Rational design of a split-Cas9 enzyme complex

Addison V. Wrighta,1, Samuel H. Sternbergb,1, David W. Taylorc, Brett T. Staahla, Jorge A. Bardalesd, Jack E. Kornfelda, and Jennifer A. Doudnaa,b,c,d,e,2

*Department of Molecular and Cell Biology, †Department of Chemistry, ‡Howard Hughes Medical Institute, and §Biophysics Graduate Group, University of California, Berkeley, CA 94720; and ¶Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Contributed by Jennifer A. Doudna, January 26, 2015 (sent for review January 21, 2015; reviewed by Dana Carroll and Fyodor D. Urnov)

Cas9, an RNA-guided DNA endonuclease found in clustered regularly interspaced short palindromic repeats (CRISPR) bacterial immune systems, is a versatile tool for genome editing, transcriptional regulation, and cellular imaging applications. Structures of Streptococcus pyogenes Cas9 alone or bound to single-guide RNA (sgRNA) and target DNA revealed a bilobed protein architecture that undergoes major conformational changes upon guide RNA and DNA binding. To investigate the molecular determinants and relevance of the interlobe rearrangement for target recognition and cleavage, we designed a split-Cas9 enzyme in which the nucleo- lobe and α-helical lobe are expressed as separate polypeptides. Although the lobes do not interact on their own, the sgRNA recruits them into a ternary complex that recapitulates the activity of full-length Cas9 and catalyzes site-specific DNA cleavage. The use of a modified sgRNA abrogates split-Cas9 activity by preventing dimerization, allowing for the development of an inducible dimerization system. We propose that split-Cas9 can act as a highly regulatable platform for genome-engineering applications.

Results

Design and Functional Validation of Split-Cas9. The nuclelease lobe of Cas9 includes the RuvC and HNH nuclelease domains, as well as a C-terminal domain that is involved in PAM recognition (Fig. 1A) (8–10). The RuvC domain comprises three distinct motifs: Motifs II and III are interrupted by the HNH domain, and motifs I and II are interrupted by a large lobe composed entirely of α-helices. This α-helical lobe, also referred to as the recognition (REC) lobe (9), forms a broad cleft that makes extensive contacts with the sgRNA and target DNA. We previously showed that the α-helical lobe undergoes a large rotation relative to the nuclelease lobe upon guide RNA binding to create a central channel where target DNA is bound (8).

Using available crystal structures as a guide, we designed a split-Cas9 in which the native structure of both lobes was kept as intact as possible (Fig. 1A). In particular, rather than simply split the full-length Cas9 sequence internally at a single junction, we constructed the nuclelease lobe by directly linking the N-terminal RuvC1 motif to the remainder of the catalytic domains from the rest of the protein scaffold.

Significance

Bacteria have evolved clustered regularly interspaced short palindromic repeats (CRISPRs) together with CRISPR-associated (Cas) proteins to defend themselves against viral infection. RNAs derived from the CRISPR locus assemble with Cas proteins into programmable DNA-targeting complexes that destroy DNA molecules complementary to the guide RNA. In type II CRISPR-Cas systems, the Cas9 protein binds and cleaves target DNA sequences at sites complementary to a 20-nt guide RNA sequence. This activity has been harnessed for a wide range of genome-engineering applications. This study explores the structural features that enable Cas9 to bind and cleave target DNAs, and the results suggest a way of regulating Cas9 by physical separation of the catalytic domains from the rest of the protein scaffold.


Reviews: D.C., University of Utah; and F.D.U., Sangamo Biosciences.

Conflict of interest statement: A.V.W., S.H.S. and J.A.D. have filed a related patent.

Freely available online through the PNAS open access option.

1A.V.W. and S.H.S. contributed equally to this work.

2To whom correspondence should be addressed. Email: doudna@berkeley.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1501698112/-/DCSupplemental.

PNAS | March 10, 2015 | vol. 112 | no. 10 | 2984–2989 | www.pnas.org/cgi/doi/10.1073/pnas.1501698112
amino acids away in primary sequence, with the intervening polypeptide comprising the α-helical lobe. Two crossover points between the lobes occur at residues ∼56 and ∼720 (Fig. 1B): the C-terminal connection is disordered in both apo-Cas9 and sgRNA/DNA-bound structures, and the N-terminal connection occurs between the RuvCI motif and the bridge helix. We connected residue E57 from RuvCI with residue G729 from RuvCII using a three-amino acid linker and removed a short, poorly conserved α-helix from the RuvCII motif that does not seem to play an important structural role in the sgRNA/DNA-bound state (Fig. 1B). The α-helical lobe spans residues G56–S714, with the N terminus encompassing the entirety of the bridge helix.

To determine whether the lobes could function as separate polypeptides, we separately overexpressed both lobes in Escherichia coli and purified them by affinity and size-exclusion chromatography (Fig. 1C and Fig. S1). We investigated whether split-Cas9 (α-helical lobe plus nuclease lobe) would recapitulate the activity of WT Cas9 using a standard cleavage assay with sgRNA and a radiolabeled double-stranded DNA (dsDNA) target (Fig. 1D). No cleavage was observed with either lobe individually, but the reconstituted split-Cas9 enzyme complex exhibited robust target DNA cleavage (Fig. 1D and Fig. S2). Split-Cas9 maintained the same site and pattern of cleavage as WT Cas9, including the "trimming" of the nontarget strand that we observed previously (4), and functioned equally well with a dual-guide RNA composed of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA) (Fig. S2). In addition, we confirmed that split-Cas9 activity was dependent on complementarity between the sgRNA and target DNA as well as the presence of a 5'-NGG-3' PAM (Fig. S2).

When we investigated the kinetics of DNA cleavage under pseudo first-order conditions using excess enzyme, we found that split-Cas9 was ~10-fold slower than WT although it reached the same endpoint after 5 min (Fig. S2 and Table S1). This reduced rate may result from slower kinetics of protein–RNA complex formation, a reduced rate of dsDNA recognition and unwinding, or a minor defect in nuclease domain activation. DNA-binding...
experiments using nuclease-inactive split dCas9 (D10A/H840A mutations) revealed a significantly weaker affinity of split-Cas9 for target DNA than WT Cas9 (Fig. S3), suggesting that slower kinetics of dsDNA binding likely limit the observed rate of cleavage. Collectively, these results demonstrate that the enzymatic activity of WT Cas9 does not require a direct linkage between the α-helical and nuclease lobes although their physical connection within RNA–protein complexes increases the affinity for the target DNA substrate. Remarkably, whereas previous work shows that the RNA-induced large-scale rearrangement of both lobes is necessary for WT Cas9 to achieve an active conformation (8), our experiments revealed that the sgRNA is entirely sufficient to recruit and dimerize the separate lobes into an assembly (8), our experiments revealed that the sgRNA is entirely sufficient to recruit and dimerize the separate lobes into an active enzyme complex. Furthermore, communication through the sgRNA enables PAM recognition, dsDNA unwinding, and DNA cleavage, despite the absence of extensive protein–protein interactions between the lobes.

sgRNA Motifs Recruit both Cas9 Lobes to Form a Ternary Complex.

We next wanted to investigate RNA molecular determinants that promote heterodimerization of the α-helical and nuclease lobes. Crystal structures of sgRNA/DNA-bound Cas9 show that the spacer (guide) and stem-loop motifs at the 5′ end of the sgRNA primarily contact the α-helical lobe whereas two hairpins at the 3′ end bind the outside face of the nuclease lobe (Fig. 2A). The nexus motif, recently shown to be critical for activity (11), occupies a central position between the lobes and forms extensive interactions with the bridge helix. Based on this interaction profile, we generated a full-length sgRNA and two shorter sgRNA constructs that were selectively truncated from either the 5′ or 3′ end (Fig. 2B and Table S2) and determined their affinities for WT Cas9, the individual α-helical and nuclease lobes, and split-Cas9 using a filter binding assay.

The full-length sgRNA is bound by WT Cas9 with an equilibrium dissociation constant ($K_d$) of $10 \pm 2$ PM whereas the lobes individually and together have $K_d$ values in the range of 0.2–0.8 nM (Fig. 2C and Table 1). The difference between WT and split-Cas9 likely reflects the increased entropic cost required to assemble a ternary versus binary complex. Interestingly, WT Cas9 bound a truncated sgRNA comprising only the 3′ hairpins ($Δ$hairpins1-2) would selectively perturb interactions with the bridge helix. Based on this interaction profile, we generated a full-length sgRNA and two shorter sgRNA constructs that were selectively truncated from either the 5′ or 3′ end (Fig. 2B and Table S2) and determined their affinities for WT Cas9, the individual α-helical and nuclease lobes, and split-Cas9 using a filter binding assay.

The full-length sgRNA is bound by WT Cas9 with an equilibrium dissociation constant ($K_d$) of $10 \pm 2$ PM whereas the lobes individually and together have $K_d$ values in the range of 0.2–0.8 nM (Fig. 2C and Table 1). The difference between WT and split-Cas9 likely reflects the increased entropic cost required to assemble a ternary versus binary complex. Interestingly, WT Cas9 bound a truncated sgRNA comprising only the 3′ hairpins ($Δ$hairpins1-2) would selectively perturb interactions with the bridge helix. Based on this interaction profile, we generated a full-length sgRNA and two shorter sgRNA constructs that were selectively truncated from either the 5′ or 3′ end (Fig. 2B and Table S2) and determined their affinities for WT Cas9, the individual α-helical and nuclease lobes, and split-Cas9 using a filter binding assay.

We similarly reasoned that removing the two hairpins from the 3′ end of the sgRNA ($Δ$hairpins1-2) would selectively perturb interactions with the nuclease lobe. Indeed, the affinity of the 3′-truncated sgRNA for the nuclease lobe decreased by over three orders of magnitude relative to full-length sgRNA ($K_d > 100$ nM).

We similarly reasoned that removing the two hairpins from the 3′ end of the sgRNA ($Δ$hairpins1-2) would selectively perturb interactions with the nuclease lobe. Indeed, the affinity of the 3′-truncated sgRNA for the nuclease lobe decreased by over three orders of magnitude relative to full-length sgRNA ($K_d > 100$ nM).

We similarly reasoned that removing the two hairpins from the 3′ end of the sgRNA ($Δ$hairpins1-2) would selectively perturb interactions with the nuclease lobe. Indeed, the affinity of the 3′-truncated sgRNA for the nuclease lobe decreased by over three orders of magnitude relative to full-length sgRNA ($K_d > 100$ nM).

We similarly reasoned that removing the two hairpins from the 3′ end of the sgRNA ($Δ$hairpins1-2) would selectively perturb interactions with the nuclease lobe. Indeed, the affinity of the 3′-truncated sgRNA for the nuclease lobe decreased by over three orders of magnitude relative to full-length sgRNA ($K_d > 100$ nM).

We similarly reasoned that removing the two hairpins from the 3′ end of the sgRNA ($Δ$hairpins1-2) would selectively perturb interactions with the nuclease lobe. Indeed, the affinity of the 3′-truncated sgRNA for the nuclease lobe decreased by over three orders of magnitude relative to full-length sgRNA ($K_d > 100$ nM).

We similarly reasoned that removing the two hairpins from the 3′ end of the sgRNA ($Δ$hairpins1-2) would selectively perturb interactions with the nuclease lobe. Indeed, the affinity of the 3′-truncated sgRNA for the nuclease lobe decreased by over three orders of magnitude relative to full-length sgRNA ($K_d > 100$ nM).
the sgRNA and either lobe. In agreement with this hypothesis, we found that the in vitro DNA cleavage activity of split-Cas9 decreased as the sgRNA concentration was increased above that of both lobes (Fig. S5), suggesting that excess sgRNA likely titrates the lobes apart from each other. Although our results leave room for optimization of split-Cas9 activity in cells, they demonstrate that the intrinsic genome-editing capabilities are retained when Cas9 comprises two individual polypeptides.

**Engineered sgRNAs Selectively Preclude Split-Cas9 but Not WT Cas9 Activity.** The potential for enhanced spatiotemporal control of genome engineering events with split-Cas9 prompted us to investigate ways in which sgRNA-mediated dimerization of the α-helical and nuclease lobes could be perturbed. In particular, we reasoned that certain 3′-truncated or modified sgRNAs, which have weak affinity for the nuclease lobe (Fig. 2E) but still support robust DNA cleavage activity of WT Cas9 (4), would selectively inactivate split-Cas9 activity through their inability to effectively recruit and dimerize both lobes into a functional enzyme complex. Thus, the activity of split-Cas9 in cells could be made dependent upon inducible protein–protein dimerization domains (Fig. S6).

When we tested sgRNA variants that lacked one or both hairpins at the 3′ end for their ability to support in vitro cleavage, split-Cas9 activity was either severely compromised or completely abolished whereas WT Cas9 activity was slightly reduced relative to a full-length sgRNA (Fig. 3C, Fig. S7, and Table S1). A recent report found that sgRNAs in which only the first hairpin is deleted function robustly in cells (11), and we found that similar designs supported DNA cleavage activity of WT Cas9 (4), but not split-Cas9 in vitro (Fig. S7). Thus, rationally designed variants of the sgRNA scaffold can be used to prevent RNA-mediated heterodimerization of the two lobes without compromising the intrinsic RNA-guided DNA-cleaving capabilities of Cas9.

**Discussion**

Here, we have successfully designed a split version of Cas9 that maintains the cleavage activity of the native enzyme. We demonstrated that the sgRNA is necessary and sufficient to dimerize the nuclease and α-helical lobes into an active complex and furthermore showed that multiple distinct sgRNA motifs interact with the lobes independently. Split-Cas9 is active for genome editing in cells, albeit at a reduced level relative to WT Cas9, and can be rendered nonfunctional through the removal of one or more of the hairpins at the 3′ end of sgRNA. Although optimization will help split-Cas9 to function effectively in cells, we have shown the potential to enable a variety of interesting and useful applications.

Split-protein systems have often been designed such that the functional unit is restored through the interactions of an exogenous pair of proteins (14). For example, DNA-binding and effector domains of modified TALEs have been fused to CRY2 and CIB1 to enable light-inducible control of gene expression (15). Alternatively, genes may be split such that a single functional polypeptide is the final product of mRNA splicing or intein-based protein splicing (16). Our strategy with split-Cas9 differs in that it faithfully recapitulates the functionality of full-length Cas9 using the very same RNA ligand that WT Cas9 requires. In this sense, sgRNA-mediated dimerization and activation of split-Cas9 may be viewed analogously to the sgRNA-mediated rearrangement and structural activation of WT Cas9.

Optimizing expression levels of the split-Cas9 components could increase their effectiveness. In particular, whereas the sgRNA should be kept limiting to avoid titrating the lobes away from the ternary complex, overall expression must be high enough to overcome the reduced affinity of both protein components for the sgRNA scaffold. Split-Cas9 could be regulated by the combinatorial use of promoters, restricting activity to highly specific subsets of tissues or creating a “coincidence detector” with two inducible promoters. Split-Cas9 could also be developed for use with adeno-associated viral vectors, where the smaller coding regions of each lobe would enable the use of effector or reporter domains that are currently prohibited by limited packaging capacity. We also suggest that split-Cas9 can be converted into a regulatable system using exogenous dimerization domains (Fig. S6). Fusing both lobes to domains that selectively dimerize upon chemical or optical induction, such as the acidic acid-inducible PYL-ABI module (17) or the blue light-inducible CRY2-CIB1 dimer (18), would allow for engineered spatiotemporal control of genome-engineering events (6). Dimerization domains may also increase the efficiency of complex formation by making lobe assembly independent of the sgRNA. We propose that the combined use of inducible dimerization domains with compromised sgRNA variants that enable DNA targeting, but not split-Cas9 assembly, would eliminate leaky activity in the absence of inducer while still allowing for robust activation, creating an extremely sensitive inducible system.

Finally, our study provides important insights into the structure–function relationship of the native Cas9 enzyme. The ability of sgRNA to act as a molecular scaffold in assembling two separate polypeptides highlights the crucial role that the sgRNA plays in orchestrating conformational rearrangements of WT Cas9. The separation of recognition and catalytic functions into two separate lobes may have evolved to eliminate nonspecific nuclease activity and control licensing of Cas9 for DNA interrogation. Our results also invite comparisons to the mechanisms of ICPN assembly and DNA targeting in other CRISPR-Cas systems, particularly the type I-E Cascade interference complex. Although Cascade is composed of 11 distinct subunits, none of which possesses nuclease activity, the guide RNA (crRNA) plays a similarly critical role in scaffolding the assembly of distinct domains into a structure that is primed to engage DNA targets (19–21). Similar principles are likely to govern the assembly and activity of other CRISPR RNA-guided DNA-targeting complexes (2). Thus, distinct CRISPR-Cas systems may have evolved similar organizational strategies in parallel that use the guide RNA for structural assembly and conformational activation.

**Materials and Methods**

**Cloning and Protein Purification.** The expression vector for purification of the nuclease lobe was generated by Around the Horn (ATH) PCR using a preexisting pET-based expression vector for S. pyogenes Cas9. The final construct encodes an N-terminal decahistidine-maltose binding protein (His10-MBP) tag, a tobacco etch virus (TEV) protease cleavage site, residues 1–57, a glycine-serine-serine linker, and residues 729–1,368. The vector for the catalytically inactive dNuclease lobe was generated by ATH PCR of a similar dCas9 (D10A/H840A) vector. The vector for expression of the α-helical lobe was generated by PCR amplification of S. pyogenes Cas9 residues 56–714 and assembly of the resulting fragment into a His10-MBP expression vector via ligation-independent cloning.

Each protein was overexpressed in E. coli BL21 Rosetta 2(DE3) (EMD Biosciences) by growing in 2xYT medium at 37 °C to an optical density of 0.5,
inducing with 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG), and growing an additional 16 h at 18 °C. Cells were lysed by sonication in a buffer containing 50 mM Tris, pH 7.5, 500 mM NaCl, 1 mM Tris-(2-carboxyethyl)phosphine (TCEP), 5% (vol/vol) glycerol, and a protease inhibitor mixture (Roche). The clarified lysate was extensively filtered and split-Cas9 ribonucleoprotein (RNP) complexes after arrest at mitosis with nocodazole (Sync) or during normal growth (Unsync). Editing efficiencies are shown at the bottom. (C) DNA cleavage time courses using WT and split-Cas9 with either a full-length sgRNA (Top) or the Δhairpins-1-2 sgRNA (Bottom). Values were averaged from three independent experiments, and error bars represent the SD. Rate constants can be found in Table S1.

DNA Cleavage Assays. All cleavage assays were performed in 1x cleavage buffer, which contains 20 mM Tris, pH 7.5, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, and 5% (vol/vol) glycerol. Preformed complexes were diluted in cleavage buffer, and reactions were initiated with the addition of radiolabeled dsDNA substrates. Final reaction concentrations were 100 nM protein:RNA complex and ~1 nM radiolabeled DNA target. The concentration of Cas9 was chosen to be sufficiently above the K₅₀ for the sgRNA such that complex assembly is unlikely to be rate-limiting, except in the case of split-Cas9 and the Δhairpins-1-2 sgRNA (K₅₀ > 100 nM). Reactions proceeded at room temperature, and aliquots were removed at selected time points and quenched with an equal volume of buffer containing 50 mM EDTA, 0.02% bromophenol blue, and 90% (vol/vol) formamide. Reaction products were resolved by 7 M urea-PAGE, gels were dried, and DNA was visualized by phosphorimaging and quantified using ImageQuant software (GE Healthcare). The percentage of DNA cleaved was determined by dividing the amount of cleaved DNA by the sum of uncleaved and cleaved DNA. Kinetic analysis was performed in GraphPad (GraphPad Software). Observed rate constants (kobs) are the average of three independent experiments ± SEM. Graphed values are the averaged time points of three independent experiments, with error bars representing the SD.

Electrophoretic Mobility-Shift Assays. All binding assays were performed in 1x binding buffer, after which radiolabeled dsDNA substrates were added to a final concentration of <0.2 nM. Reactions were incubated at room temperature for 60 min and then resolved at 4 °C on a native 8% polyacrylamide gel containing 0.5x Trisborate/EDTA (TBE) and 5 mM MgCl₂. Gels were dried, and DNA was visualized by phosphorimaging and quantified using ImageQuant software. Electrophoresis fraction of DNA eluted from the gel and resolved by 7 M urea-PAGE were analyzed. Graphed values are the averaged time points of three independent experiments, with error bars representing the SD.
transmission electron microscope operated at 120 kV at a nominal magnification of 80,000x (1.45 Å at the specimen level) using low-dose exposures (−20 e− Å−2) with a randomly set defocus ranging from −0.7 to −1.6 μm. A total of 150–200 images of each Cas9 sample was automatically recorded on a Gatan 4k × 4k CCD camera using the M3i-Raster application within LEGINON (22). Low-resolution negative-stain class averages of Lid particles from the yeast 26S proteasome (23) were used as references for template-based particle picking. The Lid complex was used as a template to avoid selection bias because it bears minimal to no structural resemblance to Cas9. Cas9 complexes were extracted using a 224 × 224-pixel box size. These particles were subjected to 2D reference-free alignment and classification using multivariate statistical analysis and multireference alignment in IMAGIC (24).

Cas9 and Split-Cas9 RNP Assembly and Nucleasection. The split-Cas9 RNP was prepared immediately before the experiment by incubating both lobes with sgRNA at molar ratios of 1.2:1:1.2 (α-helical lobe:nuclease lobe:sgRNA) for 10 min at 37 °C in 20 mM Hepes, pH 7.5, 150 mM KC1, 1 mM MgCl2, 10% (vol/vol) glycerol, and 1% TCEP. The nucleofections were carried out as previously described for Cas9, using 10, 30, and 100 pmol of RNP complex for ~2 × 10^5 cells (13). Where indicated, cells were synchronized with 200 ng/mL nocodazole for 17 h before nucleasection. Neither WT Cas9 nor the split-Cas9 lobes had nuclear localization signals, which may have led to reduced editing levels, particularly for the unsynchronized cells.

Analysis of In-Cell Genome-Editing Efficiency. Determination of the percentage of indels induced at the target region was performed as previously described (13). In brief, 640-nt regions of the EMX1 locus containing the target sites were PCR-amplified, and the resulting products were denatured, renatured, and digested with T7 endonuclease I (New England Biolabs), which cleaves mismatched heteroduplex DNA (25). The products were resolved on a 2% (wt/vol) agarose gel containing SYBR Gold (Life Technologies), and band intensities were determined using Image Lab (Bio-Rad Laboratories). Editing efficiencies were determined using the formula \( \frac{1}{2} \left[ 1 - (b + c + a) \right] \times 100 \), where \( a \) is the band intensity of DNA substrate and \( b \) and \( c \) are the cleavage products (26).

ACKNOWLEDGMENTS. We thank B. Castellano and M. Kaplan for assisting with cloning and protein purification; E. Nogales for use of EM resources and expertise in EM analysis; and members of the J.A.D. laboratory for helpful discussions. S.H.S. is a National Science Foundation Graduate Research Fellow and a National Defense Science and Engineering Graduate Research Fellow. A.V.W. is a National Science Foundation Graduate Fellowship. B.T.S. is a Roche Postdoctoral Fellow, RPF 311. J.A.D. is a Howard Hughes Medical Institute investigator.

Fig. S1. Size-exclusion chromatograms of purified α-helical (A) and nuclease (B) lobes. After cleavage of the affinity tag by TEV protease and further clean-up using ortho-Ni-NTA and ion-exchange columns (α-helical lobe) or an ortho-Ni-NTA column alone (nuclease lobe), the polypeptides were concentrated and injected onto a HiLoad 16/60 Superdex 200 gel filtration column. The α-helical (77 kDa) and nuclease lobes (81 kDa) eluted at 88.4 mL and 76.8 mL, respectively. Both polypeptides were soluble and exhibited consistent activity across multiple rounds of freeze-thawing.
Fig. S2. Split-Cas9 activity is mediated by single-guide and dual-guide RNAs and requires RNA:DNA complementarity and a PAM. (A) DNA cleavage time courses using a single-guide RNA and WT Cas9, individual α-helical and nuclease lobes, or split-Cas9. Values for WT and split-Cas9 were averaged from three independent experiments, and error bars represent the SD. Rate constants can be found in Table S1. (B) DNA cleavage time courses using a dual-guide RNA (crRNA:tracrRNA hybrid) and WT Cas9 or split-Cas9. Data are presented as in A. (C) DNA cleavage assay with split-Cas9 and DNA substrates containing a mismatched target or mutated PAM (Table S2), analyzed by denaturing PAGE. Reactions contained ~1 nM radiolabeled dsDNA and 100 nM Cas9–sgRNA complex.

Fig. S3. Split-Cas9 exhibits substantially weaker binding affinity for target DNA than WT Cas9. (A) Radiolabeled target dsDNA was incubated with increasing concentrations of Cas9–sgRNA complexes using catalytically inactive mutants of WT Cas9 and the nuclease lobe, and reaction products were resolved by native PAGE. The distinct Cas9 constructs in each titration are indicated at the top. (B) Quantified binding data from A. Split dCas9–RNA binds dsDNA with an apparent equilibrium dissociation constant of ~700 nM, which is more than three orders of magnitude greater than that determined for dCas9–RNA (K_d ~0.2 nM). However, the apparent affinity measured here is likely to be much weaker than the actual affinity because the low split dCas9–sgRNA concentrations that were tested will also favor dissociation of the ternary complex formed between the sgRNA, α-helical lobe, and nuclease lobe. Thus, the observed binding curve is likely a convolution of equilibria between the protein and sgRNA, and between the protein–sgRNA complex and dsDNA. Individual lobes together with sgRNA do not appreciably bind dsDNA at the tested concentrations.
Fig. S4. Split-Cas9 heterodimerization requires the sgRNA. (A–D) Raw electron micrographs of negatively stained α-helical and nuclease lobes alone (A and B), together (C), or together with sgRNA (D). Particles having dimensions consistent with WT Cas9–RNA complexes, and thus indicative of heterodimer formation, are observed only in the presence of sgRNA. Representative particles are circled (yellow). (Scale bars: 50 nm.)

Fig. S5. Excess sgRNA reduces the DNA cleavage activity of split-Cas9. DNA cleavage assay with varying molar ratios of protein to sgRNA, analyzed by de-naturing PAGE. Reactions contained ∼1 nM radiolabeled dsDNA, 100 nM α-helical and nuclease lobes, and 50–1,000 nM sgRNA. The extent of product formation decreases substantially as the sgRNA concentration surpasses the lobe concentration. This observation suggests that stoichiometric excesses of sgRNA titrate the individual lobes away from each other and onto independent sgRNA molecules, a hypothesis supported by the finding that distinct sgRNA motifs interact with either lobe.
Fig. S6. Strategy for inducible control of genome engineering by a split-Cas9 enzyme complex. (A) Because the α-helical and nuclease lobes dimerize in the presence of sgRNA, both WT and split-Cas9 are functional genome editing tools in cells using full-length sgRNA. (B) sgRNA variants with 3′-hairpin truncations have substantially weaker affinity for the nuclease lobe and thus do not efficiently assemble a functional split-Cas9 complex, leading to an inactive enzyme. In contrast, in vitro DNA cleavage by WT Cas9 is minimally affected by these truncations, indicating that the intrinsic activity of the Cas9–sgRNA enzyme complex does not require hairpins at the 3′ end. (C) We propose an inducible split-Cas9 system, in which exogenous dimerization domains control the assembly of a functional ternary complex between a 3′-truncated sgRNA and the α-helical and nuclease lobes. By fusing both lobes to domains that dimerize only upon some external stimulus (e.g., a small molecule; red trapezoid), split-Cas9 can be specifically activated for a desired genome-engineering outcome.
Fig. S7. The 3′-truncated sgRNA variants selectively inactivate split-Cas9. (A and B) DNA cleavage assays with WT and split-Cas9 and a panel of four different sgRNAs, analyzed by denaturing PAGE. (A) Full-length sgRNAs promote DNA cleavage activity of both WT and split-Cas9 whereas split-Cas9 activity is completely lost with an sgRNA lacking both hairpins at the 3′ end (Δhairpins1-2). (B) sgRNA variants where only one hairpin is removed show minimal effects on WT Cas9 activity but severely (Δhairpin2) or completely (Δhairpin1) inactivate split-Cas9.

Table S1. Cleavage rate constants for WT and split-Cas9 using different sgRNA constructs

<table>
<thead>
<tr>
<th>Protein</th>
<th>Rate constant (k_{obs}) for indicated sgRNA*, min^{-1}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full-length</td>
</tr>
<tr>
<td>WT Cas9</td>
<td>11.3 ± 0.9</td>
</tr>
<tr>
<td>α-Helical lobe</td>
<td>N.D.</td>
</tr>
<tr>
<td>Nuclease lobe</td>
<td>N.D.</td>
</tr>
<tr>
<td>Split-Cas9</td>
<td>1.0 ± 0.2</td>
</tr>
</tbody>
</table>

N.D., cleavage not detected; —, experiment not performed.
*Three independent experiments were performed for each condition, and the values represent the mean ± SEM.
Table S2. DNA and RNA substrates used in this study

<table>
<thead>
<tr>
<th>Description</th>
<th>Sequence*</th>
<th>Used in Figs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ1 target dsDNA</td>
<td>5’-AGCCAGAAATCTCCTGCTGACCGCAT2AAGATTGAGACGCGCTGATACAAACGCTCACT-3’</td>
<td>1D, 3C, 52A</td>
</tr>
<tr>
<td></td>
<td>3’-TCGCTTTTAAAGACGACGCTCGTTTTCTACTCTGCACCTCAATGGCTAGTGA-5’</td>
<td>52B, 53, 55, 57</td>
</tr>
<tr>
<td>λ1 target dsDNA, mutated PAM</td>
<td>5’-AGCAAGAAYCTCCTGCTGAGCCAAATTGACGACGCTAGTAACAAGTCAGCT-3’</td>
<td>52C</td>
</tr>
<tr>
<td>λ2 target dsDNA, (λ1 mismatch)</td>
<td>5’-GAGGTTAGATGCTGCAGTGATAAAAGGGAATGGCATGAGGCTGTCAAAATTGATAGC-3’</td>
<td>52C</td>
</tr>
<tr>
<td></td>
<td>3’-CTCACCTTTCCAGGTCGACTATCCCTTACGGTTACGACCCGAGACTTTTTAAATCTCG-5’</td>
<td></td>
</tr>
<tr>
<td>λ1 sgRNA, full-length</td>
<td>5’-GACGCCAUAAAGAUGAGAGCAGGCUUUGAGCCUUUGCCUGUUUGGAAACCACAGCAUGGCCGACGCACGGCGACGCUUUGGUUGACU-3’</td>
<td>1D, 2C, 2E, 3C, 52A, 52C, 53, 54D, 55, 57A</td>
</tr>
<tr>
<td>λ1 crRNA</td>
<td>5’-GACGCCAUAAAGAUGAGAGCAGGCUUUGAGCCUUUGCCUGUUUGGAAACCACAGCAUGGCCGACGCACGGCGACGCUUUGGUUGACU-3’</td>
<td>2E, 52B</td>
</tr>
<tr>
<td>TracrRNA</td>
<td>5’-GACGCCAUAAAGAUGAGAGCAGGCUUUGAGCCUUUGCCUGUUUGGAAACCACAGCAUGGCCGACGCACGGCGACGCUUUGGUUGACU-3’</td>
<td>2E, 52B</td>
</tr>
<tr>
<td>λ1 sgRNA, Δhairpin1 and -2</td>
<td>5’-GACGCCAUAAAGAUGAGAGCAGGCUUUGAGCCUUUGCCUGUUUGGAAACCACAGCAUGGCCGACGCACGGCGACGCUUUGGUUGACU-3’</td>
<td>2D, 3C, 57A</td>
</tr>
<tr>
<td>λ1 sgRNA, Δhairpin2</td>
<td>5’-GACGCCAUAAAGAUGAGAGCAGGCUUUGAGCCUUUGCCUGUUUGGAAACCACAGCAUGGCCGACGCACGGCGACGCUUUGGUUGACU-3’</td>
<td>57B</td>
</tr>
<tr>
<td>λ1 sgRNA, Δhairpin1</td>
<td>5’-GACGCCAUAAAGAUGAGAGCAGGCUUUGAGCCUUUGCCUGUUUGGAAACCACAGCAUGGCCGACGCACGGCGACGCUUUGGUUGACU-3’</td>
<td>57B</td>
</tr>
<tr>
<td>λ1 sgRNA, Δspacer--nexus</td>
<td>5’-GACGCAUAAGAUAAGAGAAGAGCAGGCUUUGAGCCUUUGCCUGUUUGGAAACCACAGCAUGGCCGACGCACGGCGACGCUUUGGUUGACU-3’</td>
<td>2D</td>
</tr>
<tr>
<td>EMX1 sgRNA, full-length</td>
<td>5’-GACGCAUAAGAUAAGAGAAGAGCAGGCUUUGAGCCUUUGCCUGUUUGGAAACCACAGCAUGGCCGACGCACGGCGACGCUUUGGUUGACU-3’</td>
<td>3B</td>
</tr>
</tbody>
</table>

*RNA guide sequences and complementary DNA target strand sequences are shown in red; PAM sites are highlighted in yellow.