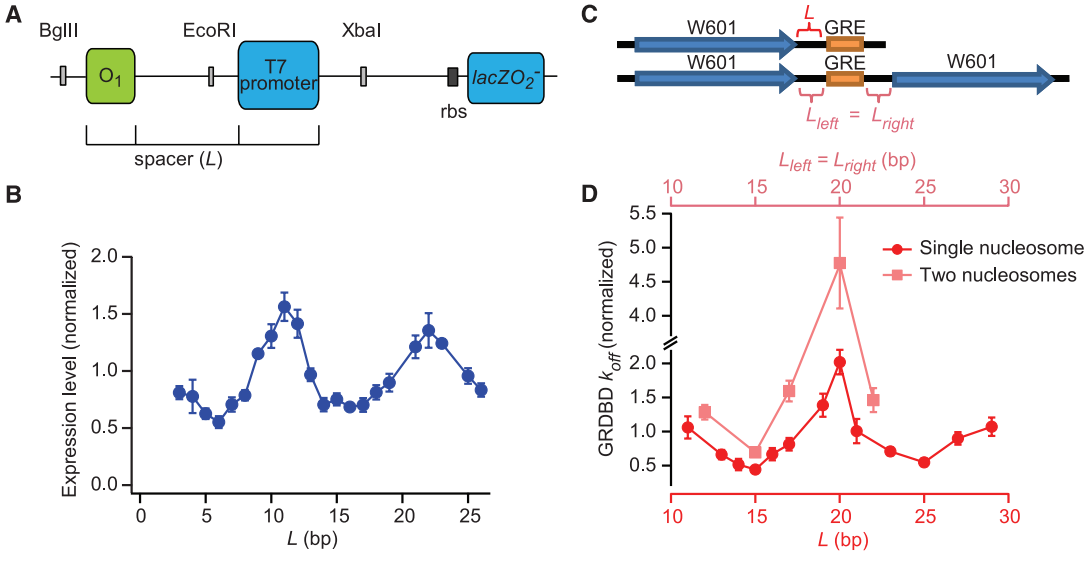

Fig. 4. Physiological relevance of DNA allostery. (**A**) *E. coli* strains constructed to examine cooperativity between LacR and T7 RNAp on the bacterial chromosome. (**B**) The expression level of *lacZ* (normalized to the average expression levels of all *Ls*) oscillates as a function of *L* with a periodicity of 10 bp, similar to the corresponding in vitro data (fig. S15). Error bars reflect SEM (n = 3 independent experiments). (**C**) Schematic for the DNA sequences used in the GRDBD-nucleosome experiment. W601 is the Widom 601 nucleosome positioning sequence (*34*). (**D**) Oscillation of the k_{off} of GRDBD as a function of *L* (±SEM). Data was normalized to k_{off} of GRDBD in the absence of histone (fig. S17).

because the efficiency of transcription initiation correlates with the binding affinity of T7 RNAp (*25*). We therefore inserted DNA templates used in vitro (fig. S15) into the chromosome of *Escherichia coli* and measured the expression level of *lacZ* using the Miller assay (Fig. 4A) (*26*). Indeed, the gene expression level oscillates as a function of L with a periodicity of ~10 bp (Fig. 4B). Similar oscillations of T7 RNAp activity were observed on plasmids in *E. coli* cells by using a yellow fluorescent protein as a reporter (fig. S16). The oscillation of gene expression levels with a 10-bp periodicity was also seen in a classic experiment on *lac* operon with a DNA loop formed by two operators (*27*). However, our T7 RNAp result illustrates that DNA allostery results in such an oscillatory phenomenon even without a DNA loop, which is consistent with a recent study in which *E. coli* RNA polymerase was used (*10*).

Pertinent to eukaryotic gene expression, DNA allostery may affect the binding affinity of transcription factors near nucleosomes that are closely positioned (*28*, *29*). We placed GRE downstream of a nucleosome (Fig. 4C) and observed a similar DNA allosteric effect in the k_{off} of GRDBD (Fig. 4D and fig. S17). To evaluate DNA allostery in an internucleosomal space, we used two nucleosomes to flank a GRE (Fig. 4C). At the same separation L, GRDBD resides on GRE for a relatively longer time with a single nucleosome nearby than it does with a pair of nucleosomes on both sides of GRE (Fig. 4D). Nonetheless, the fold change between the maximal and minimal k_{off} is larger for GRDBD with two nucleosomes (approximately sevenfold). This indicates moderately large cooperativity between the two flanking nucleosomes in modifying the binding affinity of GRDBD, which is in line with previous in vivo experiments (*30*, *31*). The fact that histones modify a neighboring transcription factor's binding suggests that allostery through DNA might be physiologically important in affecting gene regulation.

References and Notes

1. J. Monod, J. Wyman, J. P. Changeux, *J. Mol. Biol.* **12**, 88 (1965).
2. D. E. Koshland Jr., G. Némethy, D. Filmer, *Biochemistry* **5**, 365 (1966).
3. F. M. Pohl, T. M. Jovin, W. Baehr, J. J. Holbrook, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3805 (1972).
4. M. Hogan, N. Dattagupta, D. M. Crothers, *Nature* **278**, 521 (1979).
5. B. S. Parekh, G. W. Hatfield, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1173 (1996).
6. J. Rudnick, R. Bruinsma, *Biophys. J.* **76**, 1725 (1999).
7. D. Panne, T. Maniatis, S. C. Harrison, *Cell* **129**, 1111 (2007).
8. R. Moretti *et al.*, *ACS Chem. Biol.* **3**, 220 (2008).
9. E. F. Koslover, A. J. Spakowitz, *Phys. Rev. Lett.* **102**, 178102 (2009).
10. H. G. Garcia *et al.*, *Cell Reports* **2**, 150 (2012).
11. S. Kim, P. C. Blainey, C. M. Schroeder, X. S. Xie, *Nat. Methods* **4**, 397 (2007).
12. Materials and methods are available as supplementary materials on *Science* Online.
13. B. F. Luisi *et al.*, *Nature* **352**, 497 (1991).
14. M. Newman, T. Strzelecka, L. F. Dorner, I. Schildkraut, A. K. Aggarwal, *Science* **269**, 656 (1995).
15. F. K. Winkler *et al.*, *EMBO J.* **12**, 1781 (1993).
16. M. Lewis *et al.*, *Science* **271**, 1247 (1996).
17. C. G. Kalodimos *et al.*, *EMBO J.* **21**, 2866 (2002).
18. K. J. Durniak, S. Bailey, T. A. Steitz, *Science* **322**, 553 (2008).
19. S. B. Smith, L. Finzi, C. Bustamante, *Science* **258**, 1122 (1992).
20. Z. Bryant *et al.*, *Nature* **424**, 338 (2003).
21. A. A. Travers, *Annu. Rev. Biochem.* **58**, 427 (1989).
22. R. Rohs *et al.*, *Annu. Rev. Biochem.* **79**, 233 (2010).
23. A. Újvári, C. T. Martin, *J. Mol. Biol.* **295**, 1173 (2000).
24. G.-Q. Tang, S. S. Patel, *Biochemistry* **45**, 4936 (2006).
25. Y. Jia, A. Kumar, S. S. Patel, *J. Biol. Chem.* **271**, 30451 (1996).
26. J. H. Miller, *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, 1972).
27. J. Müller, S. Oehler, B. Müller-Hill, *J. Mol. Biol.* **257**, 21 (1996).
28. W. Lee *et al.*, *Nat. Genet.* **39**, 1235 (2007).
29. E. Sharon *et al.*, *Nat. Biotechnol.* **30**, 521 (2012).
30. S. John *et al.*, *Nat. Genet.* **43**, 264 (2011).
31. S. H. Meijsing *et al.*, *Science* **324**, 407 (2009).
32. H. Viadiu, A. K. Aggarwal, *Nat. Struct. Biol.* **5**, 910 (1998).
33. R. Daber, S. Stayrook, A. Rosenberg, M. Lewis, *J. Mol. Biol.* **370**, 609 (2007).
34. P. T. Lowary, J. Widom, *J. Mol. Biol.* **276**, 19 (1998).

Acknowledgments: We thank K. Wood for his early involvement and J. Hynes, A. Szabo, C. Bustamante, and J. Gelles for helpful discussions. This work is supported by NIH Director's Pioneer Award to X.S.X., Peking University for BIOPIC, Thousand Youth Talents Program for Y.S., as well as the Major State Basic Research Development Program (2011CB809100), National Natural Science Foundation of China (31170710, 31271423, 21125311).

Supplementary Materials

www.sciencemag.org/cgi/content/full/339/6121/816/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S18
Tables S1 to S11
References (*35*–*75*)

23 August 2012; accepted 7 November 2012
10.1126/science.1229223

Multiplex Genome Engineering Using CRISPR/Cas Systems

Le Cong,[1,2]* F. Ann Ran,[1,4]* David Cox,[1,3] Shuailiang Lin,[1,5] Robert Barretto,[6] Naomi Habib,[1] Patrick D. Hsu,[1,4] Xuebing Wu,[7] Wenyan Jiang,[8] Luciano A. Marraffini,[8] Feng Zhang[1]†

Functional elucidation of causal genetic variants and elements requires precise genome editing technologies. The type II prokaryotic CRISPR (clustered regularly interspaced short palindromic repeats)/Cas adaptive immune system has been shown to facilitate RNA-guided site-specific DNA cleavage. We engineered two different type II CRISPR/Cas systems and demonstrate that Cas9 nucleases can be directed by short RNAs to induce precise cleavage at endogenous genomic loci in human and mouse cells. Cas9 can also be converted into a nicking enzyme to facilitate homology-directed repair with minimal mutagenic activity. Lastly, multiple guide sequences can be encoded into a single CRISPR array to enable simultaneous editing of several sites within the mammalian genome, demonstrating easy programmability and wide applicability of the RNA-guided nuclease technology.

Precise and efficient genome-targeting technologies are needed to enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements. Although genome-editing technologies such as designer zinc fingers (ZFs) (*1*–*4*), transcription activator–like effectors (TALEs) (*4*–*10*), and homing meganucleases (*11*) have begun to enable targeted genome modifications, there remains a need for new technologies that are scalable, affordable, and easy to engineer. Here, we report the development of a class of precision genome-engineering tools based on the RNA-guided Cas9 nuclease (*12*–*14*) from the type II prokaryotic clustered regularly interspaced short palindromic repeats (CRISPR) adaptive immune system (*15*–*18*).

The *Streptococcus pyogenes* SF370 type II CRISPR locus consists of four genes, including the Cas9 nuclease, as well as two noncoding CRISPR RNAs (crRNAs): trans-activating crRNA (tracrRNA) and a precursor crRNA (pre-crRNA) array containing nuclease guide sequences (spacers) interspaced by identical direct repeats (DRs) (fig. S1) (*19*). We sought to harness this prokaryotic

[1]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA, and McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Department of Biological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [2]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA 02115, USA. [3]Harvard-MIT Health Sciences and Technology, Harvard Medical School, Boston, MA 02115, USA. [4]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. [5]School of Life Sciences, Tsinghua University, Beijing 100084, China. [6]Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA. [7]Computational and Systems Biology Graduate Program and Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [8]Laboratory of Bacteriology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: zhang@broadinstitute.org