connecting the compound-dependent CRY stabilization with period lengthening. Because CRY proteins are targets of an E3 ubiquitin ligase complex SCF^{FBXL3} and degraded through the ubiquitin-proteasome pathway (*19*–*21*), we tested the effect of KL001 on CRY1 ubiquitination in vitro in a lysate of HEK293T cells transiently overexpressing CRY1-Flag. The compound (50 μM) inhibited ubiquitination of CRY1 and showed only a little effect on the CRY1D387N mutant (fig. S13). Moreover, siRNA-mediated depletion of FBXL3 in U2OS reporter cells diminished the effects of KL001 on the period and *Per2* reporter intensity without affecting longdaysin effects (Fig. 3F and fig. S14). These results indicate that KL001 inhibits FBXL3- and ubiquitin-dependent degradation of CRY proteins and further support the selectivity of the compound.

We then used KL001 in combination with mathematical modeling to explore how KL001-mediated CRY stabilization results in period lengthening and to define the roles of the seemingly redundant CRY isoforms in the clock mechanism. We constructed a simple mathematical model of the PER-CRY negative feedback loop (Fig. 4A and fig. S15A) (*14*). The model successfully reproduced period shortening and lengthening by dose-dependent knockdown of *Cry1* and *Cry2*, respectively (*22*) (fig. S15B), and also period shortening by stabilization of cytosolic CRY2 (*23*) (fig. S15C). For period lengthening by KL001-dependent CRY stabilization, the model predicted that the stabilization occurs in the nucleus (Fig. 4B, left panel, and fig. S15D). Indeed, amounts of CRY1 and CRY2 proteins were increased and sustained, respectively, in a nuclear fraction of unsynchronized U2OS cells after KL001 treatment, although amounts of PER1 were reduced (fig. S16). Furthermore, in silico stabilization of nuclear CRY2 in a *Cry1* knockout background and nuclear CRY1 in a *Cry2* knockout background both caused period lengthening (Fig. 4B, middle and right panels). Consistent with this prediction, continuous treatment with KL001 lengthened the period in both *Cry1* knockout and *Cry2* knockout fibroblasts in a dose-dependent manner (Fig. 4C and fig. S17, A and B). Similarly, the compound caused period lengthening in CRY1 knockdown and CRY2 knockdown U2OS cells (fig. S17C) and in SCN explants from *Cry1* knockout and *Cry2* knockout mice (fig. S17D). Thus, both CRY isoforms share a similar functional role in the period regulation, despite different free-running periods in their knockouts (Fig. 4A). With both CRY1 and CRY2 feedback loops intact, the nuclear CRY1/CRY2 ratio controls the period in a bidirectional manner; that is, more CRY1 causes longer periods and more CRY2 causes shorter periods (fig. S15, B and C).

In the liver, CRY proteins negatively regulate fasting hormone–induced transcription of the *Pck1* and *G6pc* genes, which encode rate-limiting enzymes of gluconeogenesis (*4*, *5*). We therefore tested the effect of KL001 on expression of these genes in mouse primary hepatocytes. KL001 repressed glucagon-dependent induction of *Pck1* and *G6pc* genes in a dose-dependent manner without affecting their basal expression (Fig. 4D). Consistent with this result, KL001 treatment repressed glucagon-mediated activation of glucose production (Fig. 4E). This repression was specific, because basal glucose production (Fig. 4E) and cellular lactate dehydrogenase activity (fig. S18) were unaffected. Altogether, our results demonstrate the potential of KL001 to control fasting hormone–induced gluconeogenesis. Given that human genome-wide association studies identified an association of the *CRY2* gene locus with fasting blood glucose concentrations and presentation of type 2 diabetes (*24*, *25*), KL001 may provide the basis for a therapeutic approach for diabetes.

References and Notes

1. C. B. Green, J. S. Takahashi, J. Bass, *Cell* **134**, 728 (2008).
2. J. Bass, J. S. Takahashi, *Science* **330**, 1349 (2010).
3. G. Asher, U. Schibler, *Cell Metab.* **13**, 125 (2011).
4. E. E. Zhang *et al.*, *Nat. Med.* **16**, 1152 (2010).
5. K. A. Lamia *et al.*, *Nature* **480**, 552 (2011).
6. L. Yin *et al.*, *Science* **318**, 1786 (2007).
7. T. Hirota *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20746 (2008).
8. Y. Isojima *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15744 (2009).
9. T. Hirota *et al.*, *PLoS Biol.* **8**, e1000559 (2010).
10. J. W. Lee *et al.*, *Angew. Chem. Int. Ed. Engl.* **50**, 10608 (2011).
11. Z. Chen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 101 (2012).
12. L. A. Solt, D. J. Kojetin, T. P. Burris, *Future Med. Chem.* **3**, 623 (2011).
13. L. A. Solt *et al.*, *Nature* **485**, 62 (2012).
14. Materials and methods are available as supplementary materials on *Science* Online.
15. S. H. Yoo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5339 (2004).
16. T. Hirota, S. A. Kay, *Chem. Biol.* **16**, 921 (2009).
17. K. Hitomi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6962 (2009).
18. S. H. Yoo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2608 (2005).
19. S. M. Siepka *et al.*, *Cell* **129**, 1011 (2007).
20. S. I. Godinho *et al.*, *Science* **316**, 897 (2007).
21. L. Busino *et al.*, *Science* **316**, 900 (2007).
22. E. E. Zhang *et al.*, *Cell* **139**, 199 (2009).
23. N. Kurabayashi, T. Hirota, M. Sakai, K. Sanada, Y. Fukada, *Mol. Cell. Biol.* **30**, 1757 (2010).
24. J. Dupuis *et al.*; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators, *Nat. Genet.* **42**, 105 (2010).
25. M. A. Kelly *et al.*; DIAGRAM Consortium; SAT2D Consortium, *PLoS ONE* **7**, e32670 (2012).

Acknowledgments: We thank E. Peters, X. Liu, M. Garcia, C. Cho, and R. Glynne for assistance; C. Doherty for critical reading; and K. Lamia and J. Takahashi for reagents. This work was supported in part by grants from NIH (GM074868, MH051573, and GM085764 to S.A.K.; GM096873 to F.J.D.; and MH082945 to D.K.W.), Skaggs Institute for Chemical Biology (to P.G.S), the U.S. Army Research Office (W911NF-09-0001 to F.J.D.), and a Department of Veterans Affairs Career Development Award (to D.K.W.). S.A.K. and P.G.S. serve on the Board of Reset Therapeutics and are paid consultants.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1223710/DC1
Materials and Methods
Figs. S1 to S18
Tables S1 to S4
References (*26*–*43*)

23 April 2012; accepted 27 June 2012
Published online 12 July 2012;
10.1126/science.1223710

Extreme Bendability of DNA Less than 100 Base Pairs Long Revealed by Single-Molecule Cyclization

Reza Vafabakhsh[1] and Taekjip Ha[1,2]*

The classical view of DNA posits that DNA must be stiff below the persistence length [<150 base pairs (bp)], but recent studies addressing this have yielded contradictory results. We developed a fluorescence-based, protein-free assay for studying the cyclization of single DNA molecules in real time. The assay samples the equilibrium population of a sharply bent, transient species that is entirely suppressed in single-molecule mechanical measurements and is biologically more relevant than the annealed species sampled in the traditional ligase-based assay. The looping rate has a weak length dependence between 67 and 106 bp that cannot be described by the worm-like chain model. Many biologically important protein-DNA interactions that involve looping and bending of DNA below 100 bp likely use this intrinsic bendability of DNA.

Bending and looping of lengths of DNA below 100 base pairs (bp) is ubiquitous in cellular processes such as regulated gene expression in bacteria and eukaryotes (*1*, *2*), packaging of DNA in viral capsids, and DNA storage complexes in eukaryotes (*3*). Quantifying the intrinsic bendability of DNA at these biologically important length scales is essential for understanding DNA-protein interactions. According to a widely used approximation, DNA duplex is modeled as an elastic rod and its mechanical properties are described by the worm-like chain (WLC) model. Persistence length (l_p) is a measure of the bending rigidity of DNA; for a DNA

molecule that is several kilobasepair (kbp) or longer, l_p can be readily measured using single-molecule manipulation tools and is about 50 nm or 150 bp (*4*). In this framework, formation of DNA loops or sharp bends over length scales shorter than l_p incurs a large energetic cost that makes the probability of their spontaneous formation vanishingly small.

Many approaches have been developed to quantify and model the inherent flexibility and bendability of DNA at short length scales. In the cyclization assay, the ligase protein traps DNA molecules in the looped conformation, and then the looped and unlooped populations are separated based on their different gel mobility (*5*). Recent experiments using this assay and other techniques have challenged the classical picture of DNA as an elastic rod (*6*). Cyclization of DNA fragments of ~100 bp using the ligase assay yielded up to four orders of magnitude higher cyclizability (*j* factor) compared to the prediction of the WLC model (*7*). However, this extraordinary result was questioned and attributed to too high a ligase concentration used in the experiments (*8*). Small-angle x-ray scattering was used to measure end-to-end distance variations of short DNA fragments labeled with gold nanoparticles. These experiments suggested a cooperative stretching behavior over two helical turns (*9*); however, some aspects of the data are still paradoxical (*10*–*12*). Analysis of DNA images acquired using atomic force microscopy also deduced lower bending energies than WLC predicts (*13*). However, this method is indirect, is based on surface absorption of DNA molecules, and does not provide any dynamic information.

Contradictory and inconclusive results from these measurements call for a more direct approach to quantify flexibility of DNA on short length scales. In addition, most existing bulk approaches suffer from inherent limitations, such as limited range of physical conditions and formation of by-products other than monomer DNA circles, that limit their applicability to other systems. For example, due to nonspecific interactions of DNA and the ligase protein, the ligase-based assay may not be suitable for studying cyclization of very short DNA molecules (*14*). Moreover, looping events cannot be detected in real time using bulk techniques. Because of geometrical and technical limitations, single-molecule DNA-stretching approaches cannot be used to study the mechanics of very short DNA molecules. Even for a moderate length of DNA with several hundred bp, many corrections are required to account for the finite chain length and the boundary conditions (*15*). In addition, because of the relatively long persistence length of double-stranded DNA (dsDNA), even 100 femto-Newton of force makes configurations with sharp local bending inaccessible. Therefore, DNA responses measured by mechanical stretching would not include any contribution from such sharply bent conformations even if they existed in a relaxed DNA.

We developed a cyclization assay, based on single-molecule fluorescence resonance energy transfer (smFRET) (*16*, *17*), for directly monitoring the cyclization of single DNA molecules (Fig. 1A) (*18*). To avoid dimer formation during long-term observation, DNA molecules are immobilized on a polymer-coated surface through a biotin linker attached to a base at an internal DNA location. We avoided motifs such as A tracts, which are known to induce considerable intrinsic curvature (*19*). The DNA probe is a duplex with single-stranded extensions on both 5′ ends. Each DNA molecule is labeled with Cy3 (donor) and Cy5 (acceptor) fluorophores at the 5′ end of the strands. Single-stranded overhangs are complementary so that hybridization will trap the DNA molecules in the looped state. In the unlooped state, the donor and acceptor are distant from each other and the molecules show zero FRET. Looping brings the dyes close to each other, and the DNA molecules exhibit a high FRET signal. Therefore, the looped state can be clearly distinguished from the unlooped state based on the FRET value and the relative intensities of donor and acceptor (Fig. 1B).

The experiment starts in a buffer without added ions in order to strongly favor the unlooped state. Introducing a buffer containing high concentrations of Na^+ or Mg^{2+} can stabilize the looped state. Depending on the lengths of the DNA duplex and the single-stranded overhangs, different behaviors were observed during the probing time window, typically ranging from less than 1 min to up to 4 hours. DNA molecules formed stable loops, showed dynamics between looped and unlooped conformations, or exhibited no looping events.

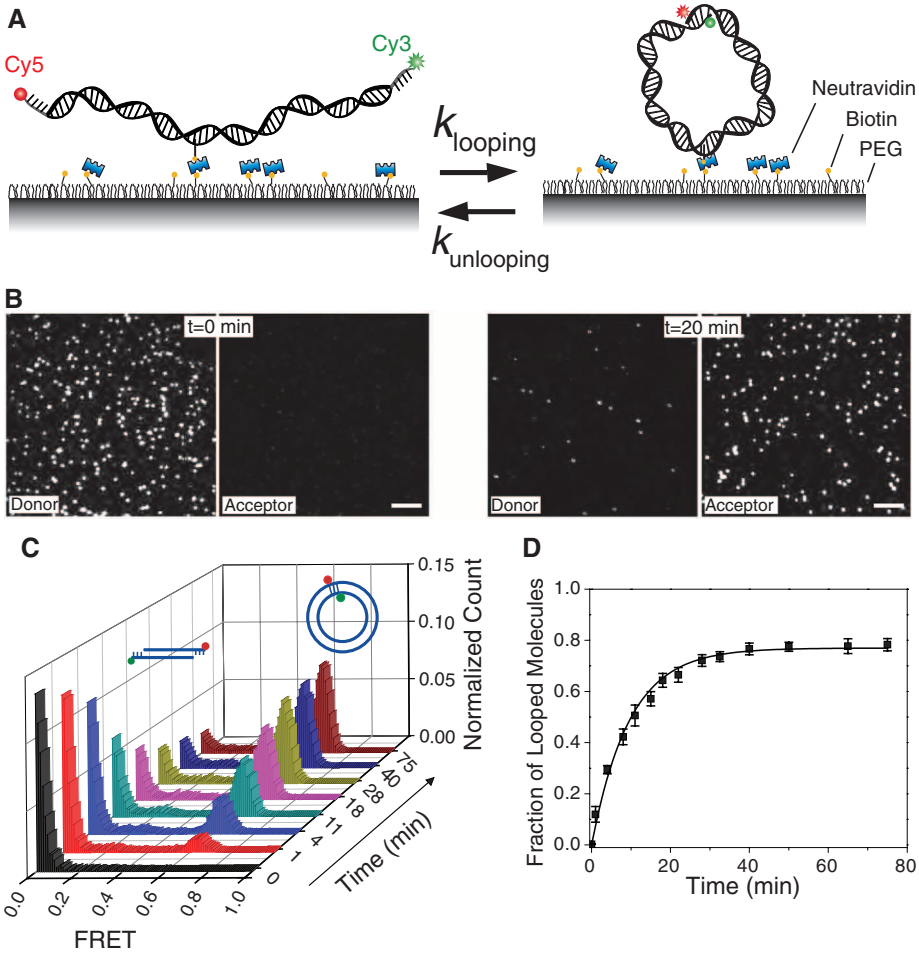

Fig. 1. (**A**) Donor (Cy3) and acceptor (Cy5) labeled DNA molecules were immobilized on the surface via biotin-neutravidin interaction. (**B**) Fluorescence images of single 91-bp DNA molecules in corresponding donor and acceptor channels are shown before (left panels) and 20 min after adding high salt (1 M NaCl) buffer (right panels). Scale bar, 5 μm. (**C**) Histograms of FRET efficiency as a function of time (t = 0 is when high salt was introduced) show the evolution of looped (high FRET) and unlooped (low FRET) populations. (**D**) Fraction of looped DNA (high FRET population) as a function of time, measured from the histograms in C. An exponential fit to this curve gives R. Error bars indicate ± SEM; n = 5.

[1]Department of Physics and the Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [2]Howard Hughes Medical Institute, Urbana, IL 61801, USA.

*To whom correspondence should be addressed. E-mail: tjha@illinois.edu

For example, for a 91-bp initial dsDNA with 10-nucleotide (nt) overhangs, looping was nearly irreversible and the looped high FRET state accumulated to saturation within about 20 min (Fig. 1, C and D). In this case, the looping rate R could be determined by fitting the time evolution of the looped population with a single exponential function (Fig. 1D).

We measured R for a series of DNA molecules with a 10-nt overhang on each end and with the circular size ranging from 67 to 106 bp (the circular size is the circumference of the DNA circle formed after looping and is therefore the sum of the initial dsDNA length and the overhang length). The measured looping time, $1/R$, varied from less than 10 min to more than 200 min (Fig. 2A, black squares). This 20-fold change in the looping times is smaller than expected; we expected that DNA much shorter than the persistent length would take considerably longer to form a loop. However, the result is qualitatively consistent with the observations that short DNA loops induced by the lac repressor and AraC protein can form efficiently in vivo and in vitro (*2*, *20*–*22*), which suggests that the protein-induced looping may use the intrinsic flexibility of the DNA. Also, our result is consistent with a recent study which predicted that bending energy for short DNA loops would be independent of the loop length (*13*). In addition to the length dependence, a variation in looping rate depending on the angular phase between the two cohesive ends would be expected. Indeed, our data displays oscillation in R with a period of about one helical turn (see data points between 93 and 106 bp).

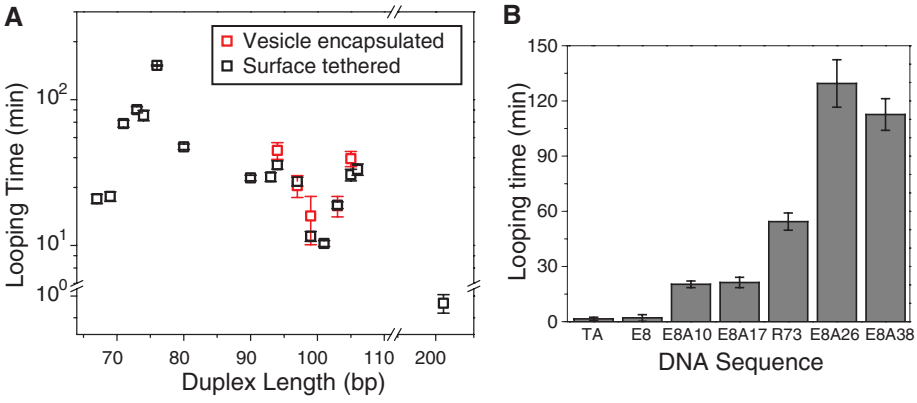

Fig. 2. (**A**) Looping time as a function of DNA circular length for surface-tethered DNA (black squares) and vesicle-encapsulated DNA molecules (red squares). (**B**) Looping time for 7 DNA sequences with 63-bp duplex length and 10-nt overhang. R73 is the standard sequence used in (A). Poly-A constructs were constructed by inserting n = 10, 17, 26, and 38 consecutive A bases in the middle of a random sequence (E8). Error bars indicate ± SEM; $n \geq 3$.

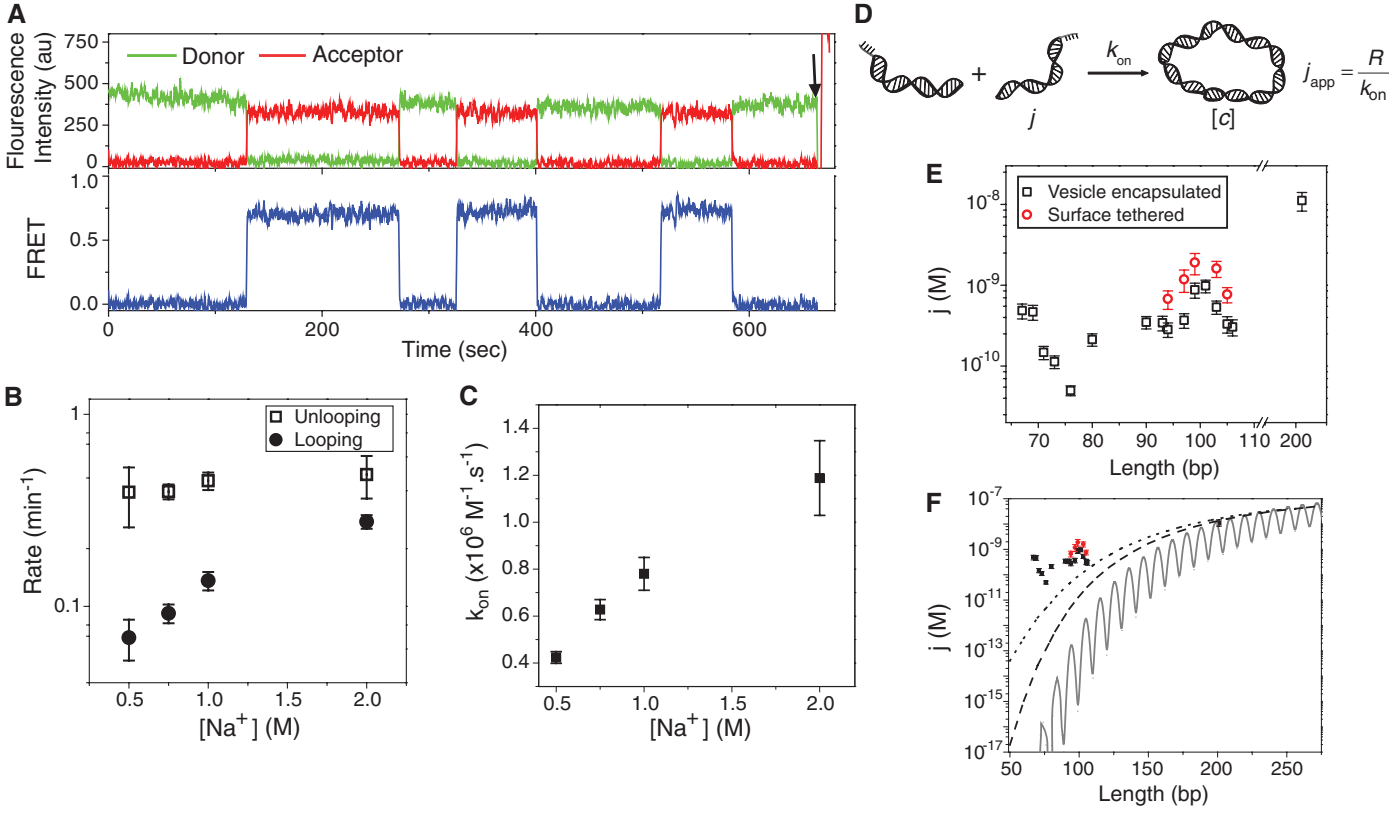

Fig. 3. (**A**) Representative fluorescence intensities (top, green for donor and red for acceptor) and corresponding FRET efficiency (bottom, blue) time traces measured from a single DNA molecule in 750 mM NaCl. The DNA has 91-bp initial dsDNA with 8-nt single-stranded overhangs. The arrow indicates a direct acceptor excitation to verify that the acceptor has not photobleached. (**B**) Looping and unlooping rates as a function of [NaCl]. The DNA has 91-bp initial dsDNA with 8-nt single-stranded overhangs. (**C**) k_{on} measured as shown in fig. S2 shows the same 3-fold increase as the looping rate with increasing [NaCl]. Data are means ± SEM ($n \geq 300$ molecules). (**D**) The model to relate R, k_{on}, and apparent j factor. (**E**) j factor for surface-tethered DNA (black squares) and vesicle-encapsulated DNA (red circles). (**F**) Measured j factor for surface-tethered DNA (black squares) and vesicle-encapsulated DNA (red squares). Solid black curve is the Shimada-Yamakawa prediction for DNA cyclization. Dashed line and dotted line are the WLC predictions for the j factor of DNA circles with free boundary condition and for DNA molecules with 5-nm capture radius, respectively. Error bars indicate ± SEM; $n \geq 3$.

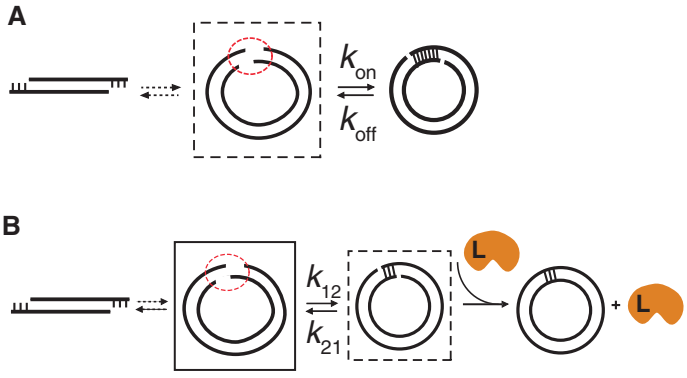

Fig. 4. (**A**) Our assay reports on the equilibrium population of the intermediate state (dashed box), with the two DNA ends in close proximity. (**B**) Schematic representation of DNA cyclization reaction steps in the ligase assay. The intermediate state with the two DNA ends in close proximity (solid box) does not get sampled in this assay. Instead, the ligase samples the equilibrium population of the annealed state (dashed box). Ligase protein is labeled L.

We performed a variety of controls to confirm that the unexpectedly weak length dependence of the looping rates was not an artifact of our experimental scheme. First, to rule out possible contributions from surface tethering and internal biotin labeling, we repeated the experiments for DNA molecules without any internal modification and confined in 200-nm-diameter phospholipid vesicles that are permeable to ions (*23*). Because infusion of 1 M NaCl ruptured the vesicles, we instead used 10 mM Mg^{2+} to stabilize the looped state. This is the same ionic condition used in the standard ligase assay. For five DNA constructs ranging from 94 to 105 bp, the looping times were similar between the vesicle encapsulated measurements and the surface measurements (Fig. 2A, red squares). Because DNA bending and torsional rigidity are sequence dependent (*24*), it remained possible that the high flexibility we observed was due to an extraordinarily flexible sequence. Therefore, we measured the looping times for a series of DNA constructs with identical loop length and overhang sequence but different internal base composition (Fig. 2B). We found that our standard sequence (R73) is not an outlier in terms of the looping rate. The DNA derived from a nucleosome positioning sequence (TA) (*25*) showed much faster (by a factor of 35) looping than our standard sequence R73 (*7*). We also examined the looping behavior of sequences with potentially curvature-inducing A tracks, $(A)_n$— where n = 0, 10, 17, 26, or 38 consecutive A bases embedded in an otherwise randomly chosen sequence—and found that the looping rate varied from a factor of more than 30 higher (n = 0) to a factor of 2 lower (n = 26) than that of R73. The two orders of magnitude difference in the looping time for these sequences, despite the fact that the final 12 bp of these duplexes on both ends are similar, rules out duplex end opening as a possible mechanism for the rapid looping observed in our assay.

Changing the length of overhang to 8 nt allowed us to observe real-time looping-unlooping dynamics (Fig. 3A) and evaluate the rates directly as a functional of salt. Although the unlooping rate did not change between 0.5 M and 2 M Na^{+}, the looping rate increased by a factor of 3 in this range (Fig. 3B). However, because we observed the same increase in the bimolecular annealing rate (Fig. 3C), we can attribute the acceleration in looping at higher salt solely to the annealing enhancement. Therefore, monovalent ion concentrations above 0.5 M do not have a detectable effect on dsDNA flexibility (3).

We also investigated the effect of the single-stranded overhang length on the stability of DNA loops. Decreasing the overhang length from 10 nt to 9 or 8 nt while maintaining the initial duplex sharply increased the unlooping rate by about two orders of magnitude without an appreciable change in the looping rate (fig. S1A). Likewise, the equilibrium fraction of the looped state became progressively smaller by shortening the overhang (fig. S1B). Therefore, the main effect of longer overhangs in our assay compared to the 4-nt overhangs typical in the ligase assay is to increase the lifetime of the looped state. When a DNA loop forms, internal elastic energy stored in the loop is expected to provide a shear force that promotes unlooping. Indeed, we found that 8 bp of duplex melts 20 times as fast in a DNA circle as in a DNA dimer, likely due to internal tension in a circle that is absent in a dimer (see supplementary text and fig. S1C).

R can be calculated as the product of the bimolecular association rate k_{on} between 10-nt-long complementary strands and the effective concentration of one end of DNA in the vicinity of the other end, which we call the apparent j factor, j_{app} (*5*) (Fig. 3D). Using a similar surface-based assay but for intermolecular annealing (see supplementary text and fig. S2), k_{on} was measured to be $0.78 \pm 0.07 \times 10^6 M^{-1}s^{-1}$ in 1 M NaCl and $0.26 \pm 0.04 \times 10^6 M^{-1}s^{-1}$ in 10 mM Mg^{2+}, both consistent with an earlier estimate for short oligonucleotide annealing (*26*). The corresponding apparent j factors, calculated using $j_{app} = R/k_{on}$, are shown in Fig. 3E. Our calculated apparent j factors, along with the prediction of the WLC model (*27*, *28*), are plotted in Fig. 3F. The solid line and the dashed line are the j factor for a semiflexible polymer with parallel and free boundary conditions, and the dotted line is the j factor for a polymer with free boundary condition and 5-nm capture radius (the two ends anneal when they are closer than 5 nm) (*29*, *30*). The measured j_{app} values matched the theoretical prediction for 201-bp DNA but deviated from the theoretical values by orders of magnitude for the shortest lengths examined, even under the most liberal boundary condition.

Our observation that the looping rate does not drop precipitously with decreasing DNA length is in stark contrast to the steep drop in the j factor predicted by the WLC model. In many biologically relevant protein-DNA interactions, such as in some genetic switches, the DNA bendability plays an important role in determining the state of the switch by controlling the concentration of one protein binding site in the vicinity of the other binding site (*1*, *31*). Our assay samples such an equilibrium in which two DNA ends are in close proximity but not annealed (dashed box in Fig. 4A). In contrast, the ligase assay samples the equilibrium of the annealed state (dashed box in Fig. 4B) (*5*). The equilibrium looped population, which is the substrate for ligase protein, is very sensitive to the unlooping rate (figs. S1A and S1B). The looping rate is biologically more relevant because it reports on how quickly two regions of DNA are brought into close proximity, whereas the unlooping rate is additionally influenced by the melting rate of the short duplex formed. Our assay could independently measure the looping rate without being affected by the loop instability caused by internal tension in the short DNA circles. Many DNA binding proteins may have evolved to use the high flexibility of the DNA to capture and further stabilize transiently bent or looped DNA conformations.

Extended WLC models have previously been developed to explain the high flexibility of short DNA by allowing for the formation of temporary bubbles (*32*) or kinks (*33*), and molecular dynamics simulations observed the emergence of kinks in small DNA minicircles (*34*, *35*). Others accommodate high bendability of short DNA by introducing nonharmonic elastic behavior (*13*). To gain insight into the mechanism of facile looping, we performed experiments on DNA constructs with a single backbone nick, double nicks, or a single-bp mismatch in the middle and observed one to two orders of magnitude higher looping rate compared with our original DNA constructs (figs. S3A and S3B). This considerable enhancement in the looping rate confirms that stable defects such as a single-bp mismatch or a nick can enhance global cyclizability of DNA, suggesting that similar but transient defects, if they are frequent enough, may explain the extreme bendability observed here. However, determining whether the high bendability of DNA at short length scales comes from transient kinks or bubbles or stems from anharmonic elasticity of DNA requires improved computational methods and further studies.

References and Notes

1. E. M. Blackwood, J. T. Kadonaga, *Science* **281**, 60 (1998).
2. O. K. Wong, M. Guthold, D. A. Erie, J. Gelles, *PLoS Biol.* **6**, e232 (2008).
3. T. J. Richmond, C. A. Davey, *Nature* **423**, 145 (2003).

4. C. G. Baumann, S. B. Smith, V. A. Bloomfield, C. Bustamante, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6185 (1997).
5. D. Shore, J. Langowski, R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4833 (1981).
6. J. P. Peters 3rd, L. J. Maher, *Q. Rev. Biophys.* **43**, 23 (2010).
7. T. E. Cloutier, J. Widom, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3645 (2005).
8. Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, A. Vologodskii, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5397 (2005).
9. R. S. Mathew-Fenn, R. Das, P. A. B. Harbury, *Science* **322**, 446 (2008).
10. N. B. Becker, R. Everaers, *Science* **325**, 538 (2009).
11. R. S. Mathew-Fenn, R. Das, T. D. Fenn, M. Schneiders, P. A. B. Harbury, *Science* **325**, 538 (2009).
12. A. J. Mastroianni, D. A. Sivak, P. L. Geissler, A. P. Alivisatos, *Biophys. J.* **97**, 1408 (2009).
13. P. A. Wiggins *et al.*, *Nat. Nanotechnol.* **1**, 137 (2006).
14. C. Yuan, X. W. Lou, E. Rhoades, H. Chen, L. A. Archer, *Nucleic Acids Res.* **35**, 5294 (2007).
15. Y. Seol, J. Li, P. C. Nelson, T. T. Perkins, M. D. Betterton, *Biophys. J.* **93**, 4360 (2007).
16. R. Roy, S. Hohng, T. Ha, *Nat. Methods* **5**, 507 (2008).
17. T. Ha *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6264 (1996).
18. Materials and methods and additional information are available as supplementary materials on *Science* Online.
19. A. M. Burkhoff, T. D. Tullius, *Nature* **331**, 455 (1988).
20. L. Han *et al.*, *PLoS ONE* **4**, e5621 (2009).
21. N. A. Becker, J. D. Kahn, L. J. Maher III, *J. Mol. Biol.* **349**, 716 (2005).
22. D. H. Lee, R. F. Schleif, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 476 (1989).
23. I. Cisse, B. Okumus, C. Joo, T. Ha, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12646 (2007).
24. P. J. Hagerman, *Annu. Rev. Biophys. Biophys. Chem.* **17**, 265 (1988).
25. P. T. Lowary, J. Widom, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1183 (1997).
26. D. Pörschke, M. Eigen, *J. Mol. Biol.* **62**, 361 (1971).
27. J. Shimada, H. Yamakawa, *Macromolecules* **17**, 689 (1984).
28. K. B. Towles, J. F. Beausang, H. G. Garcia, R. Phillips, P. C. Nelson, *Phys. Biol.* **6**, 025001 (2009).
29. J. Yan, R. Kawamura, J. F. Marko, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **71**, 061905 (2005).
30. N. Douarche, S. Cocco, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **72**, 061902 (2005).
31. R. Schleif, *Annu. Rev. Biochem.* **61**, 199 (1992).
32. J. Yan, J. F. Marko, *Phys. Rev. Lett.* **93**, 108108 (2004).
33. P. A. Wiggins, R. Phillips, P. C. Nelson, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **71**, 021909 (2005).
34. F. Lankaš, R. Lavery, J. H. Maddocks, *Structure* **14**, 1527 (2006).
35. J. Curuksu, M. Zacharias, R. Lavery, K. Zakrzewska, *Nucleic Acids Res.* **37**, 3766 (2009).

Acknowledgments: We thank R. Phillips, H. Garcia, K. Rangunathan, R. Zhou, S. Doganay, A. Jain, K. Lee, G. Lee, and S. Arslan for helpful discussions. We are grateful to the late J. Widom for generous comments on an earlier version of this paper. This work was supported by U.S. National Science Foundation grants 0646550 and 0822613 and U.S. National Institutes of Health grant GM065367 to T.H. T.H. is an employee of the Howard Hughes Medical Institute.

Supplementary Materials

www.sciencemag.org/cgi/content/full/337/6098/1097/DC1
Materials and Methods
Figs. S1 to S3
References (*36*–*39*)

1 May 2012; accepted 12 July 2012
10.1126/science.1224139

Network Context and Selection in the Evolution to Enzyme Specificity

Hojung Nam,[1]* Nathan E. Lewis,[1,3]*‡ Joshua A. Lerman,[2] Dae-Hee Lee,[1]† Roger L. Chang,[2] Donghyuk Kim,[1] Bernhard O. Palsson[1]‡

Enzymes are thought to have evolved highly specific catalytic activities from promiscuous ancestral proteins. By analyzing a genome-scale model of *Escherichia coli* metabolism, we found that 37% of its enzymes act on a variety of substrates and catalyze 65% of the known metabolic reactions. However, it is not apparent why these generalist enzymes remain. Here, we show that there are marked differences between generalist enzymes and specialist enzymes, known to catalyze a single chemical reaction on one particular substrate in vivo. Specialist enzymes (i) are frequently essential, (ii) maintain higher metabolic flux, and (iii) require more regulation of enzyme activity to control metabolic flux in dynamic environments than do generalist enzymes. Furthermore, these properties are conserved in Archaea and Eukarya. Thus, the metabolic network context and environmental conditions influence enzyme evolution toward high specificity.

Ancestral enzymes are proposed to have exhibited broad substrate specificity and low catalytic efficiency (*1*). Through mutation, duplication, and horizontal gene transfer, gene families diversified and promiscuous enzymes apparently were refined to exhibit specific and more efficient catalytic abilities (*2*, *3*). Thus, today's metabolic enzymes are commonly assumed to be "specialists," having evolved to catalyze one reaction on a unique primary substrate in an organism. However, some enzymes are "generalists" that promiscuously catalyze reactions on a variety of substrates in vivo (*2*) or exhibit multifunctionality by catalyzing multiple classes of reactions, often at different active sites (*4*). Thus, a fundamental question arises: Why do some enzymes evolve to become specialists, whereas others retain generalist characteristics? By analyzing enzyme functions and properties in experimental data and in silico metabolic network models, we show that the in vivo biochemical network context in which an enzyme resides may influence the evolution of enzyme specificity.

How many metabolic enzymes are generalists? To answer this question, we used a comprehensive reconstruction of the *Escherichia coli* K-12 MG1655 metabolic network, which accounts for the metabolic functions of 1260 gene products (28% of the predicted and experimentally validated open reading frames in *E. coli*) (*5*), which contribute to 1081 enzyme complexes analyzed in this study. In the reconstruction, we define a reaction as a unique set of substrates that are chemically transformed into a unique set of products. With this definition, we classified 677 enzymes as specialists because they catalyze one unique reaction and 404 as generalists because they catalyze multiple reactions. Thus, we estimate that 37% of metabolic enzymes in *E. coli* are generalists, most of which exhibit substrate promiscuity (fig. S1A). Furthermore, specialist and generalist enzymes catalyze 454 and 859 metabolic reactions, respectively, distributed across many metabolic subsystems (Fig. 1, A and B). Thus, contrary to the textbook view of enzymes as "specific catalysts," generalist enzymes have a prominent role in *E. coli*, catalyzing at least 65% of the nonspontaneous metabolic reactions.

We performed several network-wide analyses to provide additional support for our estimates and the classification. First, we found that almost all genes in the network have been well characterized and studied in more than 61,727 published studies (fig. S1D). Second, we found no correlation between our classification and knowledge depth, i.e., neither specialist nor generalist enzymes had been studied in more depth (fig. S1E). Third, our generalist enzymes did not likely include many latent promiscuous reactions measured in vitro that likely do not occur in vivo, because 85% of the generalist enzymes reactions (GERxns) were active in silico in common growth conditions. This is the same percentage seen for specialist enzyme reactions (SERxns) (fig. S2). Fourth, because enzyme classification may vary with further study, we tested the sensitivity of the results presented in this work. We found the results to be qualitatively robust with improvements in the metabolic network from the discovery of new enzymes, variations in enzyme classification, and the exclusion of promiscuous enzymes or multifunctional enzymes from the generalist class (fig. S3). Although transporter reactions were not included in the groups of SERxns or

[1]Department of Bioengineering, University of California San Diego, La Jolla, CA 92093–0412, USA. [2]Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA 92093–0412, USA. [3]Wyss Institute for Biologically Inspired Engineering and Department of Genetics, Harvard Medical School, Boston, MA 02115, USA.

*These authors contributed equally to this work.
†Present address: Systems and Synthetic Biology Research Center, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahak-ro, Yuseong-gu, Daejeon 305-806, Korea.
‡To whom correspondence should be addressed. E-mail: nlewis@genetics.med.harvard.edu (N.E.L.); palsson@ucsd.edu (B.O.P.)