



Cell-free gene-regulatory network engineering with synthetic transcription factors

Zoe Swank^{a,1}, Nadanai Laohakunakorn^{a,1}, and Sebastian J. Maerkl^{a,2}

^aInstitute of Bioengineering, School of Engineering, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Edited by Michael C. Jewett, Northwestern University, Evanston, IL, and accepted by Editorial Board Member David Baker February 7, 2019 (received for review September 25, 2018)

Gene-regulatory networks are ubiquitous in nature and critical for bottom-up engineering of synthetic networks. Transcriptional repression is a fundamental function that can be tuned at the level of DNA, protein, and cooperative protein–protein interactions, necessitating high-throughput experimental approaches for in-depth characterization. Here, we used a cell-free system in combination with a high-throughput microfluidic device to comprehensively study the different tuning mechanisms of a synthetic zinc-finger repressor library, whose affinity and cooperativity can be rationally engineered. The device is integrated into a comprehensive workflow that includes determination of transcription-factor binding-energy landscapes and mechanistic modeling, enabling us to generate a library of well-characterized synthetic transcription factors and corresponding promoters, which we then used to build gene-regulatory networks de novo. The well-characterized synthetic parts and insights gained should be useful for rationally engineering gene-regulatory networks and for studying the biophysics of transcriptional regulation.

cell-free synthetic biology | biophysics | transcriptional regulation | gene regulatory network | synthetic transcription factors

Cell-free systems have emerged as versatile and efficient platforms for rapid engineering, characterization, and implementation of genetic networks. It has been demonstrated that linear genetic cascades (1), logic gates (2), and oscillators (3–5) could be implemented and characterized in cell-free systems and that networks engineered in cell-free systems function in cells with remarkably similar characteristics, indicating that cell-free systems accurately emulate the cellular environment (4, 6). Besides these examples in molecular-systems engineering and characterization of complex biological systems, cell-free systems provide a viable starting point for the bottom-up synthesis of artificial cells (7, 8). Work is progressing in establishing critical cellular subsystems, including DNA replication (9), metabolism (10), ribosome synthesis (11), membrane synthesis (12), and protein structures (13). Gene-regulatory networks (GRNs) are one such critical subsystem, and here we demonstrate de novo bottom-up engineering and comprehensive characterization of synthetic GRNs in a cell-free system.

GRNs execute the genome and thus play a central role across all domains of life. Due to their importance and ubiquity, GRNs have been intensely studied, and considerable progress is being made in deciphering components, topologies, and general mechanisms of GRNs, although a complete mechanistic understanding is still lacking. Because GRNs perform many sophisticated cellular tasks, synthetic biologists use GRNs to engineer new systems (14) such as logic gates (15), toggle switches (16), band-pass filters (17), and oscillators (18). Nonetheless, past and current efforts in engineering GRNs have shown that rational design is not yet possible and that engineering GRNs still heavily relies on trial-and-error and high-throughput screening approaches (15). The inability to rationally design GRNs is in part due to the aforementioned lack of complete mechanistic understanding and because basic GRN components such as transcriptional regulators and promoters are often neither

fully characterized nor standardized. A corollary of the lack of an in-depth mechanistic understanding of these systems is that individual components are not yet readily composable. Nature provides a plethora of potential transcriptional regulators, but the number that have been tested and characterized remains rather limited. Most engineered GRNs make use of naturally occurring transcription factors, making it difficult to robustly engineer GRNs with such a nonstandard set of proteins (19). A library of well-characterized, synthetic transcription factors could alleviate many of these problems by providing a set of standardized transcription factors that are based on the same basic structural framework and whose function can be extended by generating fusion proteins in a plug-and-play format.

Native GRNs use a wide range of transcription factors that can be categorized into several structural families. The family with the largest number of members is the zinc-finger (ZF) family, followed by homeodomain, basic helix–loop–helix, and basic leucine-zipper (LZ) families (20). ZFs are of interest in biology, as they represent the largest class of transcriptional regulators and are involved in diverse biological functions. ZFs are also appealing for bottom-up engineering, as they consist of well-defined subunits that, in combination, determine DNA sequence specificity (21, 22). Many resources are therefore available that provide sequence-specificity information for a large number of native (23) and engineered (24) ZF transcription factors. An additional advantage is that ZFs are small [264 bp, 10.6 kDa

Significance

Understanding basic mechanisms and engineering new systems in biology are hampered by the challenge of quantifying and manipulating numerous molecular components and their interactions. In this work, we present an approach to tackle these difficulties, by combining a cell-free system with a microfluidic platform for high-throughput measurements. We apply the system to comprehensively characterize a library of synthetic zinc-finger transcription factors, which are common building blocks of transcriptional regulatory networks. We subsequently use the knowledge gained to engineer highly specific, tunable, and strong cooperative repressors, which can be applied to carry out logical computation on promoters.

Author contributions: Z.S., N.L., and S.J.M. designed research; Z.S. and N.L. performed research; Z.S., N.L., and S.J.M. analyzed data; and Z.S., N.L., and S.J.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. M.C.J. is a guest editor invited by the Editorial Board.

Published under the [PNAS license](#).

Data deposition: All supporting data and code are available on GitHub at <https://github.com/lbnc-epfl>.

¹Z.S. and N.L. contributed equally to this work.

²To whom correspondence should be addressed. Email: sebastian.maerkl@epfl.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1816591116/-DCSupplemental.

(Zif268)] compared with other engineerable transcriptional regulators, such as TALE [e.g., 1,161–2,397 bp, 39.9–82.6 kDa, DNA binding domain only (25)] or dCas9 (4,107 bp, 158.3 kDa), so that the coding sequence for ZFs is easily obtainable and modifiable. Due to their small size and simple structure, ZFs can be readily expressed both *in vivo* and *in vitro*. Synthetic ZFs have already been successfully used as activators in *Saccharomyces cerevisiae* (26) and human cells (27). Here, we engineer and explore the use of synthetic ZF transcriptional regulators as ideal building blocks for bottom-up design and implementation of cell-free GRNs.

In this paper, we took advantage of an existing synthetic ZF library (28) to generate a well-characterized resource of transcriptional repressors and corresponding synthetic promoters that can be used for bottom-up design, implementation, and characterization of GRNs in cell-free systems. While the mechanism of action of the simplest prokaryotic repression is competitive inhibition (29), it has long been appreciated that both *cis* modifications to the promoter, such as operator position (30), and basal promoter strength (31), as well as *trans* modifications to the transcription factor itself strongly affect repression (32, 33). These interdependencies result in a large experimental space with many degrees of freedom. To tackle this complexity, we developed a microfluidics-based method capable of performing 768 cell-free transcription–translation (TX-TL) reactions on a single device. The ability to rapidly generate ZF repressor and promoter variants by using fast-PCR assembly and the use of our high-throughput microfluidic device allowed us to perform a comprehensive characterization of repressors and promoters. We investigated the effects of binding-site position, binding-site affinity, binding-site combinations, and cooperative interactions between the repressors on transcriptional repression performance. We generated quantitative position weight matrices (PWMs) for four ZF repressors with MITOMI (34), which allowed us to rationally tune binding-site affinity and promoter output. Finally, we used the parts library and insights acquired in this study to engineer logic gates, showing that *de novo* synthetic GRNs can be rationally engineered by using a bottom-up approach. The transcription factor/promoter parts library, data, and methods described here provide a resource that should facilitate efforts to build synthetic GRNs, serve as a viable approach for building GRNs for use in artificial cells, and establish an experimental platform for studying the biophysics of transcriptional regulation.

Results

Design and Characterization of a Microfluidic Device for High-Throughput Cell-Free Experiments. The design space of even a single TF–promoter pair is large, encompassing different binding-site affinities, binding-site positions, binding-site sequences, and binding-site combinations. This complexity necessitates high-throughput methods capable of the functional characterization of hundreds to thousands of engineered variants. Current approaches in cell-free synthetic biology primarily rely on standard microtiter plates, which require a minimal reaction volume of 5–10 μ L. Such relatively large volumes quickly become cost-limiting in terms of how much cell-free reaction solution and DNA are required to perform the assays. Researchers recently made use of an acoustic liquid-handling robot that reduced reaction volumes to 2 μ L in a 384-well plate format (35). Here, we repurposed the MITOMI platform, a microfluidic device originally developed for high-throughput molecular-interaction analysis (34, 36), and applied it to the high-throughput characterization of cell-free genetic networks. The repurposed device performs 768 cell-free reactions and reduces volumes by \sim 4 orders of magnitude to \sim 690 pL per reaction.

The process involves the synthesis of DNA parts, followed by microarraying and incorporation into microfluidic unit cells,

where they serve as templates in cell-free TX-TL reactions (Fig. 1A). To expedite the synthesis of large libraries of DNA parts, we used an assembly PCR strategy to generate linear DNA templates with different promoter regions upstream of an eGFP-Del6-229 (deGFP) gene. A microarray robot was used to spot the linear templates onto an epoxy-coated glass slide,

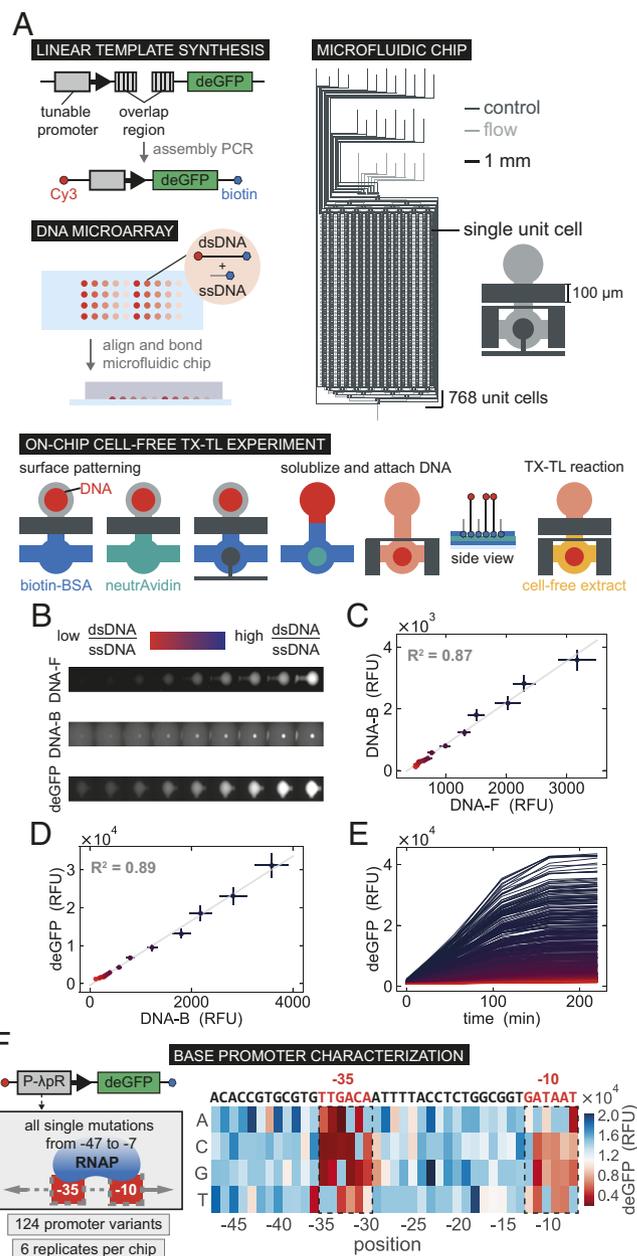


Fig. 1. High-throughput microfluidic cell-free reactions. (A) A schematic overview of the experimental design, including synthesis of DNA templates, DNA spotting, chip alignment, surface functionalization followed by DNA immobilization, and on-chip cell-free TX-TL reactions. (B) Fluorescence images of Cy3-DNA_F, -DNA_B, and deGFP expressed for a range of dsDNA:ssDNA ratios. (C) Quantification of surface-immobilized DNA (DNA_B) as a function of free DNA in solution (DNA_F). (D) deGFP expression at the final time point as a function of DNA_B concentration. All values represent means \pm SD ($n = 48$). (E) deGFP expression measured over time in all unit cells. (F) Schematic of a promoter library design and on-chip experimental throughput, followed by the deGFP expression for all single base mutations from position -47 to -7 of the λP_R promoter. RFU, relative fluorescence units.

on top of which the polydimethylsiloxane (PDMS) device was aligned. Immobilizing DNA within each reaction chamber first required surface patterning in the assay section of each unit cell, resulting in a circular area of neutravidin, to which biotinylated DNA can bind. Once DNA was surface-immobilized, cell-free extract was flowed into the device, and the unit cells were isolated from one another while the TX-TL reactions occurred. A detailed schematic of the experimental procedure is shown in *SI Appendix, Fig. S1A*.

Controlling the precise amount of DNA in each unit cell is important for quantitative experiments. By simply varying the concentration of spotted biotinylated DNA templates, we were unable to precisely control DNA concentration on-chip. We thus developed an approach based on spotting a mixture of single-stranded biotinylated DNA oligos (ssDNA) and double-stranded DNA templates (dsDNA). The amount of DNA immobilized on the surface reached saturation at a concentration of ~ 100 nM spotted DNA (*SI Appendix, Fig. S2*). We therefore held the total concentration of spotted DNA above this saturation point. Changing the ratio of dsDNA:ssDNA gave rise to a linear correlation between the concentration of dsDNA free in solution (DNA_F) and dsDNA bound to the surface (DNA_B) and was insensitive to the total amount of DNA deposited during spotting (Fig. 1 *B* and *C*). This approach allowed us to immobilize DNA over a wide concentration range, which gave rise to corresponding levels of expressed deGFP (Fig. 1 *D* and *E*). The results obtained with the high-throughput microfluidic device are reproducible with a global normalized root-mean-square deviation of $\sim 14\%$, not only when a single dsDNA template is used, but also for more complex experiments requiring multiple templates in each unit cell (*SI Appendix, Fig. S3*). Furthermore, a subset of on-chip measurements was carried out in standard microwell plate reactions, showing good correlation (*SI Appendix, Fig. S4*).

To demonstrate the high-throughput capabilities of our microfluidic chip, we created and characterized a library based on the *Escherichia coli* σ_{70} λP_R promoter. We synthesized 124 promoter variants that covered all possible single-base mutations within the -47 to -7 region of the λP_R promoter (Fig. 1*F*). Cell-free reactions for each promoter were run in six replicates on a single chip and yielded deGFP expression profiles revealing the impact of each mutation on protein expression (Fig. 1*F* and *SI Appendix, Fig. S3A*). As expected, mutations within the -10 and -35 boxes affected deGFP expression most strongly, and the

results are comparable to previous results obtained by an in vivo analysis of the lac promoter (37).

Protein synthesis eventually stopped in cell-free batch reactions, as seen in the saturation dynamics in time-course measurements (Fig. 1*E*); this is fundamentally different from cellular steady-state protein levels which result from balancing production with degradation and dilution rates. In this work, we report end-point batch reaction values and derived quantities such as fold repression. It is thus important that the end-point values correspond to protein production rates. While the relationship between the initial rate of deGFP production and its final saturated level may be complex, we observed a linear relationship between the two quantities under our experimental conditions (*SI Appendix, Fig. S5*). This is an important validation of our use of end-point protein levels and linearly derived quantities such as fold repression as proxies for synthesis rates and their ratios.

ZF Repressor and Promoter Library Design. Using the characterization of the λP_R promoter as a starting point, we applied our chip to the in-depth characterization of synthetic ZFs for use as transcriptional repressors. We adopted a ZF design based on Zif268, a three-finger Cys2–His2 protein. A large ZF repressor library can be generated by combinatorially shuffling a small number of individual ZF domains (Fig. 2*A*). We used ZF proteins drawn from a 64-member library that we previously synthesized and characterized (28) (*SI Appendix, Fig. S6*).

The affinity of a ZF repressor to DNA can be improved by increasing the number of finger domains (38–41). The same effect can also be achieved by engineering dimerizing ZFs that bind cooperatively. An early example used structure-based design to engineer a two-finger ZF which dimerized via a LZ motif to form a four-finger complex (42, 43). Three-finger ZFs have also been dimerized by using PDZ domains (26). Cooperative interactions are of interest because they potentially increase the nonlinearity of regulation, as well as decrease nonspecific binding compared with extended arrays of ZFs. To study cooperative interactions, we built several different ZFs fused to either PDZ or LZ domains (Fig. 2*B*).

In parallel, we designed corresponding repressible promoter libraries. As we used an *E. coli* cell-free system (44), we based our promoter designs on the strong λP_R promoter in combination with transcription and translation elements optimized for *E. coli* cell-free expression (45). Previous work has shown that

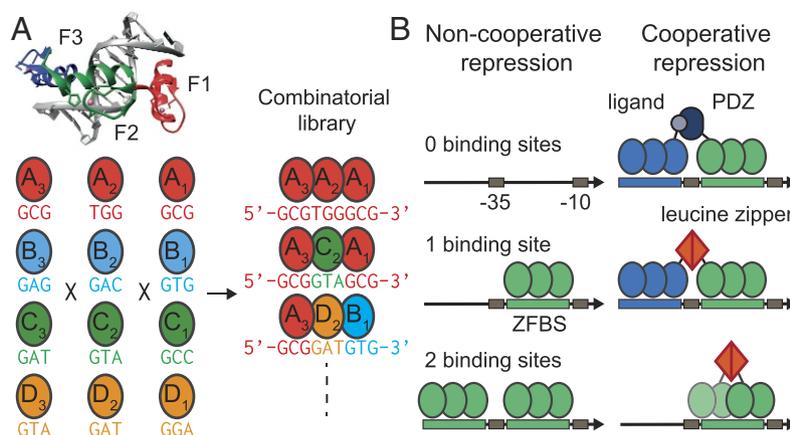


Fig. 2. ZF repressor and promoter design. (A) Our repressor design is based on the Zif268 protein, whose three ZFs (F1–F3) each recognizes a nucleotide triplet. We created a combinatorial library of repressors by shuffling individual ZFs, starting from four initial ZF proteins (here labeled with the codes AAA–DDD). (B) We designed a library of repressible promoters based on the λP_R promoter. To test the effectiveness of repression, we designed promoters containing single and dual sites with variable spacing, as well as engineering direct cooperativity between ZF proteins, which can be mediated by PDZ–ligand or LZ interactions.

the most effective position for transcriptional repression is the space between the -35 and -10 boxes (30); we thus generated a library with consensus ZF binding sites (ZFBSs) inserted into this location. Additionally, we built promoters with a second ZFBS upstream of the -35 box, allowing us to study the effect of multiple noncooperative and cooperative ZFBSs (Fig. 2B). The promoters drive expression of a deGFP reporter, a GFP protein previously optimized for cell-free translation (46). All constructs were built and tested by using linear DNA templates generated by PCR in concordance with recommended guidelines for cell-free expression (45).

Repression with Single and Multiple Binding Sites. We performed an in-depth characterization of 11 synthetic ZFs by assessing their repressive capacity in cell-free reactions and by measuring their respective dissociation constants (K_d) with MITOMI. We used MITOMI to measure the K_d values for each ZF against all possible target promoters. By localizing presynthesized histagged ZFs to the surface of each unit cell, we were able to measure the binding of DNA sequences spanning the promoter region, including the ZF binding site (Fig. 3A and *SI Appendix, Fig. S1B*). We obtained standard Gibbs free energies, $\Delta G = RT \ln(K_d)$, for each ZF–target promoter complex (Fig. 3B). A range of binding strengths was observed for the respective consensus ZF binding sequences, as well as low-affinity off-target binding. The CBD ZF was included as a negative control, as it does not bind to its own predicted binding site or any of the other targets.

To test whether the relative binding strength of each ZF related to functional gene repression, we implemented cell-free TX-TL reactions screening the same matrix of ZFs vs. promoters. Each microfluidic unit cell contained a linear template encoding the ZF to be tested and a second linear template encoding deGFP downstream of a promoter with a single ZF binding site (Fig. 3C). Binding of the expressed ZF to the target promoter would lead to down-regulation of deGFP expression. A common measure of repression performance is fold repression, or the ratio of unrepressed to repressed expression levels. Unrepressed measurements were obtained by coexpressing the target promoter template with the nonbinding ZF_{CBD} template to control for loading effects (47). Despite some off-target binding observed by MITOMI, functional repression of all ZF–target pairs was almost perfectly orthogonal (Fig. 3D), with one exception: the repression of promoter BDD by ZF_{ADD} . However the general trend of weak off-target affinities translated to no or minimal off-target repression, resulting in functional repression only for cognate pairs. Furthermore, on-target fold repression directly correlated with the measured MITOMI affinity values (Fig. 3E). Using two high-throughput microfluidic techniques, we were able to characterize the binding affinity,

repressive strength, and orthogonality of synthetic transcription factor–promoter pairs.

Promoters with a single ZF binding site achieved low to medium fold repression levels in the range of 1.5–7 (Fig. 4A). We tested whether placing an additional binding site upstream of the -35 box could further improve fold repression levels. While fold repression is a convenient measure used to describe the functionality of a given repressor–promoter pair, for applying these repressors in genetic networks, it is important to also consider basal promoter strength (unrepressed state) and leak (repressed state). These quantities are also shown in Fig. 4, where we observed that variation in binding-site sequence led to variations in basal promoter strength; this variation increased upon inclusion of the second binding site upstream of the -35 box. At the same time, the average leak from the repressed state decreased for the dual site library, resulting in higher fold-repression values. Overall, fold repression improved for almost all two-binding-site promoters, with the best promoters achieving a fold repression level of 7–10 (Fig. 4B). These results showed that good repression levels can be achieved by synthetic ZF repressors with either single or double binding-site promoters in a cell-free system.

Next, we characterized the effect of binding-site position on repression strength. We generated a library of promoters containing a single ZF binding site that was placed in various positions relative to the -35 box. The best fold repression was achieved by positioning binding sites directly proximal to the -35 box, in the range of -2 to $+4$ bps relative to the start and end of the -35 box, respectively. We also observed that repression was sensitive to single base-pair shifts in position. For instance, the site at the $+5$ position was effectively nonfunctional compared with repressing neighboring sites at $+4$ and $+6$, and the site at the -5 position exhibited significantly stronger repression than its neighbors at -4 and -6 . Based on the crystal-structure alignment of ZF and RNA polymerase (RNAP) bound to DNA containing the binding site at position $+5$, we note that it is possible for both proteins to bind simultaneously with minimal steric interference. To ascertain that the observed repression strengths were not due to changes in binding-site affinity of the ZF, as each binding site is located in a different sequence context, we measured the binding affinity of the ZF repressor to each promoter using MITOMI. The results showed only minor differences in affinity across all promoters, suggesting that the ZF repressor bound to these promoters with equal strength. Promoter repression thus appears to be primarily a function of the ability of the ZF to sterically hinder and compete with RNAP. These data are consistent with an occlusion mechanism whereby RNAP binding is competitively inhibited by ZF binding (29), and the effectiveness of the competition is dependent on the relative positions of ZF and RNAP on the promoter.

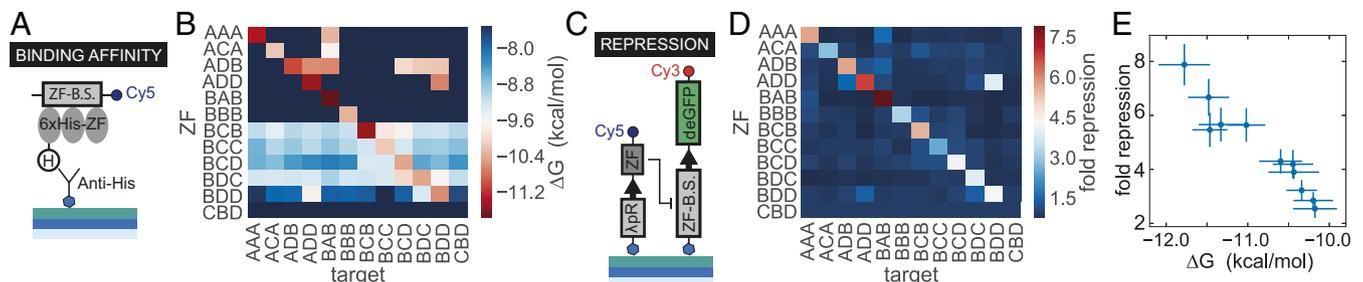


Fig. 3. ZF binding affinities, functional repression, and orthogonality. (A) Schematic depicting the MITOMI assay used to determine TF–DNA binding affinities. (B) Affinity orthogonality matrix of ΔG values for all ZFs vs. all possible DNA targets. (C) Schematic depicting the linear templates used to test functional repression in on-chip cell-free TX-TL reactions. (D) Fold repression orthogonality matrix for all ZFs vs. all possible targets. (E) Fold repression values vs. measured K_d for all ZF–promoter consensus pairs. The fold-repression data were collected from a single chip, and all values represent means \pm SD ($n = 5$). The error bars shown for the K_d values represent the 95% confidence interval for the fit to a single binding-site model.

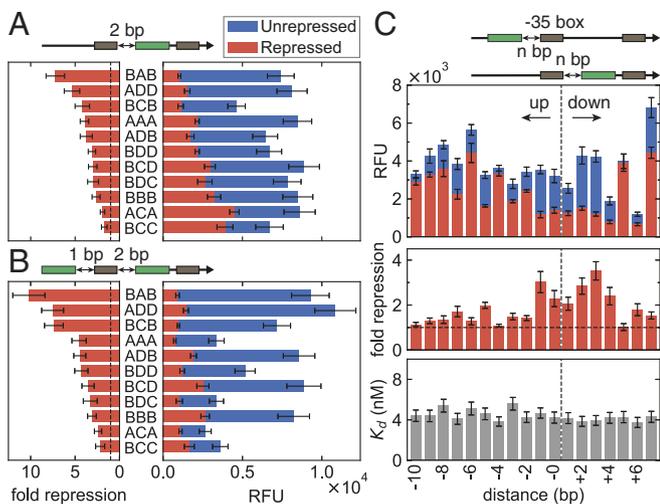


Fig. 4. Effect of binding-site number and position on repression. (A and B) Shown are endpoint unrepressed and repressed levels for the single-binding-site library (A) and the dual-site library (B). Within each library, the data are rank-ordered by fold repression values. Data were collected from two separate chips, and all values represent means \pm SD ($n = 10$). (C) A single BCB site was shifted up to 10 bp upstream and 7 bp downstream of the -35 box; position values are given by the number of nucleotides separating the 9-bp site from the -35 box. (C, Lower) The binding affinities of the ZF to its target site remain approximately constant, irrespective of target site position. Data from C, Upper were measured from a single chip, and all values represent means \pm SD ($n = 7$). The error bars shown for the K_d values represent the 95% confidence interval of the fit to a single binding-site model. RFU, relative fluorescence units.

Engineering Cooperativity. We showed that incorporating a second binding site can result in improved fold repression. However, engineering certain types of genetic circuits often requires an additional increase in the nonlinear response, as well as a decrease in the leak for a given promoter–TF pair. Nonlinearity can be increased by introducing cooperativity via protein–protein interactions. We implemented two different protein–protein domains previously demonstrated to successfully dimerize ZFs.

PDZ domains enable natural protein–protein interactions by binding specific C-terminal peptide sequences with micromolar affinity (26). We took advantage of this interaction to engineer cooperativity by linking ZF_{BCB} to a mammalian $\alpha 1$ -syntrophin PDZ domain and ZF_{ADD} to its corresponding cognate C-terminal peptide ligand (VKESLV). Furthermore, we linked ZF_{ADD} with a noncognate ligand (VKEAAA) to use as a noncooperative control.

The second type of interaction we explored was dimerization by linking ZF_{BCB} and ZF_{ADD} to GCN4 LZ domains. The GCN4 LZ has been used in a structure-based design to enable homodimerization of two-finger ZFs (42), and we thus also tested this existing structure. In both cases, a mutated LZ was used as a negative control.

Preliminary studies on a plate reader demonstrated that ZFs containing interaction domains exhibited significantly increased fold repression and decreased leak (Fig. 5 A and B). Whereas two noncooperative repressors gave a maximum fold repression of ~ 6 , this value was increased to ~ 30 for PDZ and ~ 16 for LZ-mediated cooperativity. Concurrently, leak values decreased fourfold from $\sim 4,000$ to $< 1,000$ relative fluorescence units. One critical parameter affecting PDZ cooperativity was the choice of linker, with an optimized glycine–serine linker vastly outperforming a rigid proline linker. The two-finger LZ transcriptional repressor also performed very well, achieving a fold repression ratio of ~ 28 .

To investigate cooperativity in more detail, we measured dose–response curves by titrating repressor DNA concentration. To keep a fixed load on the TX–TL machinery, the total ZF DNA concentration was kept constant by adding DNA coding for a nonbinding ZF control (ZF_{CBD}). Fig. 5C shows dose–response curves of $ZF_{BCB} - PDZ$ and $ZF_{ADD} - L$ separately, together with those for the cooperative pair, $ZF_{BCB} - PDZ + ZF_{ADD} - L$, and the noncooperative pair, $ZF_{BCB} - PDZ + ZF_{ADD} - NL$. An increase in the steepness of the dose–response curve was observed as we proceeded from a single ZF to two noncooperatively interacting ZFs, and finally to two cooperatively interacting ZFs. Similar results were obtained for the LZ designs (Fig. 5 D and E). The effect of cooperativity can be quantified by determining the sensitivity (*SI Appendix*, Fig. S7), which measures the steepness of the dose–response curve (48), as well as the effective Hill coefficient, which is obtained by fitting phenomenological Hill functions (*SI Appendix*, Fig. S8). The results of this analysis are shown in *SI Appendix*, Table S1. We observe that cooperativity increased sensitivity by nearly 50% with respect to the noncooperative repression, as well as slightly increasing the Hill coefficient.

We sought to understand this behavior quantitatively by developing a thermodynamic model that relates protein expression to the equilibrium occupancy of the promoter by RNAP (49). We extended the standard competitive model of repression to include a term for the interaction between repressor and RNAP, which is quantified by an effective interaction energy. As this energy tends to large positive values, DNA binding by either

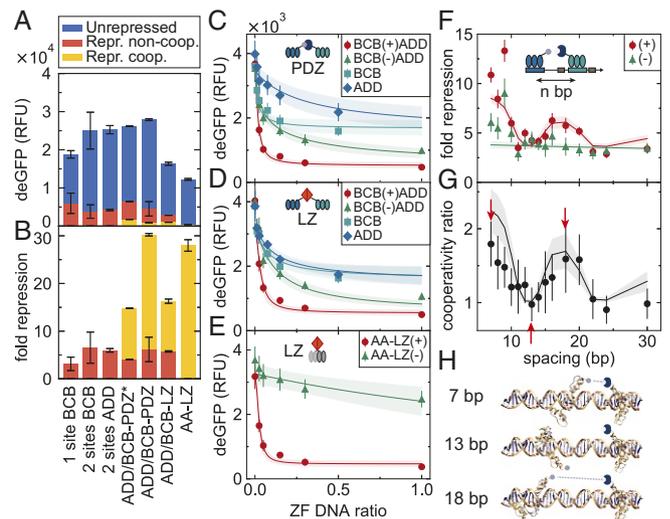


Fig. 5. Engineering cooperativity. (A and B) Comparison of unrepressed and repressed (Repr.) levels (A), as well as fold repression (B), for different cooperative (coop.) ZF designs. Three-finger ZFs dimerized by using either PDZ–ligand or GCN4 LZ domains are shown. ZFs were linked to interaction domains by using proline ($-PDZ^*$) or optimized glycine–serine linkers ($-PDZ$, $-LZ$). Additionally, two-finger ZFs were dimerized by using LZs (AA–LZ). Data in A and B were taken from plate-reader experiments; all values represent means \pm SD ($n = 3$). (C–E) The dose–response is shown for individual ZF repressors as well as noncooperative ($-$) and cooperative ($+$) pairs of ZFs, for the three-finger PDZ and two- and three-finger LZ designs. The maximum a posteriori values as well as 2-SD boundaries of the model predictions are represented as solid lines and shaded regions, respectively. Data in C–E were measured on a single chip; all values represent means \pm SD ($n = 12$). (F–H) Shifting the binding site upstream of the -35 box resulted in periodic modulation of the fold repression for the cooperative designs (F), as well as in the ratio between the cooperative and noncooperative fold repressions (G), likely due to periodic changes in relative ZF positioning (H). All data were collected from a single chip, and all values represent means \pm SD ($n = 9$). RFU, relative fluorescence units.

RNAP or the repressor is exclusive, and the model tends toward that of competitive inhibition. As the energy approaches zero, both RNAP and DNA can bind simultaneously, resulting in leaky expression at full repressor occupancy. This extension to the model was motivated by our results that a ZF with a fixed binding affinity represses with varying efficiency, depending on the position of the binding site; the changing RNAP–ZF interaction energy therefore provides a simple description of this effect. We fit the model to the dose–response curves using Markov chain Monte Carlo (MCMC) sampling (*SI Appendix, Fig. S9*), allowing us to consistently extract the posterior probability distributions of all parameters, which consist of fixed effective dissociation constants of each individual ZF, as well as the effective energies describing ZF–RNAP and ZF–ZF interactions. The fits are shown in Fig. 5 *C–E* as solid lines and shading, which represent the mean and 2-SD boundaries for model predictions, respectively. The values of all fitted parameters are given in *SI Appendix, Table S2*, and a full description of the model is given in *Materials and Methods*. We found physically sensible values for all our parameters; in particular, the cooperative interaction energies for PDZ–ligand (-2.1 ± 0.2 kcal/mol) and LZ (-1.8 ± 0.2 kcal/mol) were consistent with literature values for similar domains [~ -2 to -10 kcal/mol (50, 51)].

Since the location of the ZF binding site, and hence the relative positioning of ZF and RNAP, is an important determinant of repression efficiency, it is likely that the relative positioning of the ZF_{BCB} –PDZ and ZF_{ADD} –L binding sites would also determine their ability to interact and subsequently alter their repressive strength. Keeping the ZF_{BCB} –PDZ binding-site position fixed, we shifted the ZF_{ADD} –L binding site further and further upstream. If the two ZFs are positioned on the promoter such that the cooperative PDZ–ligand interaction is unfavorable, we would expect fold repression to be similar to that of the noncooperative ZFs. In other words, the ratio between the cooperative and the noncooperative fold repression, a quantity we call the cooperativity ratio, should go to unity when the PDZ–ligand interaction cannot occur.

We observed an effect due to this variation of spacing between the two binding sites (Fig. 5*F*), and this behavior corresponded to the relative orientation of the PDZ–ligand domains. As the binding site was shifted, ZF_{ADD} –L rotated around the DNA, modulating its alignment with ZF_{BCB} –PDZ. The cooperativity ratio fell to 1 when the interaction was unfavorably aligned, but increased again as the domains began to realign (Fig. 5*G*). The cartoon in Fig. 5*H* shows the predicted orientations of the two ZFs as the left-hand site is shifted. The ability of the ZFs to interact over distances of a few tens of base pairs is likely due to extension of the long, flexible glycine–serine linker used to join the ZF_{BCB} and the PDZ domain. It is unlikely that DNA bending plays a significant role at these distances, due to dsDNA’s much longer persistence length of ~ 150 bp.

We incorporated into our model a phenomenological exponential decay of interaction energies with distance, both between the two ZFs as well as between the ZF and the RNAP. Additionally, the ZF–ZF interaction energy was modulated by a periodic function at the frequency of the DNA helical pitch (10.5 bp per turn). Using previously inferred parameters for energies and K_D values from the dose–response measurements, we performed a fit to determine the decay constant and phase shift; the results are shown as solid lines and shading in Fig. 5 *F* and *G* and in *SI Appendix, Table S2*. Fitting a model with an explicit position dependence for the binding sites illustrates the importance of site positioning for functional repression. More generally, while simplistic, our model fits demonstrate that it is possible to understand cell-free gene expression in terms of thermodynamic occupancy.

Affinity Tuning. To test whether fold repression levels could be precisely and predictively tuned, we investigated the effect of varying binding-site affinity. To rationally tune binding-site affinity, we first generated quantitative PWMs for three ZFs: ZF_{BCB} , ZF_{AAA} , and ZF_{ADD} , covering the 9-bp core sequence plus three flanking bases on either side (Fig. 6*A* and *SI Appendix, Fig. S10 A* and *B*). The sequence logo determined for ZF_{AAA} is in concordance with the consensus sequence determined by bacterial one-hybrid and in vitro SELEX assays (52, 53). Based on our PWMs, we designed a library of promoters that included a single binding site at a fixed position between the -35 and -10 boxes, with single or double mutations within or outside the core binding sequence. As binding-site affinity decreased, we observed corresponding decreases in fold repression for all ZFs tested (Fig. 6*B* and *SI Appendix, Fig. S10C*). By converting our macroscopically measured ΔG values into microscopic interaction energies $\Delta\epsilon$, we found that the fold-repression data could be described by the same thermodynamic model presented in the previous section.

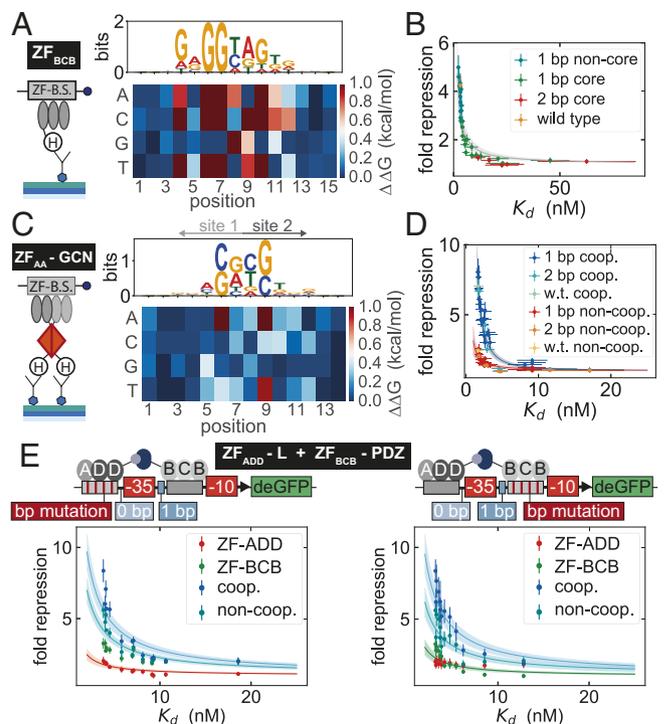


Fig. 6. Tuning repression by changing binding-site affinity. (A) Schematic of the MITOMI assay used for measuring the binding affinity of a ZF to a given DNA target. (A, Left) The sequence logo and PWM for ZF_{BCB} , where the core sequence is designated by positions 4–12. (B) The relationship between fold repression and K_d values for ZF_{BCB} . The fold repression data were collected from three separate chips, and all values represent means \pm SD ($n = 7$). (C) Schematic of the MITOMI surface used for measuring the binding affinity of the ZF_{AA} –GCN homodimer to a given DNA target. (C, Left) The sequence logo and PWM for ZF_{AA} –GCN, where the core sequence is designated by positions 3–12. (D) The relationship between fold repression and K_d values for both the cooperative (coop.) and noncooperative (non-coop.) variants of ZF_{AA} –GCN. The fold repression data were collected from a single chip, and all values represent means \pm SD ($n = 8$). (E) Fold repression vs. K_d values for the ZF_{ADD} –L– ZF_{BCB} –PDZ heterodimer pair. In E, Left, the K_d values refer to the K_d arising from the specific change made to the ADD binding site, whereas in E, Right, the K_d values are associated with the BCB binding site. The fold repression data were collected from a single chip, and all values represent means \pm SD ($n = 4$). In all cases, the error bars shown for the K_d values represent the 95% confidence interval for the fit to a single binding-site model; the solid lines are maximum a posteriori values from thermodynamic model fits; and the shaded region represents a 2-SD boundary.

Mutating either a single base outside the core site, or one core position of low information content (high entropy), enabled fine tuning of fold repression, whereas a single mutation in the core site of high information content strongly decreased fold repression. Two core mutations decreased fold repression to baseline levels. Fold repression was therefore precisely tunable over the entire dynamic range by modulating binding-site affinity, and the affinity changes required to achieve tuning were relatively small. Affinity changes of ~ 0.5 to 1 kcal/mol were sufficient to cover the entire dynamic range for each ZF repressor tested. The results are in line with previous findings that promoter tuning in *S. cerevisiae* can be accomplished by relatively subtle affinity changes in a single binding site created by mutations in flanking or single-core site mutations of high entropy (54). They also correspond to recent results obtained in *E. coli* (55).

Given that a single ZF binding site could be mutated to yield varying levels of repression, we investigated whether the same tuning could be applied to cooperative ZFs. We measured the binding affinity of the $ZF_{AA} - GCN$ homodimer vs. a library of DNA targets that consisted of all single-point mutations for the 10-bp core binding sequence plus two flanking bases on either side. The resulting sequence logo and PWM reveal the symmetric binding profile of the homodimer (Fig. 6C). Mutating a single binding site within the -35 and -10 boxes led to a change in repression levels that reflected the measured K_d s for both the cooperative and noncooperative $ZF_{AA} - GCN$ variants (Fig. 6D). As the two 6-bp binding sequences overlap, mutating a single base within the core site leads to a finer tuning of fold repression in comparison with the three-finger ZFs. Furthermore, we extended binding-site tuning to the $ZF_{ADD} - L - ZF_{BCB} - PDZ$ heterodimer pair, taking advantage of the PWMs generated for ZF_{BCB} and ZF_{ADD} . Implementing a subset of mutations to each ZF binding site yielded a range of fold repression values not only for the single ZF but also for the cooperative and noncooperative ZF pairs (Fig. 6E). As the affinity of one ZF is reduced, we see that the fold repression observed for the cooperative and noncooperative cases tends to the fold repression measured for the second ZF, whose binding site remains constant.

Logic-Gate Construction. Having established a well-characterized resource of transcriptional repressors and promoters, we applied them to designing logic gates. By combining two cooperative ZF repressors on a single promoter, we were able to create NAND gates, which are of particular interest, as they are functionally complete. An effective NAND gate should have low output only when both inputs are present (Fig. 7A). We therefore placed the binding site for a strongly binding ZF (ZF_{BCB}) 2 bp upstream of the -35 box and second binding site for different ZFs between the -35 and -10 boxes. ZF_{BCB} cannot strongly repress by itself at the -2 position, and the second ZF should also not strongly repress on its own. Only when both ZFs are bound to the promoter should they strongly repress, which can be achieved by including a cooperative interaction between the two ZFs. Using this general design, we tested NAND gates for $ZF_{BCB} - PDZ$ in combination with the remaining ZFs (Fig. 7B). As expected, NAND gate performance improved as the affinity of the $ZF_{XXX} - L$ decreased. For instance, the combination of $ZF_{BCB} - PDZ$ and $ZF_{BDD} - L$ gave rise to a functional NAND gate, whereas a combination with $ZF_{AAA} - L$ did not, due to the high affinity of $ZF_{AAA} - L$, which led to functional repression even when only $ZF_{AAA} - L$ was present.

Since we showed that binding affinity could be precisely tuned (Fig. 6), we tested whether we could improve our nonfunctional NAND gates. Based on the PWM measured for ZF_{AAA} , we mutated the $ZF_{AAA} - L$ binding-site sequence in the NAND gate promoter and showed that we could achieve tuning in this con-

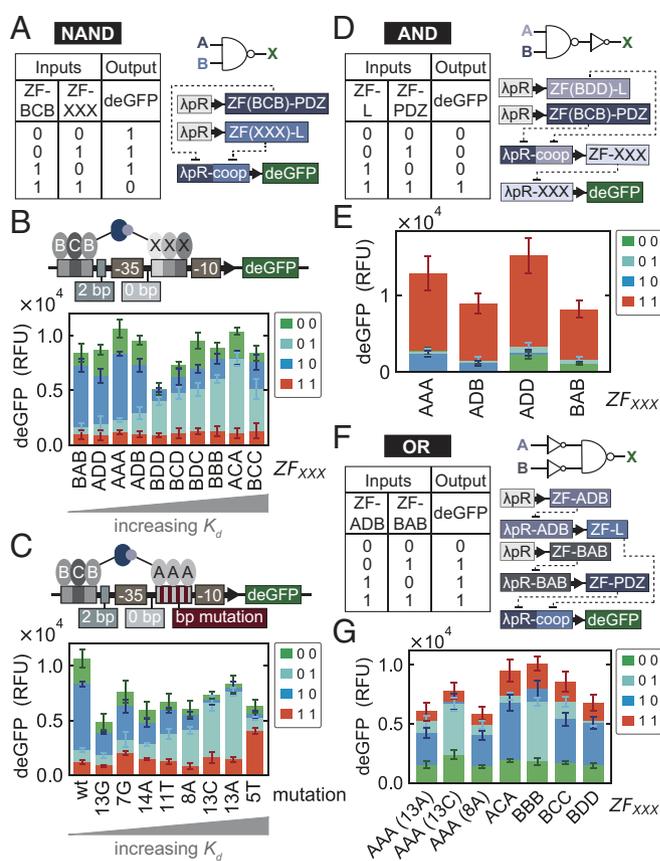


Fig. 7. Logic gates. (A) Truth table, logic-gate symbol, and biological network design for constructing a NAND logic gate. (B) NAND gate design combining $ZF_{BCB} - PDZ$ with all other $ZF_{XXX} - L$ values. (B, Lower) The output for each NAND gate tested. (C) The same NAND gate design as in A except that only $ZF_{BCB} - PDZ$ and $ZF_{AAA} - L$ are used as inputs and the $ZF_{AAA} - L$ binding-site affinity is rationally adjusted to yield a functional NAND gate. The deGFP outputs for all NAND gates were measured from a single chip, where all values presented correspond to the mean \pm SD ($n = 5$). (D) Truth table, logic gate symbols, and biological network design for constructing an AND logic gate. (E) The output for each AND gate tested. All output values were measured from a single chip and represent the mean \pm SD ($n = 8$). (F) Truth table, logic gate symbols, and biological network design for constructing an OR logic gate. (G) The output for each OR gate tested. When $ZF_{AAA} - L$ is used as part of the NAND gate, the mutation of the binding site is indicated in parentheses. All output values were measured from a single chip and represent the mean \pm SD ($n = 6$). RFU, relative fluorescence units.

text as well (*SI Appendix, Fig. S10D*). We then investigated the effect of tuning the $ZF_{AAA} - L$ binding site for all possible input combinations and showed that the NAND gate improved as we weakened $ZF_{AAA} - L$ binding affinity (Fig. 7C). Mutations +1C and +1A gave rise to functional NAND gates. Decreasing the binding-site affinity increased the output when only $ZF_{AAA} - L$ was present; however, when the mutation resulted in a $\Delta\Delta G$ of greater than ~ 0.5 kcal/mol ($\Delta\Delta T$), the cooperative binding output also suffered. Our synthetic ZF repressors can thus be used to build functional NAND gates, which can additionally be rationally optimized and precisely tuned by modifying binding-site affinities.

As a final example, we generated compound logic gates by combining NAND and NOT logic gates as linear cascades to create AND and OR gates. We created an AND gate by appending a NOT gate to the output of a NAND gate (Fig. 7D). Specifically, we combined the $ZF_{BDD} - L - ZF_{BCB} - PDZ$ NAND gate with four different ZFs. Each AND gate was tested and yielded

the expected outputs (Fig. 7E). We then generated OR logic gates by prepending two NOT gates in front of different NAND gates to invert the inputs (Fig. 7F). We used ZF_{ADB} and ZF_{BAB} as the two NOT gate inverters and a set of NAND gates, all of which gave rise to functional OR gates (Fig. 7G).

Discussion

GRNs are of central importance in both native and engineered systems. They integrate, compute, and transduce input signals, leading to specific changes in gene expression. Many components contribute to the function of GRNs, and transcription factors and their interaction with promoters are core players. Due to the complexity of even a single transcription factor–promoter interaction, it has proven difficult to quantitatively study these systems in vitro or in vivo. Although the development of new technologies is steadily enabling progress in this area, our understanding of GRNs remains limited, as exemplified by our inability to predict in vivo gene-expression levels in essentially any organism and the difficulty associated with de novo engineering of GRNs. Although methods exist for high-throughput in vitro characterization of transcription-factor-binding specificities (34, 56–58), and medium- to high-throughput approaches are used to understand gene regulation in vivo (33, 54, 55, 59), both approaches have limitations. Both an advantage and disadvantage of in vitro methods is that they generally include only the smallest number of components necessary, i.e., a transcription factor, dsDNA target, and a defined buffer solution. In vivo methods are, on the other hand, convoluted by cellular complexity. Furthermore, generating and analyzing defined libraries in vivo remains labor-intensive and difficult. Here, we explored the use of a cell-free TX-TL system to build and characterize GRNs in an environment that bridges the gap between in vitro and in vivo methods. This cell-free approach also has the advantage of allowing complex assays to be performed in high-throughput, in a well-controlled and accessible environment. As a consequence, the ability to study functional transcriptional regulation in an in vitro system has allowed us to delve into much greater depth than comparable in vivo methods have been able to achieve (54, 60, 61).

We chose to build GRNs from the bottom up using ZF transcription factors for several reasons. First, in regard to GRN engineering, researchers have long been hampered by the relatively small number and poor characterization of available transcriptional regulators. Khalil et al. (26) have engineered ZF regulators, showing that they are viable tunable transcriptional regulators in vivo. We built on this concept, generating additional ZF regulators and interaction domains. More importantly, we quantified the binding-energy landscapes of several synthetic ZF regulators and were able to show that repression can be precisely tuned with small changes in affinity. These small changes were achieved by mutating the flanking bases lying outside of the consensus core sequence or by mutating one consensus core base of low information content. Hitherto, only coarse tuning has been accomplished through varying the number of consensus-sequence binding sites, leading to rather large differences in output (26, 27). The ability to predictively and precisely tune expression levels as demonstrated here is important in engineered GRNs, where individual nodes of the network need to be matched in expression levels. For example, we show here that the ability to precisely adjust individual binding-site affinities is crucially important for optimizing logic gate function.

With the advent of TALEs and dCas9, ZFs might be considered outdated technology, but there are a number of reasons why ZF TFs remain an appealing tool for GRN engineering. ZFs have several advantages, such as small size, relatively easy gene synthesis, and good expressability. The biggest advantage of dCas9 and TALEs is their programmability, allowing them

to be precisely targeted to any DNA sequence. Conversely, for ZFs, it remains relatively difficult to rationally design a particular binding-site preference. For genome editing and in vivo targeting approaches, in which the target sequence is defined and immutable, programmability is crucial. In the context of bottom-up GRN design, this ability becomes less important as target sequences can be easily adjusted to a particular TF specificity. We argue that it is actually more important to be in possession of a well-characterized TF binding-energy landscape that can be obtained for ZF TFs using current methods (28).

A second argument in support of using ZF transcription factors over TALEs and dCas9 is the simple but important fact that ZFs are native transcriptional regulators and the most abundant class of transcriptional regulators in vivo. Cas9, to the best of our knowledge, has not been shown to be involved in gene regulation in native systems, while TALEs are injected into plant host cells to modulate gene expression by pathogenic bacteria (62). If cell-free approaches are to be used to understand the function of native systems, it is important to build GRNs with native transcription factors. For example, the protein–DNA interaction kinetics are very different in that dCas9 (63) and TALE (64) tend to have very slow DNA dissociation rates, while native transcriptional regulators have fast dissociation rates (65), which may make engineering dynamic GRNs using TALEs and dCas9 difficult.

To improve fold repression and to add more control over the system, we engineered cooperative binding into our ZF TFs by including PDZ or LZ protein–protein interaction domains. These interactions improved repression from ~ 10 -fold to up to ~ 30 -fold and were functional for both two- and three-finger ZFs. We showed that the relative placement of binding sites for two cooperative TFs is a major determinant of interaction capacity and, consequently, repression strength. Repression was achieved when the TFs were located on the same face of the DNA, and repression strength followed the helical twist of DNA. Cooperative interactions consequently allowed us to engineer functionally complete NAND gates. In all cases, we were able to explain our data with thermodynamic models. Combining these models with binding-energy landscapes thus provides a viable and useful approach to rationally engineer GRNs.

One outstanding problem encountered during this study is the issue of composability. Although transcription-factor binding sites were only introduced in regions outside the -10 and -35 boxes of the original λP_R promoter, many of the synthetic promoters had considerably different baseline (nonrepressed) expression levels. In the future, it will clearly be important to better understand and predict basal promoter strength from the underlying sequence, which would lead to models that allow introduction of transcription-factor binding sites without affecting basal promoter output. Here, we have seen that basal promoter strength itself can be finely tuned over a relatively large range of expression levels (Fig. 1). It should therefore be possible to adjust promoter strength as desired: We demonstrate a basic example of this idea by tuning the basal expression level of a repressible promoter (*SI Appendix, Fig. S11*). Ultimately, understanding the outcome of multiple base changes in close context with each other remains a complex issue. Evaluating a greater number of sequences and systematically addressing all factors affecting transcription efficiency, similar to the approach taken by Cambray et al. (66) toward translation, could lead to an improved understanding of promoter-sequence design principles.

To characterize and measure our synthetic ZF transcription factors and promoters in detail, we repurposed a high-throughput microfluidic device that allowed us to measure 768 cell-free reactions in parallel. Eliminating cloning and transformation steps by relying on PCR-based assembly strategies allowed us to measure a large number of defined

transcription-factor and promoter variants. Over 13,000 on-chip cell-free TX-TL reactions were performed, encompassing replicates for ~2,000 unique reactions. We furthermore took >8,000 MITOMI measurements to provide binding-energy landscapes for four synthetic ZF transcription factors. Together, these technologies allowed us to establish a quantitative and in-depth dataset and insights into transcriptional regulation that should be of general interest. The approach taken here nonetheless does not per se require these state-of-the-art technologies and is easily transferable to standard laboratory equipment. Cell-free lysate can now be easily and cheaply generated, yielding sufficient material so that medium-scale screens in 384-well plates are feasible (44). Commercial liquid-handling equipment can also be used to scale up throughput. Binding-energy landscapes can be generated by many approaches, including PBMs (56), MITOMI (67), SELEX-seq (57), and HiPFA (58). While our binding-energy landscapes are based on direct affinity measurements, it may be sufficient to use PWMs from indirect measurements as found in other high-throughput techniques.

Rapid progress is being made in the development and application of cell-free synthetic biology. Cell-free systems are being used to tackle fundamental problems in molecular engineering and are being applied to molecular diagnostics (68), therapeutics (69), and synthesis (70) and are even being used for educational purposes (71). Cell-free systems are an appealing alternative to cellular systems, as they eliminate many of the complexities associated with working with cells. Cell-free systems are also a rapid prototyping platform for engineering molecular systems destined to be applied in cellular hosts (4). As engineered systems become more complex, it will become increasingly important that a large number of standardized characterized components become available. It will be equally important to develop a comprehensive mechanistic understanding of these components and systems to allow parts to be standardized and rationally assembled with-

out requiring extensive trial-and-error cycles or large screens, which may not be feasible for large systems. As work progresses on cellular subsystems such as gene regulation, DNA replication, ribosome biogenesis, metabolic networks, and membrane and protein superstructures, it will be intriguing to contemplate whether it may be possible to integrate these individual systems to create a synthetic cell or cell-like mimic. Work in this area will not only provide tools and methods aiding the engineering of synthetic systems, but is likely to provide insights into the function of native systems as well. Before being used as tools for protein synthesis and synthetic biology, cell-free systems have already had a rich history in deciphering fundamental aspects of biochemistry, including DNA replication (72) and the genetic code (73). It is likely that they will continue to provide fundamental insights into complex systems such as transcriptional regulation.

Materials and Methods

Full details are given in *SI Appendix*. High-throughput cell-free experiments were conducted on a PDMS microfluidic device. DNA templates were assembled by PCR from plasmids and dsDNA gene fragments and spotted onto a glass slide by using a robotic microarrayer. *E. coli* (BL21 Rosetta) cell-free extract, which contains native enzymes, substrates, and an energy-regeneration mechanism, was used in all experiments. Binding-energy landscapes were obtained by microfluidic MITOMI measurements. Our thermodynamic model was fit to experimental data by using a MCMC method to determine the posterior probability distributions of model parameters. All supporting data and code are available on GitHub (74).

ACKNOWLEDGMENTS. We thank S. Clamons and M. Yun (Murray laboratory, Caltech) for providing the cell-free TX-TL extract; S. Clamons and R. Murray for helpful discussions; and M. Kabani, E. Pankevich, and S. Bassler for their experimental contributions to this project. This work was supported by Human Frontier Science Program Grant RGP0032/2015; the European Research Council under the European Union's Horizon 2020 research and innovation program Grant 723106; and the École Polytechnique Fédérale de Lausanne.

- Noireaux V, Bar-Ziv R, Libchaber A (2003) Principles of cell-free genetic circuit assembly. *Proc Natl Acad Sci USA* 100:12672–12677.
- Shin J, Noireaux V (2012) An *E. coli* cell-free expression toolbox: Application to synthetic gene circuits and artificial cells. *ACS Synth Biol* 1:29–41.
- Niederholtmeyer H, Stepanova V, Maerkl SJ (2013) Implementation of cell-free biological networks at steady state. *Proc Natl Acad Sci USA* 110:15985–15990.
- Niederholtmeyer H, et al. (2015) Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife* 4:e09771.
- Karzbrun E, Tayar AM, Noireaux V, Bar-Ziv RH (2014) Programmable on-chip DNA compartments as artificial cells. *Science* 345:829–832.
- Chappell J, Jensen K, Freemont PS (2013) Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Res* 41:3471–3481.
- Schwille P, et al. (2018) MaxSynBio: Avenues towards creating cells from the bottom up. *Angew Chem Int Ed* 57:13382–13392.
- Forster AC, Church GM (2006) Towards synthesis of a minimal cell. *Mol Syst Biol* 2:1–10.
- van Nies P, et al. (2018) Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. *Nat Commun* 9:1583.
- Otrin L, et al. (2017) Toward artificial mitochondrion: Mimicking oxidative phosphorylation in polymer and hybrid membranes. *Nano Lett* 17:6816–6821.
- Jewett MC, Fritz BR, Timmerman LE, Church GM (2013) In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Mol Syst Biol* 9:678.
- Bhattacharya A, Brea RJ, Devaraj NK (2017) De novo vesicle formation and growth: An integrative approach to artificial cells. *Chem Sci* 8:7912–7922.
- Furusato T, et al. (2018) De novo synthesis of basal bacterial cell division proteins FtsZ, FtsA, and ZipA inside giant vesicles. *ACS Synth Biol* 7:953–961.
- Brophy JAN, Voigt CA (2014) Principles of genetic circuit design. *Nat Methods* 11:508–520.
- Nielsen AlecAK, et al. (2016) Genetic circuit design automation. *Science* 352:aac7341.
- Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403:339–342.
- Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R (2005) A synthetic multicellular system for programmed pattern formation. *Nature* 434:1130–1134.
- Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403:335–338.
- Stanton BC, et al. (2014) Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat Chem Biol* 10:99–105.
- Vaquerez JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: Function, expression and evolution. *Nat Rev Genet* 10:252–263.
- Beerli RR, Barbas CF III (2002) Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol* 20:135.
- Tebas P, et al. (2014) Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *The New Engl J Med* 370:901–910.
- Najafabadi HS, et al. (2015) C2h2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat Biotechnol* 33:555–562.
- Fu F, Voytas DF (2013) Zinc finger database (ZiFDB) v2.0: A comprehensive database of c2h2 zinc fingers and engineered zinc finger arrays. *Nucleic Acids Res* 41:D452–D455.
- Moore R, Chandras A, Bleris L (2014) Transcription activator-like effectors: A toolkit for synthetic biology. *ACS Synth Biol* 3:708–716.
- Khalil AS, et al. (2012) A synthetic biology framework for programming eukaryotic transcription functions. *Cell* 150:647–658.
- Lohmueller JJ, Armel TZ, Silver PA (2012) A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic Acids Res* 40:5180–5187.
- Blackburn MC, Petrova E, Correia BE, Maerkl SJ (2015) Integrating gene synthesis and microfluidic protein analysis for rapid protein engineering. *Nucleic Acids Res* 44:e68.
- Ptashne M, et al. (1976) Autoregulation and function of a repressor in bacteriophage lambda. *Science* 194:156–161.
- Cox RS III, Surette MG, Elowitz MB (2007) Programming gene expression with combinatorial promoters. *Mol Syst Biol* 3:1–11.
- Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I-I-2 regulatory elements. *Nucleic Acids Res* 25:1203–1210.
- Lanzer M, Bujard H (1988) Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci USA* 85:8973–8977.
- Sharon E, et al. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat Biotechnol* 30:521–530.
- Maerkl SJ, Quake SR (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science* 315:233–237.
- Moore SJ, et al. (2018) Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. *Proc Natl Acad Sci USA* 115:E4340–E4349.
- García-Cordero JL, Maerkl SJ (2016) Mechanically induced trapping of molecular interactions and its applications. *J Lab Automation* 21:356–367.

37. Kinney JB, Murugan A, Callan CG, Jr, Cox EC (2010) Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc Natl Acad Sci USA* 107:9158–9163.
38. Kamiuchi T, et al. (1998) Artificial nine zinc-finger peptide with 30 base pair binding sites. *Biochemistry* 37:13827–13834.
39. Kim J-S, Pabo CO (1998) Getting a handhold on DNA: Design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc Natl Acad Sci USA* 95:2812–2817.
40. Moore M, Klug A, Choo Y (2001) Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc Natl Acad Sci USA* 98:1437–1441.
41. Pomerantz JL, Wolfe SA, Pabo CO (1998) Structure-based design of a dimeric zinc finger protein. *Biochemistry* 37:965–970.
42. Wolfe SA, Ramm EI, Pabo CO (2000) Combining structure-based design with phage display to create new Cys2His2 zinc finger dimers. *Structure* 8:739–750.
43. Wolfe SA, Grant RA, Pabo CO (2003) Structure of a designed dimeric zinc finger protein bound to DNA. *Biochemistry* 42:13401–13409.
44. Sun ZZ, et al. (2013) Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J Visualized Exp* 79:e50762.
45. Sun ZZ, Yeung E, Hayes CA, Noireaux V, Murray RM (2014) Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synth Biol* 3:387–397.
46. Shin J, Noireaux V (2010) Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J Biol Eng* 4:8.
47. Siegal-Gaskins D, Tuza ZA, Kim J, Noireaux V, Murray RM (2014) Gene circuit performance characterization and resource usage in a cell-free 'breadboard'. *ACS Synth Biol* 3:416–425.
48. Bintu L, et al. (2005b) Transcriptional regulation by the numbers: Applications. *Curr Opin Genet Development* 15:125–135.
49. Bintu L, et al. (2005) Transcriptional regulation by the numbers: Models. *Curr Opin Genet Development* 15:116–124.
50. Saro D, et al. (2007) A thermodynamic ligand binding study of the third PDZ domain (PDZ3) from the mammalian neuronal protein PSD-95. *Biochemistry* 46:6340–6352.
51. Jana NK, Deb S, Bhattacharyya B, Mandal NC, Roy S (2000) A study of energetics of cooperative interaction using a mutant lambda-repressor. *Protein Eng* 13:629–633.
52. Meng X, Brodsky MH, Wolfe SA (2005) A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat Biotechnol* 23:988–994.
53. Wolfe SA, Greisman HA, Ramm EI, Pabo CO (1999) Analysis of zinc fingers optimized via phage display: Evaluating the utility of a recognition code. *J Mol Biol* 285:1917–1934.
54. Rajkumar AS, Dénervaud N, Maerkl SJ (2013) Mapping the fine structure of a eukaryotic promoter input-output function. *Nat Genet* 45:1207–1215.
55. Barnes SL, Belliveau NM, Ireland WT, Kinney JB, Phillips R (2018) Mapping DNA sequence to transcription factor binding energy in vivo. *PLoS Comput Biol* 15:e1006226.
56. Bulyk ML, Huang X, Choo Y, Church GM (2001) Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc Natl Acad Sci USA* 98:7158–7163.
57. Zhao Y, Granás D, Stormo GD (2009) Inferring binding energies from selected binding sites. *PLoS Comput Biol* 5:e1000590.
58. Jung C, et al. (2018) True equilibrium measurement of transcription factor-DNA binding affinities using automated polarization microscopy. *Nat Commun* 9:1605.
59. Mogno I, Kwasniewski JC, Cohen BA (2013) Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. *Genome Res* 23:1908–1916.
60. Amit R, Garcia HG, Phillips R, Fraser SE (2011) Building enhancers from the ground up: A synthetic biology approach. *Cell* 146:105–118.
61. Garcia HG, Phillips R (2011) Quantitative dissection of the simple repression input-output function. *Proc Natl Acad Sci USA* 108:12173–12178.
62. Boch Jens, et al. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326:1509–1512.
63. Boyle EA, et al. (2017) High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc Natl Acad Sci USA* 114:5461–5466.
64. Cuculis L, Abil Z, Zho H, Schroeder CM (2016) TALE proteins search DNA using a rotationally decoupled mechanism. *Nat Chem Biol* 12:831–837.
65. Geertz M, Shore D, Maerkl SJ (2012) Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc Natl Acad Sci USA* 109:16540–16545.
66. Cambray G, Guimaraes JC, Arkin AP (2018) Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*. *Nat Biotechnol* 36:1005–1015.
67. Maerkl SJ, Quake SR (2009) Experimental determination of the evolvability of a transcription factor. *Proc Natl Acad Sci USA* 106:18650–18655.
68. Keith P, et al. (2016) Rapid, low-cost detection of Zika virus using programmable biomolecular components. *Cell* 165:1255–1266.
69. Keith P, et al. (2016b) Portable, on-demand biomolecular manufacturing. *Cell* 167:248–259.e12.
70. Goering AW, et al. (2016) In vitro reconstruction of nonribosomal peptide biosynthesis directly from DNA using cell-free protein synthesis. *ACS Synth Biol* 6:39–44.
71. Stark JC, et al. (2018) BioBits bright: A fluorescent synthetic biology education kit. *Sci Adv* 4:eaat5107.
72. Fuller RS, Kornberg A (1983) Purified dnaA protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication. *Proc Natl Acad Sci USA* 80:5817–5821.
73. Nirenberg MW, Heinrich Matthaei J (1961) The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci USA* 47:1588–1602.
74. Swank Z, Laohakunakorn N, Maerkl SJ (2019) Data and code from 2019.Swank-analysis GitHub. Available at <https://github.com/lbnc-epfl/2019.Swank-analysis>. Deposited December 14, 2018.

1

2 **Supplementary Information for**

3 **Cell-free gene regulatory network engineering with synthetic transcription factors**

4 **Zoe Swank, Nananai Laohakunakorn, Sebastian J. Maerkl**

5 **Corresponding Author Name.**

6 **E-mail: sebastian.maerkl@epfl.ch**

7 **This PDF file includes:**

8 Supplementary text

9 Figs. S1 to S11

10 References for SI reference citations

11 Supporting Information Text

12 Supplementary Methods

13 **Microfluidic chip fabrication.** The molds for each device layer were fabricated using standard photolithography. For the control
14 layer, a silicon wafer was primed in an oxygen plasma processor for 7 minutes (TePla 300) and SU-8 photoresist (GM 1070,
15 Gersteltec Sarl) was spin coated onto the wafer yielding a height of 30 μm . Following a soft bake the wafer was exposed to
16 light (365 nm illumination, 20 mW/cm^2 light intensity) using a chrome mask for 10 s on a Süss MJB4 mask aligner. After a
17 post exposure bake the wafer was developed with PGMEA (propylene glycol monomethyl ether acetate) to remove unexposed
18 SU-8 and a hard bake was performed to remove unwanted cracks in the SU-8 structures. For the flow layer, a silicon wafer was
19 treated with HMDS (hexamethyldisilazane) vapor in a YesIII primer oven and AZ 9260 photoresist (Microchemicals GmbH)
20 was spin coated on the wafer to a height of 14 μm . After baking and a one hour relaxation period, the coated wafer was
21 illuminated with a broadband light using an MABA6 mask aligner with a total dose of 660 mJ/cm^2 split into two exposures of
22 18 s with a 10 s wait period between (20 mW/cm^2 light intensity). The wafer was developed with AZ 400K developer and
23 baked at 175 $^\circ\text{C}$ for two hours to re-flow and anneal the AZ structures.

24 Each of the wafers was subsequently treated with TMCS (trimethylchlorosilane) and coated with PDMS (polydimethylsiloxane,
25 Sylgard 184, Dow Corning). For the control layer PDMS with an elastomer to crosslinker ratio of 5:1 was prepared and poured
26 over the wafer to yield a height of ~ 0.5 cm, while for the flow layer, PDMS with a 20:1 elastomer to crosslinker ratio was spin
27 coated at 1800 rpm to yield a height of ~ 50 μm . Both PDMS coated wafers were placed in the oven at 80 $^\circ\text{C}$ for 20 minutes.
28 After an initial curing the control layers were cut out and the holes for each control line were punched using a 900 mm pin.
29 Each control layer was then aligned by hand on top of a flow layer using a Nikon stereo microscope and the aligned devices
30 were placed in the oven at 80 $^\circ\text{C}$ for 90 minutes, allowing the two layers to bond together.

31 **Preparation of cell-free extract.** BL21 Rosetta *E. coli* cell-free extract was a gift from the Murray lab (Caltech) and prepared
32 according to a published protocol (1). As described in the cited protocol, the addition of purified gamS protein is added to the
33 final reaction mixture to prevent the degradation of linear DNA templates by nucleases.

34 **Preparation of DNA templates.** All promoters were generated by PCR from a pBEST-OR2-OR1-Pr-UTR1-deGFP-T500 plasmid
35 (2) (Addgene #40019). Oligos containing the promoter variants were ordered from IDT. Assembly PCR was used to synthesize
36 the complete template including a given promoter variant and the downstream deGFP gene. In the first step two linear DNA
37 products are amplified with PCR where each product contains a 20 bp overlap region and one of the products contains the
38 change to the base promoter. The overlap region was generally between the -10 box and the RBS. In the second step the two
39 DNA products were annealed together and the entire template was amplified using global primers which were labeled with
40 a fluorescent molecule (Cy3) on the 5' end and biotin on the 3' end. ZF repressor templates were generated by PCR from
41 gBlock gene fragments (IDT) that included the λP_R promoter. Global primers labeled with Cy5 at the 5' end and biotin at
42 the 3' end were used to synthesize the ZF templates. To amplify ZF gene templates with cognate or non-cognate ligands a
43 separate 3' primer was used to incorporate the ligand. However, to produce ZF - PDZ or LZ gene templates, an assembly PCR
44 was performed to link the sequence encoding the ZF with that encoding the PDZ or LZ domain. The two-finger homodimer
45 AA - GCN was amplified from a gBlock gene fragment. All linear DNA templates contained 5' protection sequences of 250 bp
46 (promoters) or 130 bp (zinc fingers); this slows degradation of DNA in the cell-free extract (3). Complete sequences are given
47 in Tables S3-S6. Given that each linear template is tagged by a fluorescent molecule at the 5' end and a biotin molecule at the
48 3' end we can assume that all DNA which is immobilized at the surface of a given unit cell and we visualize via fluorescence is
49 therefore full length.

50 **Setting up high-throughput cell-free experiments.** The concentration of linear template DNA was quantified by absorbance
51 (NanoDrop, ThermoFisher) and PCR products were directly added to a 384 microwell plate along with short biotinylated
52 ssDNA oligos for spotting. For characterization of the λP_R promoter library, a single linear template was spotted. Simple
53 repression assays involved spotting both a ZF template along with the deGFP target template and dose response experiments
54 for cooperative ZFs required spotting up to 5 linear templates per unit cell. Taking the PDZ-ligand heterodimer case as an
55 example, both $ZF_{BCB} - PDZ$ and $ZF_{CBD} - PDZ$ as well as $ZF_{ADD} - L$ and $ZF_{CBD} - NL$ templates were combined in
56 varying ratios, in addition to the deGFP template. All ZF templates and the target template were spotted for the NAND logic
57 gate experiments, however for the AND and OR gates the first ZF template was pre-expressed in a separate reaction while all
58 other down-stream components of the logic gate were spotted. Linear DNA templates were spotted with a QArray2 (Genetix)
59 microarrayer using an MP3 pin (Arrayit) onto epoxy coated glass slides. The DNA was diluted in 4% BSA in MilliQ H_2O to
60 prevent the DNA from binding the epoxy functional groups and to aid visualization of the drops for alignment with the PDMS
61 chip. Once the PDMS chip was aligned on top of the DNA microarray the chip was bonded to the glass by incubating the
62 assembled device for 1 hour at 80 $^\circ\text{C}$. Chips were then stored overnight at 45 $^\circ\text{C}$ and used the following day or up to 2-5 days
63 later.

64 To prime the chip, control lines were filled with PBS and pressurized at up to 138 kPa. While isolating the spotted DNA
65 with the neck valve closed, the lower half of the unit cell was patterned with BSA-biotin and neutravidin (Thermo Fisher) (4).
66 First, BSA-biotin (2 mg/mL) was flowed for 15 minutes, then neutravidin (1 mg/mL) for 15 minutes, after which the button
67 was closed and BSA-biotin was flowed again for 10 minutes. Between each of these three steps PBS was flowed for 5 minutes
68 to wash away any unbound molecules. The pressure applied to the flow lines during this process was ~ 24 kPa. This surface

69 chemistry resulted in a circular area coated with neutravidin, whereas the remaining surface of the unit cell was passivated with
70 BSA-biotin. Afterwards the spotted DNA was solubilized with PBS by closing the outlet and opening the neck valve while
71 PBS was flowed into the device. To avoid cross-contamination of DNA between unit cells the neck valve was closed and the
72 lower half was washed with PBS. The sandwich valves were then closed and the neck was released to allow the DNA to diffuse
73 into the lower half of the unit cell. After an incubation of 90 minutes the button was opened and the DNA was allowed to
74 attach to the neutravidin coated area. The unbound DNA could then be washed away while the button was pressurized. The
75 surface immobilized DNA was imaged by fluorescence microscopy. A 10 μL reaction volume of cell-free extract was prepared
76 according to a previous protocol (1) and flowed into the device. For the AND and OR logic gates, the pre-synthesis reaction
77 made up 25% of the fresh cell-free reaction mixture. The chip was separated into four parts using a multiplexer to test all
78 logic gate inputs on a single chip. All unit cells were isolated from one another by the sandwich valves before the button
79 was released and the chip was incubated at 29 $^{\circ}\text{C}$. The production of deGFP in each unit cell was monitored over time on
80 an automated fluorescence microscope (Nikon). The pneumatic valves, microscope, and camera were controlled by a custom
81 LabVIEW program throughout the experiment.

MITOMI measurements. DNA targets were either amplified with PCR or Klenow reactions and diluted over a range of 2 nM to
1 μM . The DNA was then spotted and the chip was aligned in the same way as described in the previous section. The same
surface chemistry was also performed, followed by an additional two steps. First, a biotinylated anti-His antibody (Qiagen, 15
nM) was flowed for 20 minutes, enabling its binding to the neutravidin coated region. Second, a 30 μL cell-free reaction, in
which a His-tagged ZF was expressed, was flowed for 25 minutes, immobilizing the ZF at the surface of each unit cell. The
DNA targets were then solubilized and allowed to diffuse into the lower half of the unit cell where their binding to the ZF could
be quantified via fluorescence microscopy. All image analysis was done using a custom MatLab script that either calculated the
mean of the fluorescence signal at the button area or in the solution of the unit cell. K_d s were determined by fitting the data
with a single binding site model

$$DNA_B = \frac{B_{max}DNA_F}{DNA_F + K_d},$$

82 where B_{max} is the maximum specific binding, DNA_F is the free DNA in solution, DNA_B is the DNA bound to protein at the
83 surface and K_d is the dissociation constant. Absolute K_d values in molar units were subsequently determined according to a
84 calibration made with known concentrations of Cy5-tagged DNA on-chip.

85 **Thermodynamic models for repression.** Following (5) we make the assumption that gene expression is proportional to RNAP
86 occupancy, and that RNAP binding to promoters is at thermodynamic equilibrium. This allows the final protein expression
87 level to be written as a function of the equilibrium occupancy of the promoters by RNAP, which itself is a function of the
88 available microstates. Thus the model can be constructed by enumerating all possible binding states of RNAP and any
89 transcription factor it interacts with. For concentrations of nonspecific sites and RNAPs given by N_{NS} and P , respectively, a
90 third assumption $P \ll N_{NS}$ allows the occupancy to be written in the following simple form,

$$p_{bound} = \frac{1}{1 + \frac{N_{NS}}{PF_{reg}} e^{\beta\Delta\epsilon_{pd}}}, \quad [1]$$

92 where the effects of the microstate distribution are subsumed into a regulation function F_{reg} . $\Delta\epsilon_{pd}$ is the energy difference
93 between RNAP binding to specific versus nonspecific sites. All energies are given in units of $k_B T$, and $\beta = 1/k_B T$.

Typically repression is modeled using a purely competitive model where RNAP and the repressor competes for exclusive
binding on the promoter. Motivated by our experimental observations, we extend this standard formulation of repression by
enumerating four possible states for repressor binding, with the following energies:

$$\begin{aligned} \text{no binding} &: 0 \\ \text{repressor only} &: \Delta\epsilon_{rd} \\ \text{RNAP only} &: \Delta\epsilon_{pd} \\ \text{repressor + RNAP} &: \Delta\epsilon_{rd} + \Delta\epsilon_{pd} + \epsilon_{rp} \end{aligned}$$

94 In our model, the repressor and RNAP can thus both bind at the same time, and interact with an energy ϵ_{rp} . As ϵ_{rp} becomes
95 large and positive, the repression tends to that of competitive inhibition, where either species excludes the other from binding.
96 This formulation allows for a continuous transition between competitive and noncompetitive mechanisms of inhibition: as
97 repression becomes noncompetitive, simultaneous binding is possible and thus the promoter exhibits a non-zero leak at full
98 repression.

99 We can then write down the regulation functions. For a single repressor of concentration R we have

$$F_{reg,1} = \frac{1 + \frac{R}{N_{NS}} e^{-\beta\Delta\epsilon_{rd}} e^{-\beta\epsilon_{rp}}}{1 + \frac{R}{N_{NS}} e^{-\beta\Delta\epsilon_{rd}}}. \quad [2]$$

101 For repressors binding to two sites, we have

$$102 \quad F_{reg,2} = \frac{1 + \frac{R_1}{N_{NS}} e^{-\beta \Delta \epsilon_{r1d}} e^{-\beta \epsilon_{r1p}} + \frac{R_2}{N_{NS}} e^{-\beta \Delta \epsilon_{r2d}} e^{-\beta \epsilon_{r2p}} + \frac{R_1 R_2}{N_{NS}^2} e^{-\beta(\Delta \epsilon_{r1d} + \Delta \epsilon_{r2d})} e^{-\beta(\epsilon_{r1p} + \epsilon_{r2p})} e^{-\beta \epsilon_{r12}}}{1 + \frac{R_1}{N_{NS}} e^{-\beta \Delta \epsilon_{r1d}} + \frac{R_2}{N_{NS}} e^{-\beta \Delta \epsilon_{r2d}} + \frac{R_1 R_2}{N_{NS}^2} e^{-\beta(\Delta \epsilon_{r1d} + \Delta \epsilon_{r2d})} e^{-\beta \epsilon_{r12}}} \quad [3]$$

103 where the two repressors can interact with an energy ϵ_{r12} . Positive values of ϵ_{r12} result in positive cooperativity, where the
104 binding of one repressor facilitates the binding of the other.

105 We make two transformations to Equation 3. First, in dose response experiments, we assume that repressor concentration
106 R is proportional to repressor DNA concentration $R = B d_R$, as supported by on-chip DNA titration measurements (Figure
107 1). This allows us to convert repressor concentrations to DNA concentrations, which is the experimentally varying quantity.
108 Second, we simplify all our expressions by defining an effective dissociation constant

$$109 \quad K_D = \frac{N_{NS}}{B} e^{+\beta \Delta \epsilon_{rd}}, \quad [4]$$

which gives the DNA, rather than the repressor concentration for half-maximum occupancy. The effective K_D s in the model
are related to standard physical dissociation constants $K_d = e^{\Delta G^\ominus / k_B T}$ by a multiplicative factor. These transformations result
in the simplified equations

$$F_{reg,1} = \frac{1 + \frac{d_R}{K_D} e^{-\beta \epsilon_{rp}}}{1 + \frac{d_R}{K_D}}$$

$$F_{reg,2} = \frac{1 + \frac{d_{R1}}{K_{D1}} e^{-\beta \epsilon_{r1p}} + \frac{d_{R2}}{K_{D2}} e^{-\beta \epsilon_{r2p}} + \frac{d_{R1} d_{R2}}{K_{D1} K_{D2}} e^{-\beta(\epsilon_{r1p} + \epsilon_{r2p})} e^{-\beta \epsilon_{r12}}}{1 + \frac{d_{R1}}{K_{D1}} + \frac{d_{R2}}{K_{D2}} + \frac{d_{R1} d_{R2}}{K_{D1} K_{D2}} e^{-\beta \epsilon_{r12}}}.$$

110 Finally, the protein level is given by a direct proportionality with the occupancy $y = A p_{bound}$, and fold repressions are given by
111 ratios of protein levels. Thus,

$$112 \quad y = \frac{A}{1 + \frac{C_0}{F_{reg}}} \quad [5]$$

with

$$C_0 = \frac{N_{NS}}{P} e^{\beta \Delta \epsilon_{pd}}.$$

For binding site tuning experiments, the independent variable is an experimentally measured K_d , in units of nM. The
regulation functions become

$$F_{reg,1} = \frac{1 + \frac{R}{K_d} e^{-\beta \epsilon_{rp}}}{1 + \frac{R}{K_d}}$$

$$F_{reg,2} = \frac{1 + \frac{R_1}{K_{d1}} e^{-\beta \epsilon_{r1p}} + \frac{R_2}{K_{d2}} e^{-\beta \epsilon_{r2p}} + \frac{R_1 R_2}{K_{d1} K_{d2}} e^{-\beta(\epsilon_{r1p} + \epsilon_{r2p})} e^{-\beta \epsilon_{r12}}}{1 + \frac{R_1}{K_{d1}} + \frac{R_2}{K_{d2}} + \frac{R_1 R_2}{K_{d1} K_{d2}} e^{-\beta \epsilon_{r12}}},$$

113 where R is now an effective repressor concentration, in units of nM.

To extend the model to take into account the distance-dependent effects, as well as the dependence of cooperativity on the
helical positioning of the ZF on the DNA backbone, we posit the phenomenological description

$$\epsilon_{r1p} = \epsilon_{r1p0} e^{-\lambda \Delta x} \quad [6]$$

$$\epsilon_{r12} = \epsilon_{r120} e^{-\lambda \Delta x} [1 + \cos(k \Delta x - \phi)], \quad [7]$$

114 where λ is a distance decay constant, Δx the spacing between the two ZF binding sites, k the wavenumber corresponding to
115 the DNA helical pitch of 10.5 bp/turn, and ϕ a phase shift. The model was parameterized using values obtained from dose
116 response measurements before fitting to the helical data. Finally, to take into account varying experimental conditions between
117 those measurements and the helical effect experiments, a third parameter R_0 was introduced as a multiplicative factor on DNA
118 concentrations, $d_{R1}^* = R_0 d_{R1}$ and $d_{R2}^* = R_0 d_{R2}$. This parameterizes global changes such as a different total DNA concentration
119 and protein production rates between the dose response and helical experiments.

120 **Markov chain Monte Carlo inference of model parameters.** Model fits to experimental data are carried out using Markov chain
121 Monte Carlo (MCMC) sampling, using the python package `emcee` (6) which is an implementation of Goodman and Weare's
122 affine-invariant ensemble sampling method (7). We first found a maximum likelihood estimate (MLE) of parameters using the
123 BFGS algorithm from the `scipy.optimize` package. We define our negative log-likelihood function as

$$124 \quad \ln p(y|x, params) = -\frac{1}{2} \sum_n \left[\frac{(y_n - \hat{y}_n(x, params))^2}{s_n^2} + \ln(2\pi s_n^2) \right] \quad [8]$$

125 where $\hat{y}_n(x, params)$ is the model prediction and y_n the experimental data with errors s_n . In order to combine results from
126 different experiments (for example, separate dose response curves), we added together the log-likelihood functions without
127 normalization.

128 Uninformative, broad normal priors were used, centered on the MLE parameter values. 50 Markov chains were initialized in
129 a tight ball around these MLE values, and allowed to run for 10,000 iterations. The first 5,000 points were considered as part
130 of the burn-in period and discarded; equilibration of the Markov chains was verified by inspection of the traces. Sampling the
131 equilibrated Markov chains returns the posterior probability distributions of the parameters, which can be used to generate an
132 ensemble of potential models (and hence, a distribution of fits as shown in Figure 5). The posterior distributions, as well as
133 pairwise distributions of parameters are shown in Figure S9.

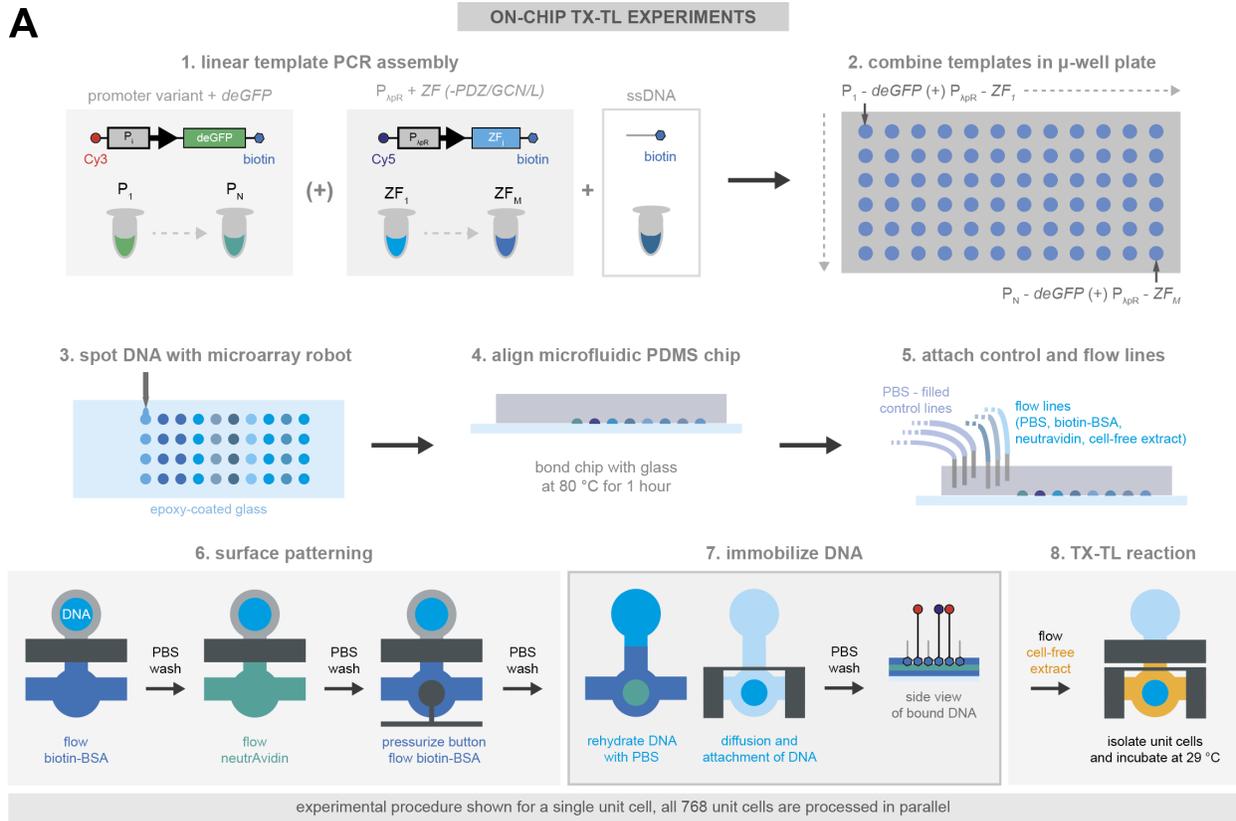
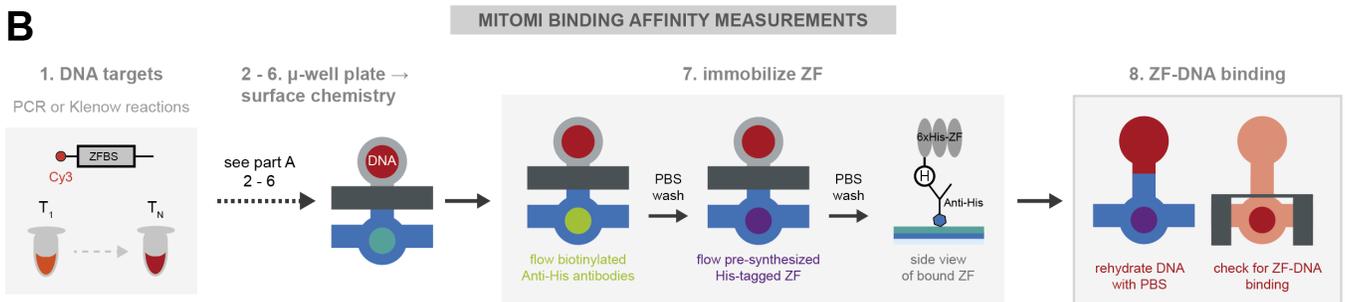
A**B**

Fig. S1. Overview of experimental protocol. Schematic showing the details of the experimental procedure for on-chip cell-free TX-TL reactions (**A**) and for measuring the binding affinity of ZF-DNA complexes with MITOMI (**B**).

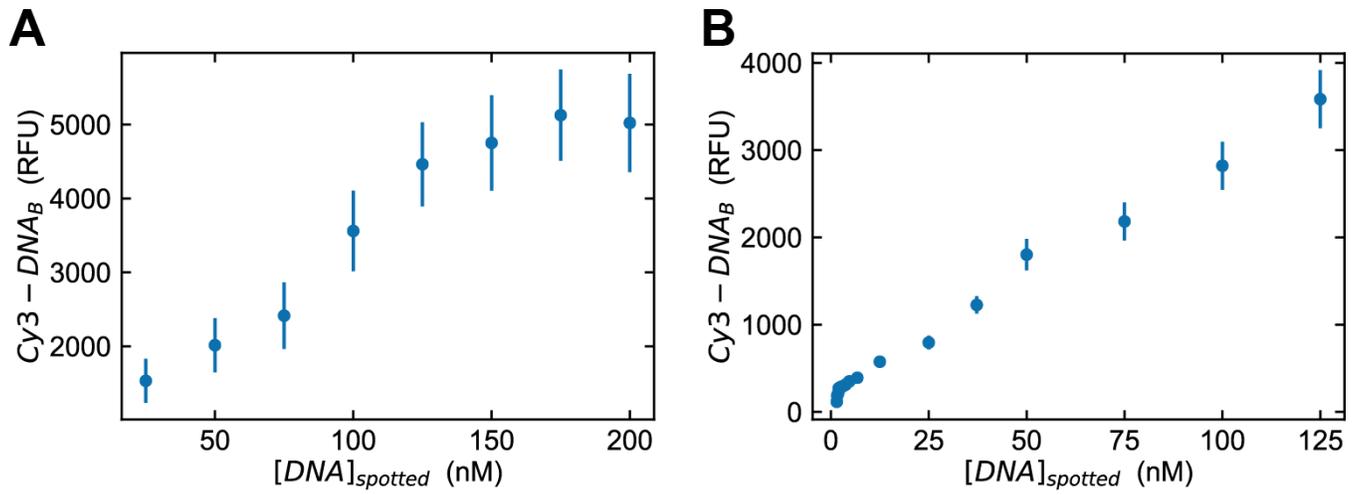


Fig. S2. DNA template spotting. DNA concentration measured on-chip versus the concentration of DNA in the spotting plate, when a single dsDNA template was diluted (**A**) versus a mixture of dsDNA template and ssDNA (**B**).

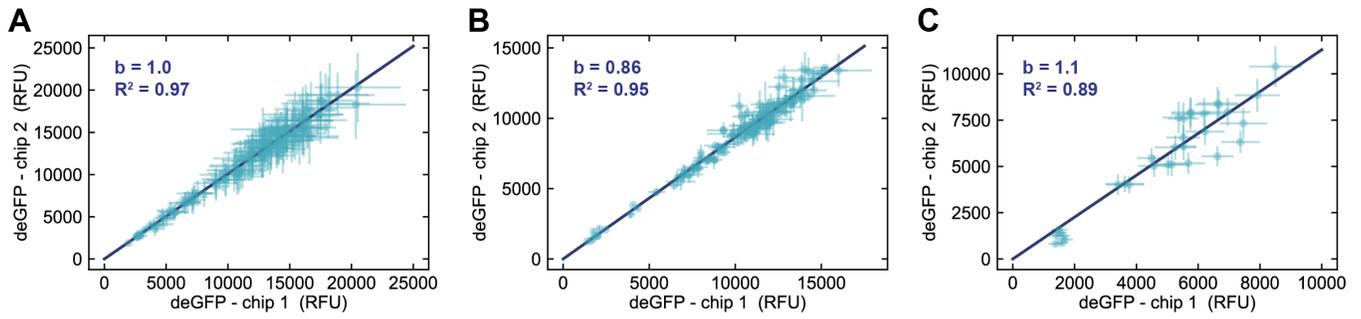


Fig. S3. Chip-to-chip reproducibility. (A) deGFP expression values obtained from two separate chips for the λP_R promoter mutagenesis library. The same DNA microplate was used to spot both chips. (B) A comparison of two separate chips measuring the output of all possible ZF – promoter pairs. Similar to when only a single DNA template is present per unit cell ((A)), two templates could be added with good reproducibility between chips. (C) Data from two separate chips with two or more DNA templates per unit cell (NAND logic gate). Each chip was prepared using different DNA microplates, showing a slightly increased variability for DNA templates derived from different PCR reactions.

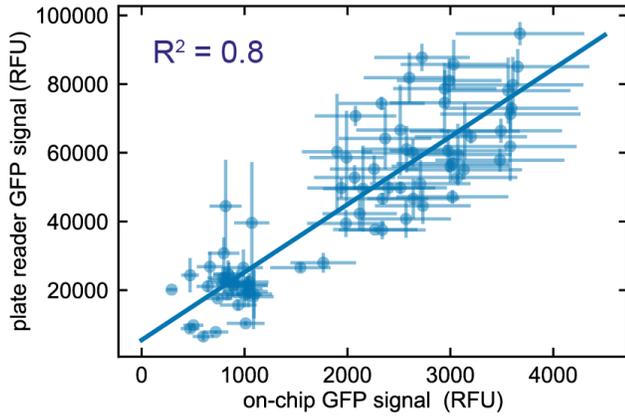
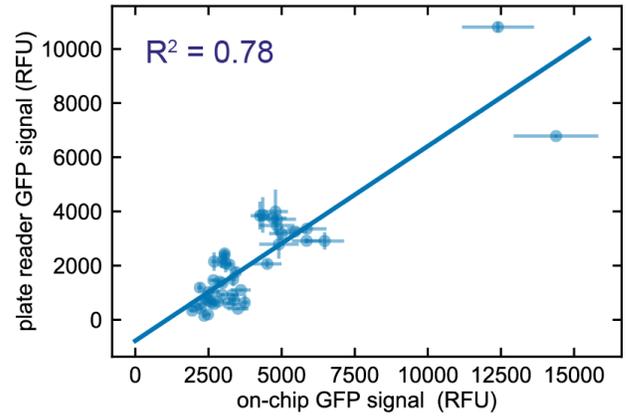
A**B**

Fig. S4. Comparison of on-chip and plate reader measurements. (A) Data collected from standard micro-well plate reader reactions versus data collected on-chip for a 9x9 ZF-target orthogonality matrix for targets with two binding sites in the promoter region. The 9 ZFs tested included AAA, ADB, ADD, BAB, BCB, BCC, BCD, BDD and CBD. (B) Plate reader versus on-chip data for cooperative, non-cooperative and control ZF PDZ-L heterodimers tested with a library of promoters that had variable spacing between the two ZF binding sites.

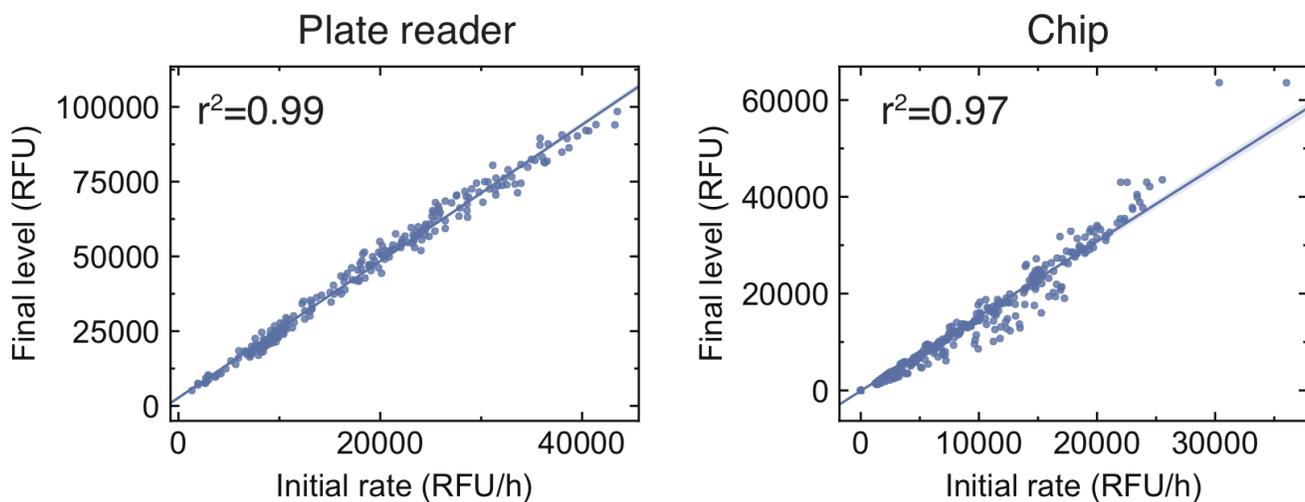


Fig. S5. Relationship between initial rates and final deGFP values. We observe a linear relationship between the initial rate of deGFP production, as obtained by linear fits to the time course at early times (~30 minutes), and the final steady state level of deGFP, on both plate reader and chip experiments. This suggests that final levels of deGFP are proportional to the production rate, and validates the use of endpoint protein levels as a proxy for transcription rates.

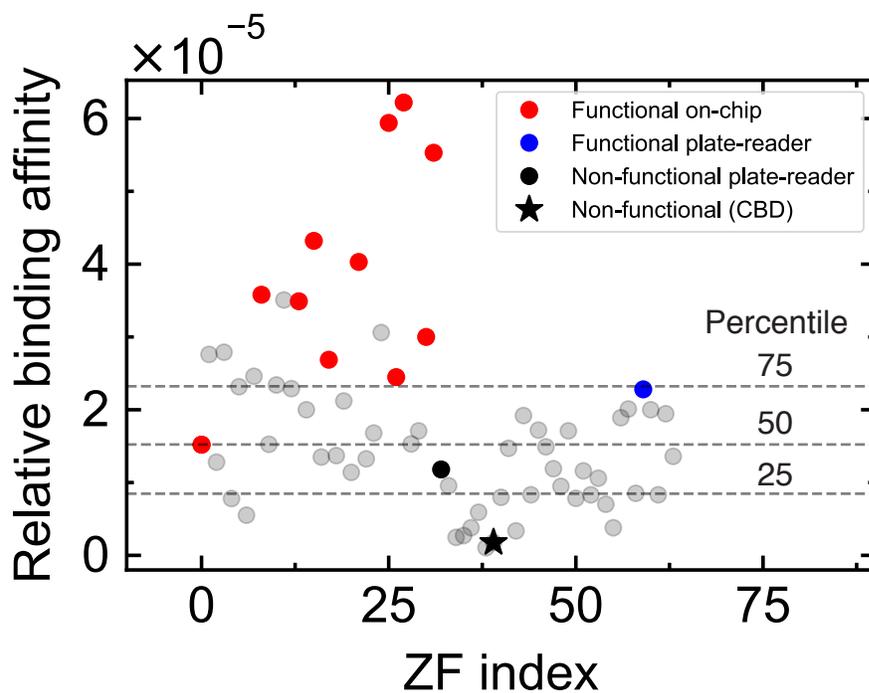


Fig. S6. Functional ZFs within the combinatorial library. This figure shows the on-target binding affinity for all 64 members of our combinatorial library (data from (8)). In red are the 11 functional ZFs characterised on-chip. Additionally, the blue and black points are two other ZFs characterised in a separate plate-reader experiment. The black star corresponds to ZF_{CBD} . Horizontal dashed lines represent the 25th, 50th, and 75th percentile of binding affinity, and we observe that of the ZFs tested, all those with binding affinities greater than the 50th percentile exhibited functional repression. We therefore hypothesize that of the untested ones, those with affinities greater than approximately 1.5×10^{-5} in relative units should be functional. This would correspond to a total of 32 functional repressors.

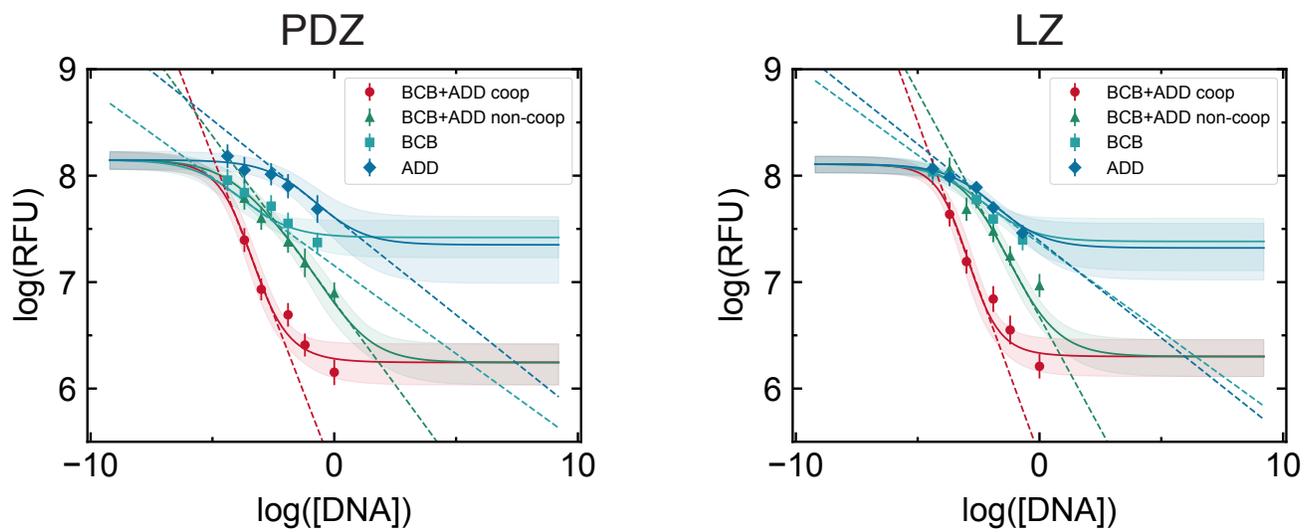


Fig. S7. Sensitivity calculations from dose response curves. The slope of the dose response curve, as measured in the linear regime of a log-log plot, is defined as the sensitivity; this quantity increases in the presence of cooperative interactions. Values are given in Table S1.

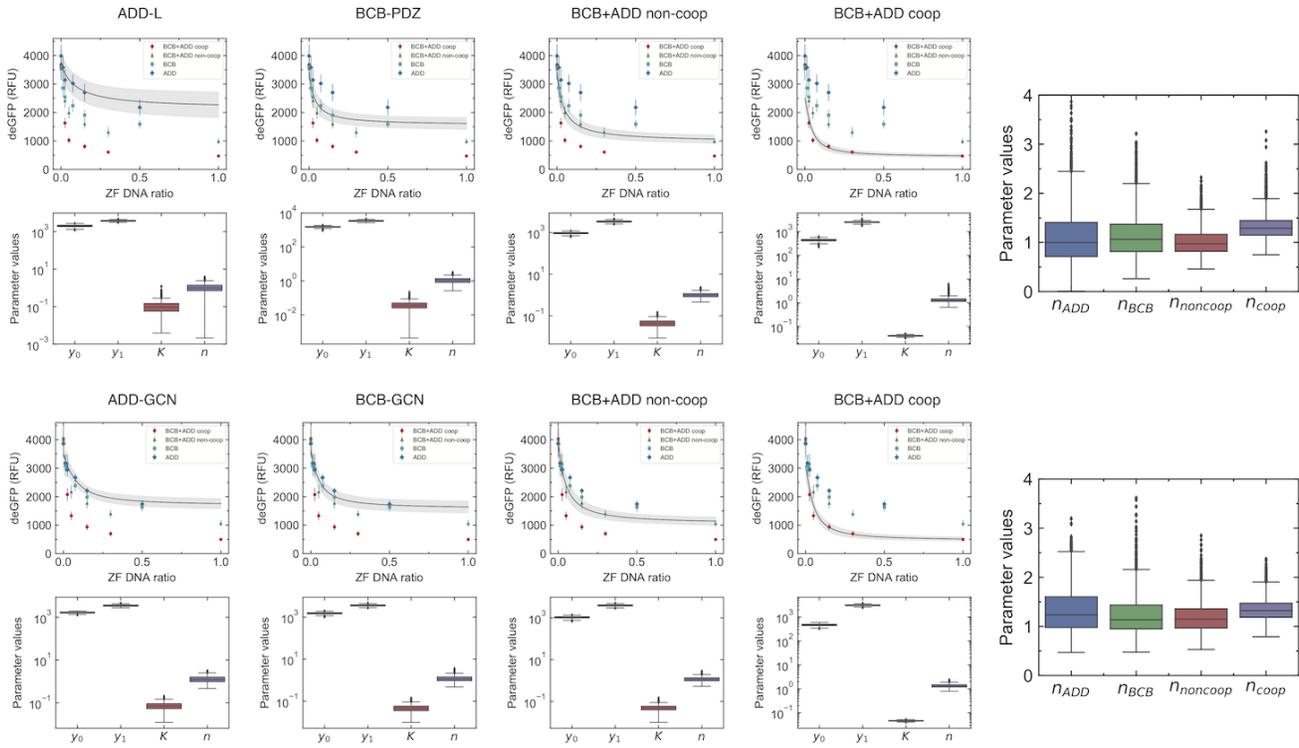
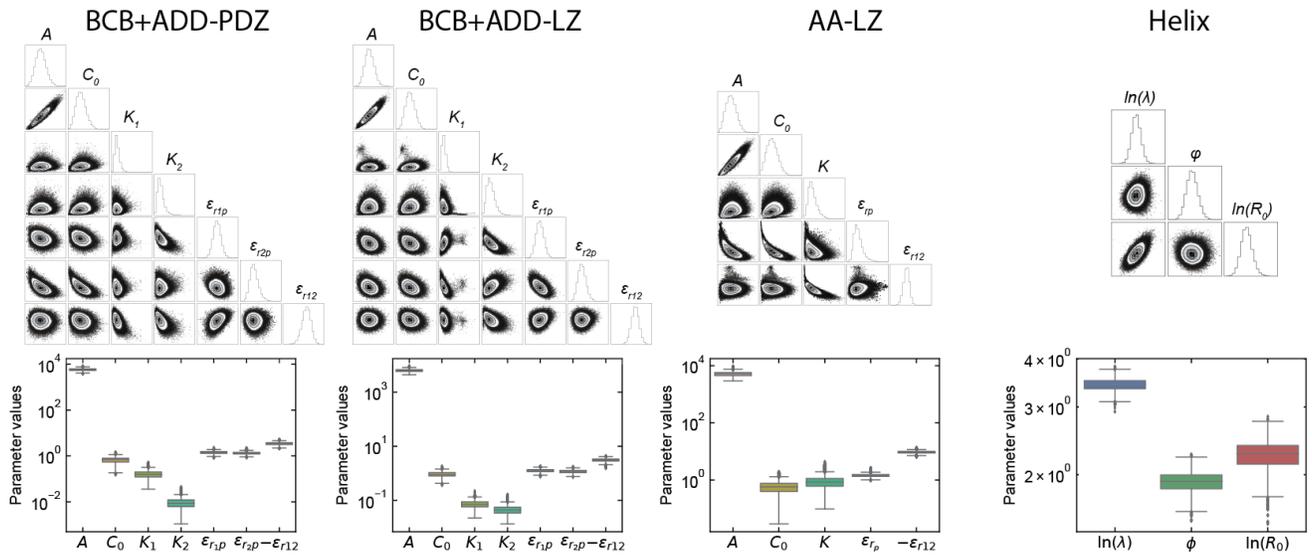


Fig. S8. Hill function fits to dose response curves. Standard Hill functions take the form $y = y_0 + \frac{y_1 - y_0}{1 + \left(\frac{R}{K_D}\right)^n}$, where y_0 is the leak, y_1 the maximum expression, R the repressor concentration, K_D the dissociation constant, and n the Hill coefficient. Unlike the thermodynamic model which proposes a mechanism to consistently fit the entire data set, Hill functions must be independently fit to each dose response curve. We first fit the single ZF data, followed by the non-cooperative BCB+ADD curve. Hill functions describe two-site binding using an effective K_D and varying n . We make the assumption that in both the cooperative and non-cooperative case, the effective K_D is the same. Thus the cooperative BCB+ADD curve is fit using the K_D value obtained from the non-cooperative BCB+ADD curve. Although the Hill coefficient is predicted to increase in the presence of cooperativity, we observe a minimal change in our data, within the error of the fits.

Dose response fits



Site tuning fits

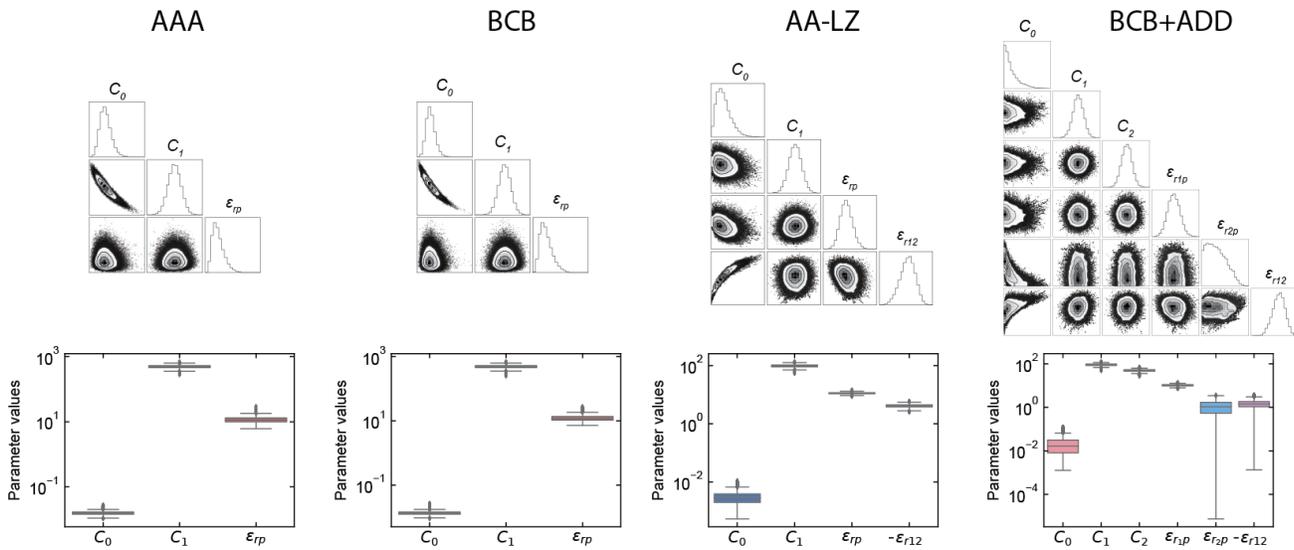


Fig. S9. Markov chain Monte Carlo inference of model parameters. Matrix plots show pairwise posterior probability distributions of model parameters for each of the models fitted. The individual posterior distribution for each parameter can also be visualized using box plots; the top and bottom of each box represents the 75th and 25th percentile, respectively.

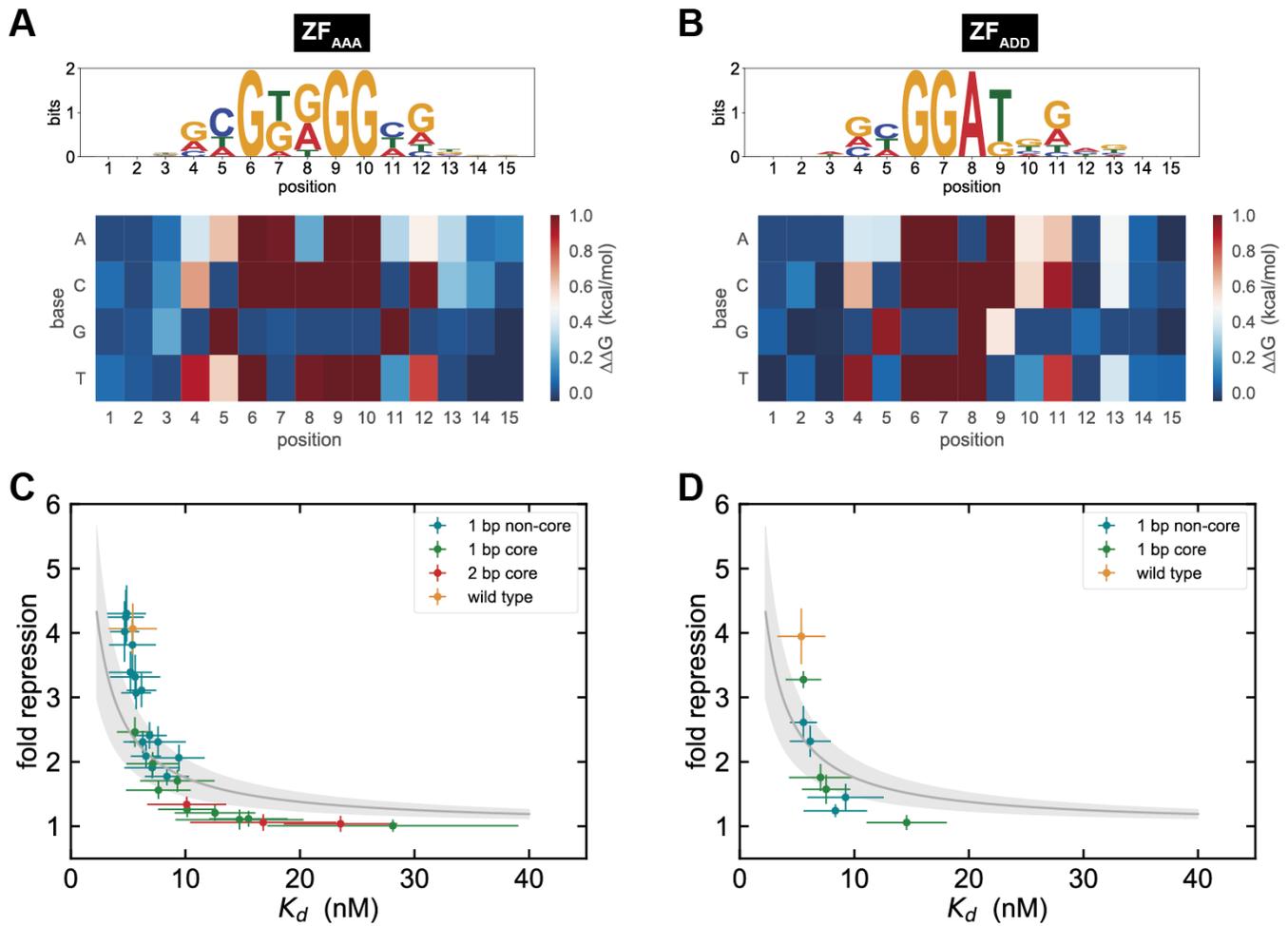


Fig. S10. PWMs ZF_{AAA} and ZF_{ADD} and binding site tuning for ZF_{AAA} . Sequence logos and PWMs measured by MITOMI for ZF_{AAA} (A) and ZF_{ADD} (B). Fold repression versus K_d measured for mutations to a ZF_{AAA} binding site within a single binding site promoter (C) and a cooperative promoter (D). Solid lines represent the model fits.

-35 -10

ACACCGTGCGTGT**TTGACA**ATTTTACCTCTGGCGGT**GATAAT** promoter λP_R

---X---XXX---X---XXX-X---XX- location of selected mutations

ACACCGTGCGTGT**TTGACA**AT**GAGGTAGTG**TGCGGT**GATAAT** promoter pZF_{BCB}

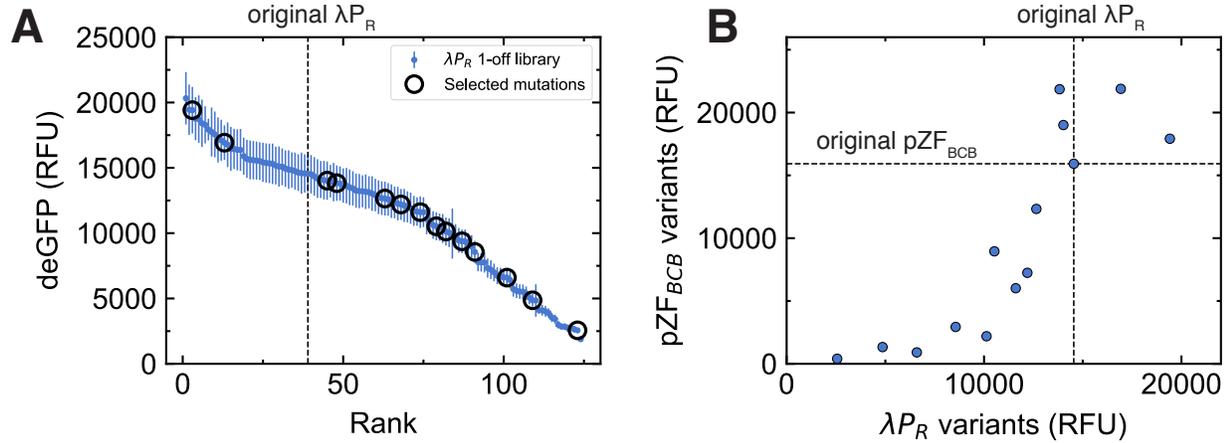


Fig. S11. Tuning promoter strength. In Figure 1 we presented the one-off library data which contains all single base mutations from positions -47 to -7 of the λP_R promoter. **(A)** These data can be rank-ordered, and we observe that a large dynamic range of promoter output is accessible. A number of mutations were selected which both increased and decreased expression relative to the original promoter. **(B)** Introducing a single *BCB* binding site between the -35 and -10 boxes generates the pZF_{BCB} promoter. Applying the selected mutations to this new promoter changes the output in a correlated way when compared to the λP_R promoter; the new promoter can thus be roughly tuned in a predictive fashion.

134 **References**

- 135 1. Sun ZZ, et al. (2013) Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic
136 Biology. *Journal of Visualized Experiments* 79:e50762.
- 137 2. Shin J, Noireaux V (2010) Efficient cell free expression with the endogenous E. coli RNA polymerase and sigma factor 70.
138 *J. Biol Eng* 4(8).
- 139 3. Sun ZZ, Yeung E, Hayes CA, Noireaux V, Murray RM (2014) Linear DNA for Rapid Prototyping of Synthetic Biological
140 Circuits in an Escherichia coli Based TX-TL Cell-Free System. *ACS Synthetic Biology* 3(6):387–397.
- 141 4. Maerkl SJ, Quake SR (2007) A systems approach to measuring the binding energy landscapes of transcription factors.
142 *Science* 315(5809):233–237.
- 143 5. Bintu L, et al. (2005) Transcriptional regulation by the numbers: models. *Current Opinion in Genetics and Development*
144 15:116–124.
- 145 6. Foreman-Mackey D, Hogg DW, Lang D, Goodman J (2013) emcee: The MCMC Hammer. (arXiv:1202.3665v4).
- 146 7. Goodman J, Weare J (2010) Ensemble samplers with affine invariance. *Comm. App. Math. and Comp. Sci.* 5(1):65–80.
- 147 8. Blackburn MC, Petrova E, Correia BE, Maerkl SJ (2015) Integrating gene synthesis and microfluidic protein analysis for
148 rapid protein engineering. *Nucleic Acids Research* 44(7):e68.

Supplementary Tables

Table S1: **Quantitative measures of cooperative interactions.**

Design: PDZ	Sensitivity	Hill coefficient
BCB+ADD cooperative	0.592±0.007	1.3±0.2
BCB+ADD non-cooperative	0.312±0.003	1.0±0.2
BCB	0.165±0.001	1.1±0.4
ADD	0.183±0.001	1.1±0.4

Design: LZ	Sensitivity	Hill coefficient
BCB+ADD cooperative	0.627±0.009	1.3±0.2
BCB+ADD non-cooperative	0.422±0.002	1.1±0.3
BCB	0.166±0.001	1.2±0.4
ADD	0.182±0.001	1.3±0.5

Table S2: **Inference of parameter values.** The posterior probability distribution of each parameter was quantified by reporting the value for the 50th percentile (large numbers) as well as uncertainty bounds given by the 16th and 84th percentiles (small numbers)

Dose response fits

Parameters: PDZ	Value	Units
A	+5.76e + 03 ^{+8.02e+02} _{-7.19e+02}	RFU
C ₀	+6.62e - 01 ^{+2.28e-01} _{-1.99e-01}	
K ₁	+1.66e - 01 ^{+7.12e-02} _{-5.09e-02}	
K ₂	+8.24e - 03 ^{+4.67e-03} _{-3.01e-03}	
ε _{r1p}	+1.43e + 00 ^{+1.81e-01} _{-1.78e-01}	k _B T
ε _{r2p}	+1.32e + 00 ^{+1.88e-01} _{-1.57e-01}	k _B T
ε _{r12}	-3.46e + 00 ^{+4.36e-01} _{-4.66e-01}	k _B T

Parameters: LZ	Value	Units
A	+6.26e + 03 ^{+7.54e+02} _{-6.95e+02}	RFU
C ₀	+8.84e - 01 ^{+2.18e-01} _{-1.95e-01}	
K ₁	+6.86e - 02 ^{+2.50e-02} _{-2.02e-02}	
K ₂	+4.27e - 02 ^{+1.92e-02} _{-1.45e-02}	
ε _{r1p}	+1.28e + 00 ^{+1.86e-01} _{-1.74e-01}	k _B T
ε _{r2p}	+1.19e + 00 ^{+1.77e-01} _{-1.54e-01}	k _B T
ε _{r12}	-3.05e + 00 ^{+3.92e-01} _{-3.99e-01}	k _B T

Parameters: AA-LZ	Value	Units
A	+5.13e + 03 ^{+8.87e+02} _{-8.27e+02}	RFU
C ₀	+5.70e - 01 ^{+2.77e-01} _{-2.48e-01}	
K	+8.42e - 01 ^{+4.79e-01} _{-3.16e-01}	
ε _{rp}	+1.44e + 00 ^{+1.98e-01} _{-1.35e-01}	k _B T
ε _{r12}	-9.48e + 00 ^{+7.85e-01} _{-8.39e-01}	k _B T

Parameters: Helical	Value	Units
ln(λ)	-3.41e + 03 ^{+1.18e-01} _{-1.16e-01}	
φ	-1.92e + 00 ^{+1.21e-01} _{-1.14e-01}	
ln(R ₀)	-2.24e + 00 ^{+1.78e-01} _{-1.72e-01}	

Binding site tuning fits

Parameters: AAA	Value	Units
C_0	$+1.55e - 02^{+1.93e-03}_{-1.57e-03}$	
R	$+4.94e + 02^{+5.04e+01}_{-5.09e+01}$	nM
ϵ_{rp}	$+1.13e + 01^{+2.92e+00}_{-1.94e+00}$	$k_B T$

Parameters: BCB	Value	Units
C_0	$+1.35e - 02^{+1.62e-03}_{-1.33e-03}$	
R	$+4.93e + 02^{+5.03e+01}_{-5.05e+01}$	nM
ϵ_{rp}	$+1.20e + 01^{+2.78e+00}_{-1.85e+00}$	$k_B T$

Parameters: AA-LZ	Value	Units
C_0	$+2.82e - 03^{+1.64e-03}_{-1.15e-03}$	
R	$+9.78e + 01^{+1.00e+01}_{-1.01e+01}$	nM
ϵ_{rp}	$+1.12e + 01^{+7.33e-01}_{-6.70e-01}$	$k_B T$
ϵ_{r12}	$-4.12e + 00^{+4.72e-01}_{-5.37e-01}$	$k_B T$

Parameters: BCB/ADD	Value	Units
C_0	$+1.64e - 02^{+2.38e-02}_{-1.10e-02}$	
R_1	$+9.08e + 01^{+9.06e+00}_{-8.80e+00}$	nM
R_2	$+5.00e + 01^{+4.81e+00}_{-5.09e+00}$	nM
ϵ_{r1p}	$+1.04e + 01^{+9.03e-01}_{-8.68e-01}$	$k_B T$
ϵ_{r2p}	$+1.06e + 00^{+9.13e-01}_{-7.10e-01}$	$k_B T$
ϵ_{r12}	$-1.45e + 00^{+5.34e-01}_{-5.94e-01}$	$k_B T$

Table S3: **ZF expression templates.** 3-finger ZFs were expressed from lambda P_R promoters (blue, underlined) with conserved 5'UTR (grey, underlined), and could be modified to contain interaction domains and ligands (red).

P _R -ZF	Cy5gacatggtgaagactatcgccaccatcagccagaaaaccgaatTTTgctgggtgggcta acgatatccgcctgatgCGTgaacgtgacggacgtaaccaccgacatgtgtgtgctgtt ccgctgggcatgctgagctaaccaccgtgCGTg <u>ttgacaatTTTtacctctggcggtgataat</u> ggttgCagctagcaataatTTTgtttaaactTTtaagaaggagatataccATG----ZF---- TAATAagaatcaggggataacgcaggaaaga-BIOT
P _R -ZF + cognate ligand VKESLV	Cy5gacatggtgaagactatcgccaccatcagccagaaaaccgaatTTTgctgggtgggcta acgatatccgcctgatgCGTgaacgtgacggacgtaaccaccgacatgtgtgtgctgtt ccgctgggcatgctgagctaaccaccgtgCGTg <u>ttgacaatTTTtacctctggcggtgataat</u> ggttgCagctagcaataatTTTgtttaaactTTtaagaaggagatataccATG----ZF---- GGCAGCGGCAGCGTAAAGAAAGCCTGGTGAATAAgaatcaggggataacgcaggaaaga -BIOT
P _R -ZF + non- cognate ligand VKEAAA	Cy5gacatggtgaagactatcgccaccatcagccagaaaaccgaatTTTgctgggtgggcta acgatatccgcctgatgCGTgaacgtgacggacgtaaccaccgacatgtgtgtgctgtt ccgctgggcatgctgagctaaccaccgtgCGTg <u>ttgacaatTTTtacctctggcggtgataat</u> ggttgCagctagcaataatTTTgtttaaactTTtaagaaggagatataccATG----ZF---- GGCAGCGGCAGCGTAAAGAAAGCGCGCGTAAATAAgaatcaggggataacgcaggaaaga -BIOT
P _R -ZF-PDZ	Cy5gacatggtgaagactatcgccaccatcagccagaaaaccgaatTTTgctgggtgggcta acgatatccgcctgatgCGTgaacgtgacggacgtaaccaccgacatgtgtgtgctgtt ccgctgggcatgctgagctaaccaccgtgCGTg <u>ttgacaatTTTtacctctggcggtgataat</u> ggttgCagctagcaataatTTTgtttaaactTTtaagaaggagatataccATG----ZF---- AAAGACGGTGGCGGCGGTTCTCCAGCGCCAGCGCCACCGGAAGCGCTGCAGCGCCGCGCG TGACCGTGCGCAAAGCGGATGCGGGCGGCTGGGCATTAGCATTAAAGCGCGCCGCGAAAA CAAATGCCGATTCTGATTAGCAAAATTTTTAAAGGCCTGGCGGCGGATCAGACCGAAGCG CTGTTTGTGGCGATGCGATTCTGAGCGTGAACGGCGAAGATCTGAGCAGCGCACCCATG ATGAAGCGGTGCAGCGCTGAAAAAACGGCAAAGAAGTGGTGTGGAAGTAAATATAT GAAAGAAAgcttgatatacgaattcctgcagccccggggatcccatggtacgcgtgctag aggcatcaaataaaacgaaaggctcagtcgaaagactgggcttttcgtttatctgtttgtt tgtcggatgaacgctctcctgagtaggacaaatccgccccctagacctagcgcttcggctg cggcgagcggatcagctcactcaaaggcggttaatacggttatc-BIOT
P _R -ZF-GCN4	Cy5gacatggtgaagactatcgccaccatcagccagaaaaccgaatTTTgctgggtgggcta acgatatccgcctgatgCGTgaacgtgacggacgtaaccaccgacatgtgtgtgctgtt ccgctgggcatgctgagctaaccaccgtgCGTg <u>ttgacaatTTTtacctctggcggtgataat</u> ggttgCagctagcaataatTTTgtttaaactTTtaagaaggagatataccATG----ZF---- AAAGACGGTGGCGGCGGTTCTCCAGCGCCAGCGCCACCGCATCGCGATTTTCAGCATATTC TGCCGATTCTGGAAGATAAAGTGAAGAAGTCTGAGCAAAAATCATCTGAAAAACGA AGTGGCGCGCCTGAAAAAAGTGGTGGCGGAACGCTAATAAgaatcaggggataacgcag gaaaga-BIOT
P _R -ZF- GCN4mut (L67A- L68A-L74A- L81A)	Cy5gacatggtgaagactatcgccaccatcagccagaaaaccgaatTTTgctgggtgggcta acgatatccgcctgatgCGTgaacgtgacggacgtaaccaccgacatgtgtgtgctgtt ccgctgggcatgctgagctaaccaccgtgCGTg <u>ttgacaatTTTtacctctggcggtgataat</u> ggttgCagctagcaataatTTTgtttaaactTTtaagaaggagatataccATG----ZF---- AAAGACGGTGGCGGCGGTTCTCCAGCGCCAGCGCCACCGCATCGCGATTTTCAGCATATTC TGCCGATTCTGGAAGATAAAGTGAAGAAGCGGCGAGCAAAAATCATCTGAAAAACGA AGTGGCGCGCGCAAAAAGTGGTGGCGGAACGCTAATAAgaatcaggggataacgcag gaaaga-BIOT

Table S4: **Three-finger ZF coding sequences.** Sequences for each ZF are given here, excluding only the start and stop codons. Variable recognition helices for each finger (1-3) are highlighted. ZFs bind in the 3'-5' direction by convention, and thus to construct, for example, ZF_{ADD} the fingers must be arranged in reverse order F1(D)-F2(D)-F3(A).

ZF _{AAA}	GAACGTCCGTACGCTTGCCCGGTTGAATCTTGCGACCGTCGTTTCTCT <u>CGTTCTGACGAAC</u> <u>TGACCCGT</u> CATATTAGAATTCATACTGGACAAAAACCATTCCAATGTAGAATTTGTATGAG AAATTTCTCT <u>CGTTCTGACCACCTGACCACC</u> CACATCCGTACCCACACCCGGTGAAAAACCG TTCGCTTGCGACATCTGCGGTCGTAATTCGCT <u>CGTTCTGACGAACGTAACCGT</u> CACACCA AAATCCACCTGCGTCAG
ZF _{BBB}	GAACGTCCGTACGCTTGCCCGGTTGAATCTTGCGACCGTCGTTTCTCT <u>CGTAACTTCATCC</u> <u>TGCAGCGT</u> CATATTAGAATTCATACTGGACAAAAACCATTCCAATGTAGAATTTGTATGAG AAATTTCTCT <u>GACCGTGCTAACCTGCGTCGT</u> CACATCCGTACCCACACCCGGTGAAAAACCG TTCGCTTGCGACATCTGCGGTCGTAATTCGCT <u>CGTCACGACCAGCTGACCCGT</u> CACACCA AAATCCACCTGCGTCAG
ZF _{CCC}	GAACGTCCGTACGCTTGCCCGGTTGAATCTTGCGACCGTCGTTTCTCT <u>GACTCTCCGACCC</u> <u>TGCGTCGT</u> CATATTAGAATTCATACTGGACAAAAACCATTCCAATGTAGAATTTGTATGAG AAATTTCTCT <u>CAGCGTTCTTCTCTGGTTTCGT</u> CACATCCGTACCCACACCCGGTGAAAAACCG TTCGCTTGCGACATCTGCGGTCGTAATTCGCT <u>GAACGTGGTAACCTGACCCGT</u> CACACCA AAATCCACCTGCGTCAG
ZF _{DDD}	GAACGTCCGTACGCTTGCCCGGTTGAATCTTGCGACCGTCGTTTCTCT <u>GACAAAACCAAAC</u> <u>TGCGTGTT</u> CATATTAGAATTCATACTGGACAAAAACCATTCCAATGTAGAATTTGTATGAG AAATTTCTCT <u>GTTTCGTCACAACCTGACCCGT</u> CACATCCGTACCCACACCCGGTGAAAAACCG TTCGCTTGCGACATCTGCGGTCGTAATTCGCT <u>CAGTCTACCTCTCTGCAGCGT</u> CACACCA AAATCCACCTGCGTCAG
Example: ZF _{ADD}	GAACGTCCGTACGCTTGCCCGGTTGAATCTTGCGACCGTCGTTTCTCT <u>GACAAAACCAAAC</u> <u>TGCGTGTT</u> CATATTAGAATTCATACTGGACAAAAACCATTCCAATGTAGAATTTGTATGAG AAATTTCTCT <u>GTTTCGTCACAACCTGACCCGT</u> CACATCCGTACCCACACCCGGTGAAAAACCG TTCGCTTGCGACATCTGCGGTCGTAATTCGCT <u>CGTTCTGACGAACGTAACCGT</u> CACACCA AAATCCACCTGCGTCAG

Table S5: **Two-finger ZF coding sequence.**

ZF _{AA}	AAACCGTTTCAGTGCCGATTTGCATGCGCAACTTTAGCCGACGATCATCTGACCACCC ATATTCGCACCCATACCGGCGAAAAACCGTTTTCGCTGCGATATTTGCGGCCGAAATTCGC CCGCAGCGATGAACGCAAACGC
------------------	---

Table S6: Reporter and target promoter design. deGFP was used as a reporter. The upstream regulatory region consisted of variations of the original lambda P_R promoter containing 9-bp ZF binding sites (N) and their flanking bases (X).

P _R -deGFP	Cy3ccagccagaaaacgacctttctgtggtgaaaccggatgctgcaattcagagcggcagc aagtgggggacagcagaagacctgaccgccgagagtggatgtttgacatggtgaagacta tcgaccatcagccagaaaaccgaattttgctgggtgggctaacgatatccgcctgatgcg tgaacgtgacggacgtaaccaccgacatgtgtgtgctgttccgctgggcatgctgagct aacaccgtgctgttgacaatttacctctggcggtgataatggttgacgctagcaataat ttgtttaactttaagaaggagataaccATGGAGCTTTTCACTGGCGTTGTTCCCATCCT GGTCGAGCTGGACGGCGACGTAACCGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC GATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCCACCACCGGAAGCTGCCCGTGC CCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGTTCAGCCGCTACCCCGA CCATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGC ACCATCTTCTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGCAACATCCT GGGCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCATGGCCGACAAGCAG AAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGC TCGCCGACCACTACCAGCAGAACACCCCATCGGGCAGCGCCCGTGTCTGCCGACAA CCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGGATCACATG GTCCTGCTGGAGTTCGTGACCGCCCGGGATCTAAActcgagcaaagcccgcgaaaggcg ggcttttctgtgctgaccgatgcccttgagagccttcaaccagtcagctccttccggtgg gcgcggggcatgactatcgctcgccgacttatgactgtcttctttatcatgcaactcgtag gacaggtgccggcagcgtcttccgcttccctgctcactgactcgtcgtcgctcggtcgttc ggctgcggcgagcggtatcagctcactcaaggcggtataacggttatccacagaatcagg ggataacgcaggaaaga-BIOT
P _{ZF} Single site	Cy3ccagccagaaaacgacctttctgtggtgaaaccggatgctgcaattcagagcggcagc aagtgggggacagcagaagacctgaccgccgagagtggatgtttgacatggtgaagacta tcgaccatcagccagaaaaccgaattttgctgggtgggctaacgatatccgcctgatgcg tgaacgtgacggacgtaaccaccgacatgtgtgtgctgttccgctgggcatgctgagct aacaccgtgctgttgacaXNNNNNNNNXggcggtgataat
P _{ZF} Two sites	Cy3ccagccagaaaacgacctttctgtggtgaaaccggatgctgcaattcagagcggcagc aagtgggggacagcagaagacctgaccgccgagagtggatgtttgacatggtgaagacta tcgaccatcagccagaaaaccgaattttgctgggtgggctaacgatatccgcctgatgcg tgaacgtgacggacgtaaccaccgacatgtgtgtgctgttccgctgggcatgctgagct aacXNNNNNNNNXttgacaXNNNNNNNNXggcggtgataat
P _{ZF} Cooperative	Cy3ccagccagaaaacgacctttctgtggtgaaaccggatgctgcaattcagagcggcagc aagtgggggacagcagaagacctgaccgccgagagtggatgtttgacatggtgaagacta tcgaccatcagccagaaaaccgaattttgctgggtgggctaacgatatccgcctgatgcg tgaacgtgacggacgtaaccaccgacatgtgtgtgctgttccgctgggcatgctgagct aacXNNNNNNNNXttgacaXNNNNNNNNXggcggtgataat
P _{ZF-AA}	Cy3ccagccagaaaacgacctttctgtggtgaaaccggatgctgcaattcagagcggcagc aagtgggggacagcagaagacctgaccgccgagagtggatgtttgacatggtgaagacta tcgaccatcagccagaaaaccgaattttgctgggtgggctaacgatatccgcctgatgcg tgaacgtgacggacgtaaccaccgacatgtgtgtgctgttccgctgggcatgctgagct aacaccgtgctgttgacaCGCCACGCGTGGGCGTgataat

Table S7: **Three-finger ZF target sequences.** Each ZF binds a 9bp target; for each target we conserve single flanking bases on either side where possible.

Target AAA	5'-a GCG TGG GCG t-3'
Target BBB	5'-t GAG GAC GTG t-3'
Target CCC	5'-a GAT GTA GCC t-3'
Target DDD	5'-t GTA GAT GGA g-3'
Example: target ADD	5'-a GCG GAT GGA g-3'

Table S8: **MITOMI PWM targets.** The bases which were mutated are colored blue and the bold bases represent the core binding sequence.

Target ZF _{AAA}	5'-GGCCAA AATGCGTGGGCGT CGTTTTCCGGCGGTATGAC-3'
Target ZF _{ADD}	5'-GGCCAA AAAGCGGATGGAGGC TTTTCCGGCGGTATGAC-3'
Target ZF _{BCB}	5'-GGCCAA AATGAGGTAGTGTGC TTTTCCGGCGGTATGAC-3'
Target ZF _{AA}	5'-GGCCAA ATCCCACGCGTGGG AGTTTTCCGGCGGTATGAC-3'