PNAS
www.pnas.org

Supplementary Information for

Massively parallel CRISPRi assays reveal concealed thermodynamic determinants of dCas12a binding

David A. Specht, Yasu Xu and Guillaume Lambert

Guillaume Lambert.
E-mail: lambert@cornell.edu

This PDF file includes:

Supplementary text
Figs. S1 to S7
Legends for Dataset S1 to S3

Other supplementary materials for this manuscript include the following:

Datasets S1 to S3

Supporting Information Text

1. Massively parallel CRISPR interference assay

Growth rate from sequencing counts.

Each experiment contains a total of n different crRNA:DNA combinations. The total number of transformant after plasmid assembly is $M_T = \sum_{i=1}^{n} x_i$, where x_i is the number of cells with a specific crRNA-target DNA combination i. If the assembly of each feature x_i does not depend on the underlying DNA sequence, the distribution of x_i will be given by a Poisson distribution with rate $\frac{M_T}{n}$.

The sample is then grown under two different selection conditions (K and SK) until it reaches an optical density OD_{600}=1. The time needed for each the population to reach OD_{600}=1 in each condition is given by τ_K and τ_S, respectively. Plasmids are collected when each flask reaches OD=1 and the region containing the crRNA and target DNA coding sequence is amplified using NEBNext Multiplex Oligos for Illumina (Index Primers Set 1).

Each sample is then sequenced using either the Miseq or iSeq 100 platform, and the number of times each crRNA-target DNA combination i is present in the population is denoted by K_i and S_i for cells grown under K and SK conditions, respectively. Each feature i will grow at a rate Λ_s^i under SK selection and at a constant rate $\Lambda_k^i = \Lambda_0$ under K selection.

We define the relative fraction of each feature i in each condition according to:

$$\begin{cases} k_i = \dfrac{K_i}{\sum_i^n K_i} = \dfrac{x_i e^{\Lambda_0 \tau_k}}{M_1} \\ s_i = \dfrac{S_i}{\sum_i^n S_i} = \dfrac{x_i e^{\Lambda_S^i \tau_s}}{M_1} \end{cases} \quad [1]$$

where M_1 is the number of cells at the end of the experiment when the flask reaches OD=1. To find the effective growth rate Λ_s^i under SK-condition for all other features i, we can re-arrange eqn. 1 to get

$$\Lambda_s^i = \frac{log\left(\frac{s_i}{k_i}\right) + \Lambda_0 \tau_k}{\tau_s} \quad [2]$$

Determining the growth time τ_k under K-selection. A subset of the population will not fully repress the tetA-sacB cassette and will not grow in SK conditions. This happens, for example, when the crRNA/target DNA Hamming distance $\mathcal{H}$ is 6. If we let s_- and k_- bet the population fraction under K and SK selection conditions of the non-growing subpopulation x_-, we get

$$\begin{cases} k_- &= \dfrac{x_- exp(\Lambda_0 \tau_k)}{M_1} \\ s_- &= \dfrac{x_-}{M_1} \end{cases} \quad [3]$$

Using this, we can find the time τ_k cells were growing under K-selection from the ratio $\frac{s_-}{k_-}$ and obtain

$$\tau_k = -\frac{1}{\Lambda_0} log\left(\frac{s_-}{k_-}\right) \quad [4]$$

Determining the growth time τ_s under SK-selection. Next, consider the subpopulation that is expected to grow at the same rate in either condition, which should occur when the crRNA/target DNA Hamming distance $\mathcal{H}$ is zero. Using the population fractions in the K and SK conditions for this subpopulation (labeled s_+ and k_+, respectively), we can use Eqn. 1 to find τ_s:

$$\tau_s = \frac{1}{\Lambda_0} log\left(\frac{s_+}{k_+}\right) + \tau_k \quad [5]$$

Here, we assume that the growth rate under SK conditions Λ_s^i is equal to the growth rate under K conditions Λ_0 for all matched combinations –i.e. for crRNA-target DNA combinations Hamming distance $\mathcal{H}$ equal to zero.

Determining the growth rate Λ_s under SK-selection. Having derived expressions for τ_k and τ_s, we can expand eqn. 2 to get an expression for the relative growth rate $\lambda_s^i \equiv \frac{\Lambda_s^i}{\Lambda_0}$ in terms of the $s_\pm$ and $k_\pm$ population fractions:

$$\lambda_s^i \equiv \frac{\Lambda_s^i}{\Lambda_0} = \frac{log\left(\frac{s_i}{k_i}\right) - log\left(\frac{s_-}{k_-}\right)}{log\left(\frac{s_+}{k_+}\right) - log\left(\frac{s_-}{k_-}\right)} \quad [6]$$

Determining the growth rate Λ_t under TK-selection. Using a similar approach, we also derive an analogous expression of the growth rate Λ_t^i under TK-selection, assuming the t_+ cells grow in TK when $\mathcal{H}$=6 and the t_- cells do not grow when $\mathcal{H}$ = 0. Specifically, we get:

$$\lambda_t^i \equiv \frac{\Lambda_t^i}{\Lambda_0} = \frac{log\left(\frac{t_i}{k_i}\right) - log\left(\frac{t_-}{k_-}\right)}{log\left(\frac{t_+}{k_+}\right) - log\left(\frac{t_-}{k_-}\right)} \quad [7]$$

Determining the initial population sizes $\mathbf{x_i}$. While the size of each founding population x_i cancels out in our analysis, we still need to find its probabilistic distribution in order to compute its expected variance from the measured number of counts. Specifically, if we consider the measured population fraction under K-selection:

$$k_i = \frac{x_i exp(\Lambda_0 \tau_k)}{M_1}$$

we note that since the factor $\frac{exp(\Lambda_0 \tau_K)}{M_1}$ is common to all sequences, we can set $x_i = k_i$.

2. Thermodynamic model of CRISPR-Cas binding

Measuring dCas12a occupancy from the grand canonical ensemble. We use an auxiliary reporter system based on sacB expression to measure the effective dCas12a occupancy θ_c for the tested targets. The system is composed of a simple CRISPR interference (CRISPRi) promoter architecture, where a protospacer adjacent motif (PAM) and a target DNA overlaps with the -35 or -10 consensus site of the promoter (Fig. 2). Binding of an RNA-guide CRISPR-Cas endonuclease with deactivated nuclease sites to the promoter prohibits initiation of RNA transcription by RNA polymerase (RNAP). In this scenario, the binding energy of the CRISPR-Cas protein to its target DNA site is ϵ_c, the RNAP binding energy is ϵ_p, the PAM site binding energy is ϵ_{PAM}, and the grand partition function of this system is [1, 2]

$$\mathcal{Z} = 1 + \lambda_p e^{-\beta\epsilon_p} + \lambda_c e^{-\beta\epsilon_{PAM}} + \lambda_c e^{-\beta\epsilon_c} \quad [8]$$

where $\beta = k_B T$, μ_p and μ_c are the RNAP and dCas12a chemical potentials $\lambda_p = e^{\beta\mu_p}$ and $\lambda_c = e^{\beta\mu_c}$ are the RNAP and dCas12a fugacities.

Using $\mathcal{Z}$, we derive an expression for the fold-change FC, defined as the ratio of the average number of absorbed RNAP molecules with and without repressor molecules [1, 2], and get

$$FC \equiv \frac{\langle N_c \rangle_{R\neq 0}}{\langle N_c \rangle_{R=0}} = \frac{1}{1 + \lambda_c e^{-\beta\epsilon_c}} \quad (\lambda_p e^{-\beta\epsilon_p}, \lambda_c e^{-\beta\epsilon_{PAM}} \ll 1) \quad [9]$$

Here, we used the weak promoter limit $\lambda_p e^{-\beta\epsilon_p} \ll 1$ because the RNAP typically initiates RNA transcription immediately after binding to the promoter and does not occupy the promoter for a long time (i.e. it binds to the promoter in a manner that appears as though its binding energy is very weak). Similarly, we used a weak PAM binding limit $\lambda_c e^{-\beta\epsilon_{PAM}} \ll 1$ because we assume that the dCas12a protein typically does not remain in the PAM-bound state for a long time (approximately 0.13s according to single-molecule studies [3]) and will only transitions into a stable ternary complex if sufficient crRNA-DNA homology is found.

The average dCas12a occupancy is

$$\theta_c = \frac{\lambda_c e^{-\beta\epsilon_c}}{1 + \lambda_c e^{-\beta\epsilon_c}} \quad [10]$$

$$(\lambda_p e^{-\beta\epsilon_p}, \lambda_c e^{-\beta\epsilon_{PAM}} \ll 1) \quad [11]$$

In terms of the fold-change, θ_c becomes

$$\theta_c = 1 - FC \quad [12]$$

Starting with the fold-change under equilibrium conditions, an easy to measure quantity obtained by directly measuring levels of protein or RNA in Fig. 2B and C, we can extract the effective dCas12a occupancy. In Fig. 2C, the fold-change for a perfectly matching crRNA-DNA hybrid measured using ddPCR is 1/207, meaning that the quantity $\lambda_c e^{-\beta\epsilon_c} = 206$ and the dCas12a occupancy θ_c is 122/123 = 99.5%. In addition, dCas12a levels (101 molecules per μL of cell extract) are comparable with levels of KanR (129 per μL of cell extract), suggesting that the number of dCas12a proteins should exceed the number of DNA targets present on the pSC101 plasmid (copy number = 5) in each cell. Additional ddPCR measurements also confirm that crRNA levels (4890 per μL of cell extract) are much higher than the number of dCas12a proteins and should be present at saturating concentrations.

We can also extract the average PAM site occupancy θ_{PAM} according to

$$\theta_{PAM} = \frac{\lambda_c e^{-\beta\epsilon_{PAM}}}{\mathcal{Z}} = \frac{\lambda_c e^{-\beta\epsilon_{PAM}}}{1 + \lambda_c e^{-\beta\epsilon_c}} \quad [13]$$

It is important to note that the presence of competitor sites may sequester some of the available dCas12a proteins and effectively reduce the PAM and dCas12a occupancies. However, we find no potential PAM site and target DNA sequence within the *E. coli* genome that contains significant sequence overlap (>15bp) with any of the crRNA sequences tested. Note that equations 11 and 13 are only valid for the equilibrium measurements of GFP protein and mRNA levels presented in Fig. 2B and C.

Measuring dCas12a occupancy from transition state theory. We next describe dCas12a binding/unbinding considering the rate of transition between these two states. First, to describe how dCas12a recognizes and binds to its DNA target after the PAM attachment step, we note that the initial attachment step involves recognition of a PAM site by a crRNA-loaded dCas12a endonuclease. This recognition step depends on the specific PAM sequence, and leads to a conversion into a fully-bound state at a rate ν if sufficient crRNA-target DNA homology is found. Then, once stably bound, the only way for Cas12a to unbind its target DNA is to be destabilized by the DNA replication machinery [4], which occurs at a rate Λ.

Using this kinetic approach, the number of bound dCas12a to a DNA site with PAM binding energy ϵ_{PAM}, a (PAM)→(stable complex) transition probability given by the reconfiguration rate ν, and a DNA replication rate Λ, is given by

$$\dot{N}_c = \nu N_{PAM} - \Lambda N_c \quad [14]$$

where N_c and N_{PAM} are the number of dCas12a proteins bound to their target DNA and bound to the PAM site, respectively.

At steady state, we note that $N_{tot}\theta_{PAM} = N_{PAM}$ to derive an expression for θ_c and get

$$\theta_c = \frac{\nu}{\Lambda}\theta_{PAM} \quad [15]$$

To find θ_{PAM}, we can extract the sequence-specific binding energy $\epsilon_{PAM} = \sum \epsilon_b^i$ by keeping the same target DNA and measuring θ_c for different PAM sequences. Specifically, a PAM sequence which deviates from the canonical PAM site sequence will have a binding energy given by $\epsilon'_{PAM} = \epsilon_{canon} + \Delta\epsilon_{PAM}$, which will decrease the PAM occupancy θ_{PAM} by a factor $e^{-\beta\Delta\epsilon_{PAM}}$ in the weak PAM binding limit .

To find ν, we first assume that the (PAM)→(stable complex) transition occurs in a number of n discrete steps, and each step i can only transition to either state $i-1$ or $i+1$. In this case, the transition rate from state 1 state n is simply given by:

$$\nu = \kappa \cdot e^{-\beta \sum_i^{n-1} \Delta\epsilon_i} \quad [16]$$

where $\Delta\epsilon_i = \epsilon_{i+1} - \epsilon_i$ and κ is the basal rate at which dCas12a will bind to its target given a fully matching PAM and crRNA sequences. κ is assumed to be constant for all experimental conditions tested.

Note that Eqn. 16 does not consider the sequence of events leading to the transitions from i to n because addition is commutative, which means that this expression may not accurately describe the reconfiguration rate in the presence of multiple crRNA-target DNA mismatches. On the other hand, if only a single mismatch exists at location i, the new binding energy will be shifted by a fixed position-dependent energy $\Delta\epsilon^*$ and the new rate ν will be given by

$$\nu^* = \nu e^{-\beta\Delta\epsilon^*} \quad [17]$$

Combining equations 15 and 17, we obtain an explicit formulation of the relative binding probabilities between two targets:

$$\frac{\theta'_c}{\theta_c} = \frac{\Lambda}{\Lambda'} e^{-\beta\Delta\epsilon_{PAM}} \cdot e^{-\beta\Delta\epsilon^*} \quad [18]$$

When measured in *E. coli* bacteria, both the DNA replication rate and the thermodynamic determinants of dCas12a binding will impact θ_c'/θ_c. We describe in the next section how we untangle both effects using our massively parallel CRISPRi assay.

In mammalian cells, on the other hand, DNA replication rates are not affected by CRISPR-Cas binding (i.e. $\Lambda = \Lambda'$). In this case, only the thermodynamic determinants associated with the PAM attachment probability and the reconfiguration rate will have an impact on DNA binding probabilities. Since DNA binding is also directly related to DNA cleavage activity, the relative indel frequency between two targets is thus given by

$$\text{Cleavage activity} \propto \frac{\theta_c'}{\theta_c} = e^{-\beta \Delta\epsilon_{PAM}} \cdot e^{-\beta \Delta\epsilon^*} \quad [19]$$

Thus, knowing the basic thermodynamic determinants of dCas12a binding can help determine the relative cleavage activity between any two DNA targets.

dCas12a occupancies from growth rate. The tetA-sacB cassette is under the control of a dCas12a repressible promoter whose fold-change expression (Eqn. 9) is given by

$$FC = \frac{1}{1 + \lambda_c e^{-\beta\epsilon}} \equiv f(\epsilon)$$

where the fugacity λ_c converges to the concentration of dCas12+crRNA binary complex in the [crRNA] $\gg 1$ limit and ϵ is the effective binding energy of the stabilized dCas12+crRNA+DNA ternary complex.

We describe the kinetics of this system using this system of ODE equations:

$$\begin{cases} \dfrac{dS}{dt} & = \gamma f(\epsilon) - \Lambda S \\ \Lambda(S) & = \dfrac{\Lambda_0}{1 + \dfrac{sS(t)}{k_{1/2}}} \end{cases} \quad [20]$$

where S is the number of sacB molecules, s is the sucrose concentration, γ is the sacB production rate, and the growth rate Λ is given by sucrose-dependent Monod kinetics with $k_{1/2}$ as the half-velocity constant and maximum growth rate Λ_0.

At steady-state, we get:

$$\begin{cases} S_0 & = \dfrac{\gamma f(\epsilon)}{\Lambda} \\ \Lambda(S_0) & = \dfrac{\Lambda_0}{1 + \dfrac{sS_0}{k_{1/2}}} \end{cases} \quad [21]$$

In Fig. S2, we measure the growth rate of cells with $f(\epsilon) = 1$ (i.e. with a crRNA-target DNA Hamming distance $\mathcal{H}$ equal to 6) as a function of sucrose concentration to find that the growth rate is equal to $\Lambda_0/2$ when $s_{1/2} = 0.6\%$ sucrose. Solving for $k_{1/2}$, we obtain

$$k_{1/2} = \frac{2\gamma s_{1/2}}{\Lambda_0} \quad [22]$$

Solving for Λ, we obtain

$$\lambda = \frac{\Lambda}{\Lambda_0} = 1 - \frac{s\gamma f(\epsilon)}{k_{1/2}\Lambda_0} = 1 - \frac{1}{2}\frac{s}{s_{1/2}} f(\epsilon) \quad [23]$$

Since the experiments were performed at a sucrose concentration of $s = 4.5\%$, we can rearrange Eqn. 23 to get

$$\lambda = 1 - \alpha f(\epsilon) \quad [24]$$

where $\alpha = 4.5\%/0.6\%/2 = 3.75$.

The dCas12a occupancy θ_c is related to the fold-change and growth rate $\lambda = \dfrac{\Lambda}{\Lambda_0}$ according to

$$\theta_c = 1 - f(\epsilon) = 1 - \frac{1-\lambda}{\alpha} = \nu \frac{\theta_{PAM}}{\Lambda}. \quad [25]$$

Hence, given two arbitrary crRNA-target DNA configurations, the ratio of their dCas12a occupancies is

$$\frac{\theta_c'}{\theta_c} = \frac{\alpha - 1 + \lambda'}{\alpha - 1 + \lambda} = \frac{\lambda}{\lambda'}\frac{\nu'}{\nu}\frac{\theta'_{PAM}}{\theta_{PAM}}. \quad [26]$$

Noting that $\alpha = 3.75$, we obtain

$$e^{-\beta\Delta\epsilon_{PAM}} \cdot e^{-\beta\Delta\epsilon^*} = \frac{2.75 + \lambda'}{2.75 + \lambda}\frac{\lambda'}{\lambda} \quad [27]$$

$$\approx \frac{\lambda'}{\lambda} \quad (\text{when } \lambda \approx 1) \quad [28]$$

where $\Delta\epsilon_{PAM}$ and $\Delta\epsilon^*$ are the PAM and mismatch binding energy differences, respectively, and λ is the growth rate of a reference combination with fully matching PAM and crRNA sequences.

Eqn. 28 allows us to untangle the contribution of the thermodynamic determinants of dCas12a binding from growth-dependent effects due to tetA-sacB expression. The magnitude of this effect can directly be tuned by changing the sucrose concentration in each experiment. For instance, sucrose concentrations near $s_{1/2}$ are less sensitive to sequence-specific changes and growth measurements will correlate with sequence differences that greatly affect binding. Similarly, sucrose concentrations much larger than $s_{1/2}$ will be more apt at differentiating between binding affinities that are very close to one another –for examples between two sequences with fully-matching crRNA-target DNA sequences, which were compressed near λ=1 in all figures presented in this paper. Here, we purposely tuned the sucrose concentration (4.5%) to most sensitively respond to an intermediate level of binding that is representative of 1 or 2 crRNA-target DNA mismatches.

Note that Eqn. 28 is only valid in the regime where the ratio $(2.75 + \lambda')/(2.75 + \lambda)$ is close to 1. This means dCas12a occupancy for sequences displaying a low growth rate will fall outside the bounds for which Eqn. 28 is valid. In addition, important physiological effects (such as the amount of dCas12a protein and mRNA present inside each cell) may negatively impact binding dynamics at slower growth rates, and the relative binding energy contribution described by Eqn. 28 may not be accurate.

Nevertheless, the thermodynamic determinants of dCas12a binding (parametrized as ϵ_{PAM} and $\Delta\epsilon^*$) can be used to evaluate the relative binding rate between different DNA targets with an intermediate level of binding (less than ∼2 crRNA-target DNA mismatches).

Measuring dCas12a occupancy for TK selection. When a crRNA regulates the expression of tetA or targets an essential gene, the growth rate Λ depends on the amount of tetA proteins T and tetracycline concentration $[tet]$ in the cell. If we measure the growth rate of cells that constitutively express tetA as a function of tetracycline concentration (Fig. S2), we get that the half-max growth rate occurs for $[tet]_{1/2}$=0.14 μg/mL. This means that the experiment were carried out at a 10 μg/mL tetracycline concentration, any fold-change $f(\epsilon)$ smaller than $\frac{0.14\mu g/mL}{10\mu g/mL} = 0.014$ will prevent cells from growing. Therefore, even a partially-repressed tetA-sacB cassette will decrease the amount of tetA in the cell and can in turn drastically reduce the growth rate. We are therefore unable to detect differences in dCas12a binding by monitoring the growth rate alone, and we mainly use the TK growth rates as a means to confirm repression trends observed under SK selection.

3. Model predictions

Fitting the PAM attachment and mismatch costs.

According to Equation 19, two identical DNA targets that are flanked by different PAM sequence will have the same reconfiguration rate ν but different PAM attachment energies, which in turn will yield different growth rates under SK conditions. In our model, the PAM attachment energy is defined as $\epsilon_{PAM} = \sum_i \epsilon_b^i$, where ϵ_b^i is the specific binding energy of a base of type b=(T,C,G,A) at location i=(1..6). In Fig. 3F and G, we use SK growth rates for PAM sites of the form NNNTTV to compute the position-dependent binding energies ϵ_b^i.

First, we computed a baseline value for all $e^{-\beta\epsilon_b^i}$, the Boltzmann weight of base b=(T, C, G, A) at location i, by averaging all the growth rates of the PAM of the form NNNTTV. Specifically,

$$e^{-\beta\epsilon_b^i} = \begin{cases} e^{-\beta\epsilon_b^6} = \langle\lambda(bTTTTV)\rangle \\ e^{-\beta\epsilon_b^5} = \langle\lambda(TbTTTV)\rangle \\ e^{-\beta\epsilon_b^4} = \langle\lambda(TTbTTV)\rangle \\ e^{-\beta\epsilon_b^3} = 1 \text{ if b=T, 0 otherwise} \\ e^{-\beta\epsilon_b^2} = 1 \text{ if b=T, 0 otherwise} \\ e^{-\beta\epsilon_b^1} = \langle\lambda(NTTTTb)\rangle \end{cases} \quad [29]$$

where the brackets $\langle\cdot\rangle$ signify averages over either V=(C, G, A) or N=(T, C, G, A).

This process first yielded this unoptimized energy "matrix":

TTV PAM site	-6	-5	-4	-3	-2	-1
$e^{-\beta\epsilon_T}$	0.91	0.91	0.91	1.00	1.00	0.14
$e^{-\beta\epsilon_C}$	0.94	0.91	0.48	0.00	0.00	0.92
$e^{-\beta\epsilon_G}$	0.92	0.87	0.38	0.00	0.00	0.92
$e^{-\beta\epsilon_A}$	0.90	0.91	0.25	0.00	0.00	0.91

We then optimized the energy matrix for the TTV PAM sites by performing 1,000 optimization steps where we 1) added normally distributed noise $\mathcal{N}(\mu = 0, \sigma^2 = 0.001)$ to each value of the energy matrix, 2) used the new matrix to compute new predicted values for $\lambda(PAM)$ for the TTV PAM sites, 4) performed a least-square fit of the predicted vs. measured growth rate values, and 5) updated the value of the energy matrix only if the least-square fit was smaller than in the previous iteration. The optimized matrix is shown in Fig. 3G.

A similar procedure was followed to extract an optimized energy matrix for the combined TTTV and TTTTV PAM data, obtaining

(T)TTTV PAM site	-6	-5	-4	-3	-2	-1
$e^{-\beta\epsilon_T}$	0.96	1.00	1.03	1.00	1.00	0.23
$e^{-\beta\epsilon_C}$	1.10	1.25	0.41	0.00	0.00	0.99
$e^{-\beta\epsilon_G}$	0.96	0.98	0.64	0.00	0.00	0.99
$e^{-\beta\epsilon_A}$	0.97	0.98	0.29	0.00	0.00	0.99

The predicted values for the TTV PAM sited were computed from the energy matrix shown in Fig. 3G and those of the form (T)TTTV were computed from the *(T)TTTV PAM site* matrix above. The combined model has a predicted-measured Person correlation of 0.943.

We performed the same procedure to generate the mismatch energy matrix in Fig. 4E. In this case, we first computed the baseline position-dependent energy costs $e^{-\beta\Delta\epsilon_i}$ from the values of the growth rate for individual mutations (Fig. 4C). This unoptimized energy matrix is given by

Position	1	2	3	4	5	6	7	8	9	10
$e^{-\beta\Delta\epsilon_i^*}$	0.39	0.41	0.28	0.26	0.78	41	0.25	0.61	0.52	0.25
Position	11	12	13	14	15	16	17	18	19	20
$e^{-\beta\Delta\epsilon_i^*}$	0.36	0.24	0.17	0.18	0.23	0.65	0.75	0.83	1.05	1.07

We used this initial set of values for $e^{-\beta\Delta\epsilon_i^*}$ to evaluate the accuracy of the model prediction in computing the growth rate for crRNA that contain two mismatches from $\lambda(2MM) = e^{-\beta\sum_i \epsilon_i^*}$. In this case, the Pearson correlation between the predicted and measured growth rate values is 0.769. By subjecting the position-dependent energy matrix to the fitting/optimization procedure described above for 10,000 steps, we obtained the set of values for $e^{-\beta\Delta\epsilon_i^*}$ shown in Fig. 4F, which yields a Pearson correlation of 0.869.

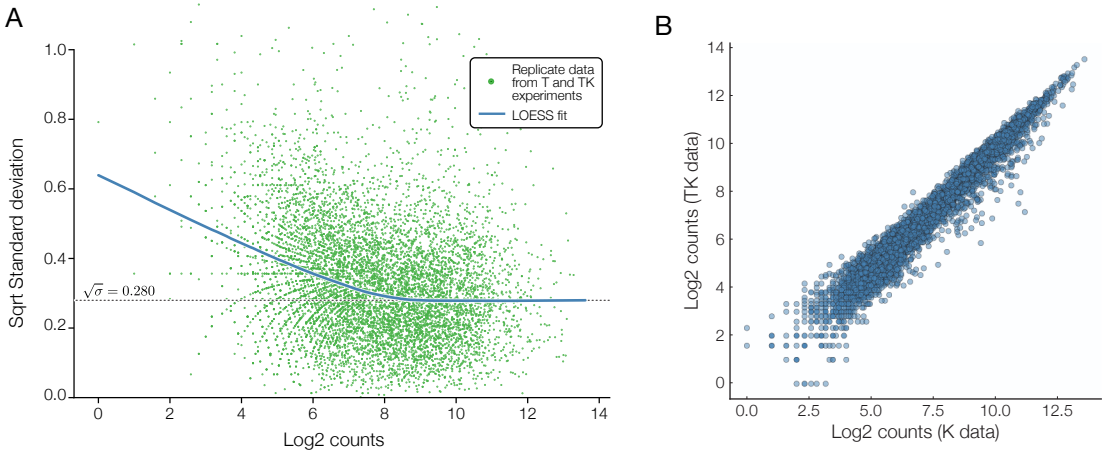

Fig. S1. Error calculations. A) Error bars are calculated using a LOESS fit [5] of the mean/variance relationship between experimental replicates of the fold change, inspired by the error estimation in [6]. In this figure, the standard deviation is computed from independent replicates sourced from the K and TK selection conditions with 6 mismatches in the seed sequence. B) Comparing the raw counts from the K and TK independent replicates.

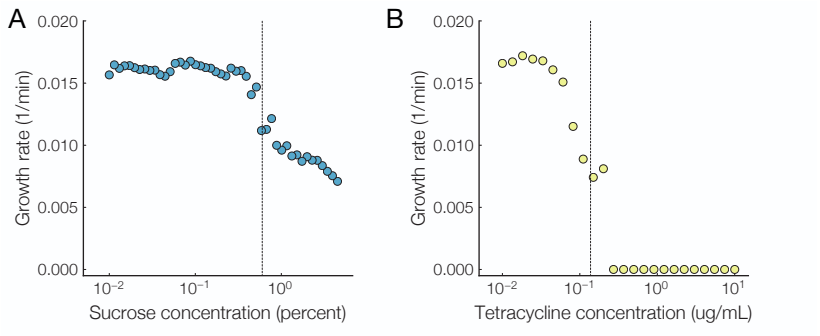

Fig. S2. Growth rate under sucrose and tetracycline. A) Sucrose-dependent growth rate for cells that fully express the tetA-sacB cassette. Transition between growth/no-growth occurs at 0.6%. B) Tetracycline-dependent growth rate for cells that fully express the tetA-sacB cassette. Transition between no-growth/growth occurs at 0.14 μg/mL.

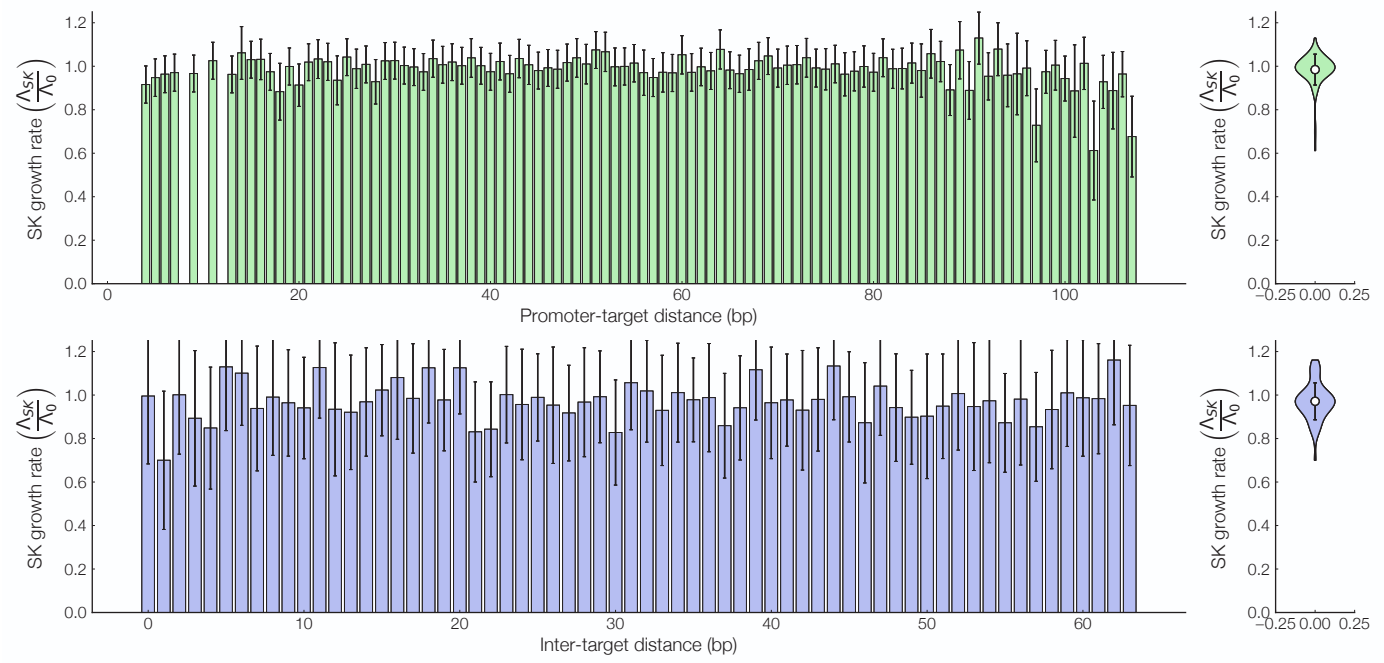

Fig. S3. Promoter-target and intra-target distances. Growth rate in SK conditions for inverter constructs which contain (A) a single target or (B) two targets located after the +1 location of the output promoter. Violin plots of the distribution of all constructs is shown on the left. Error bars = std. dev.

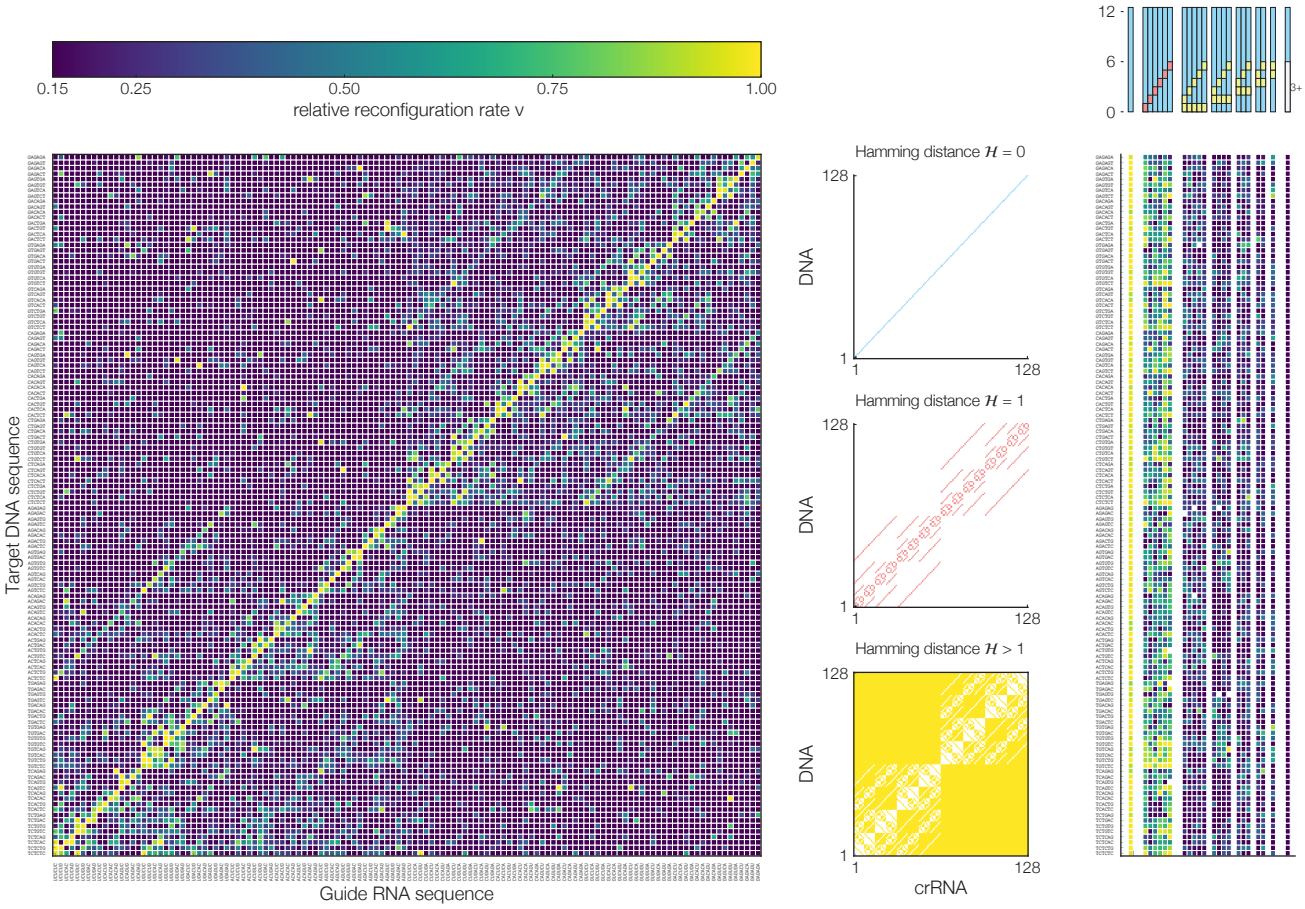

Fig. S4. Cross-talk map target DNA dependent mismatch map of the relative reconfiguration rate under 4.5% sucrose (SK) conditions for (SWSWSW+WSWSWS)x(SWSWSW+WSWSWS) seed constructs (bases 1-6).

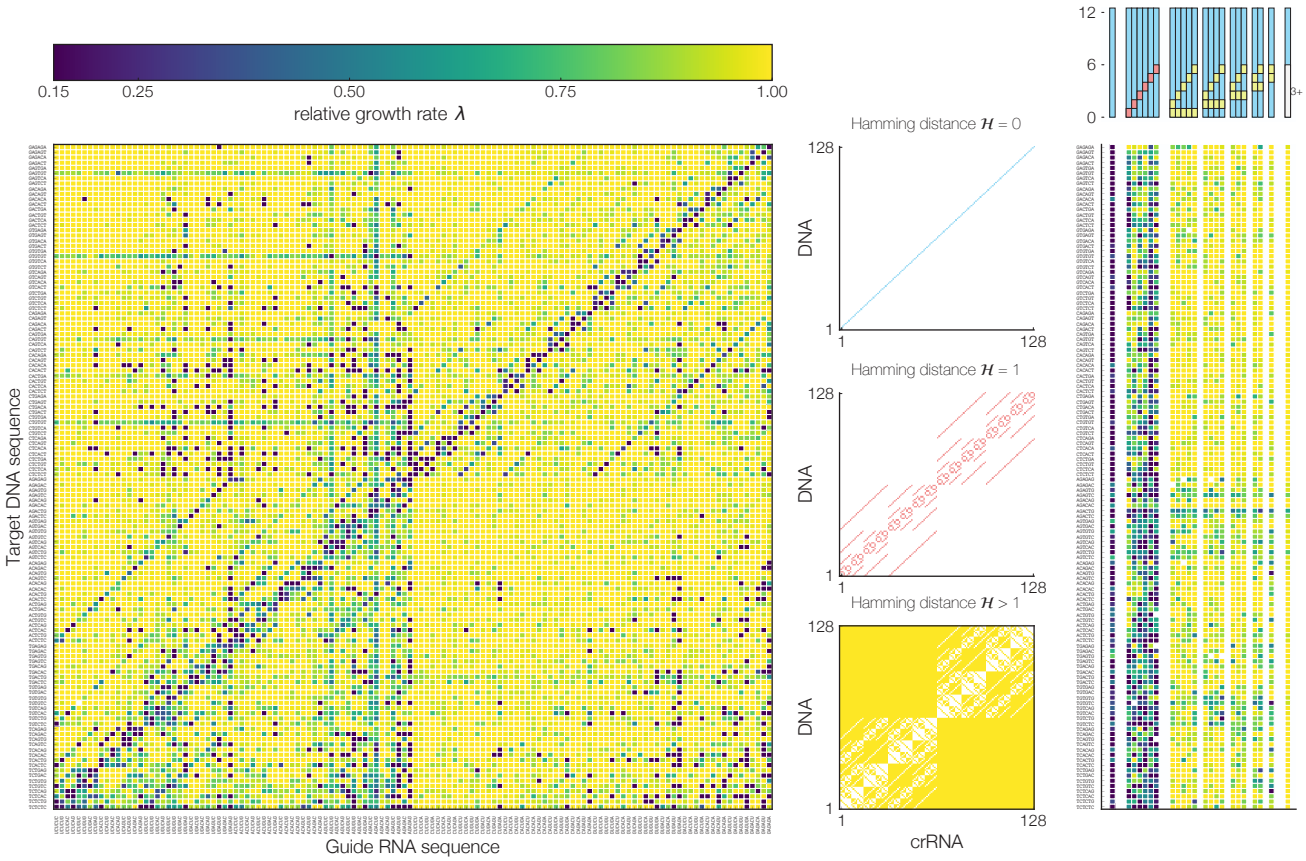

Fig. S5. Cross-talk map target DNA dependent mismatch map of the relative growth rate under 10 μg/mL tetracycline (TK) conditions for (SWSWSW+WSWSWS)x(SWSWSW+WSWSWS) seed constructs (bases 1-6).

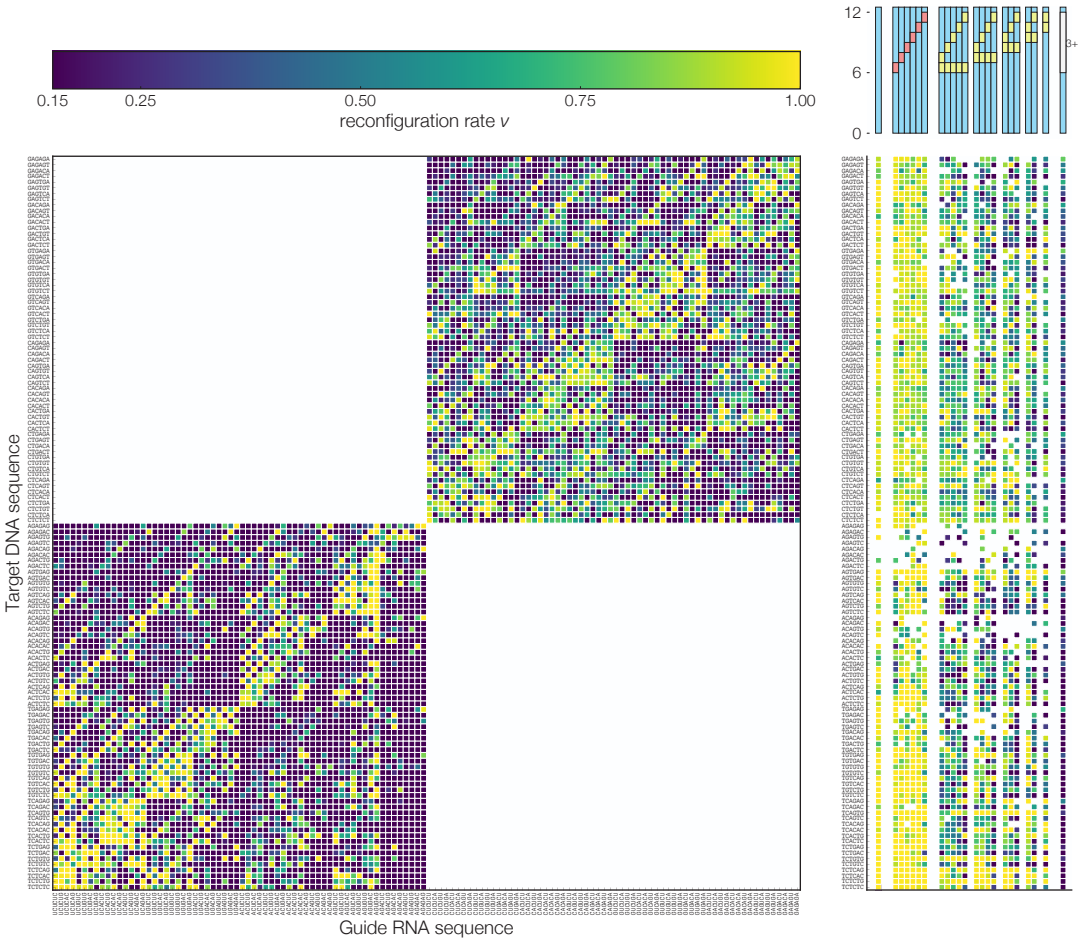

Fig. S6. Cross-talk map target DNA dependent mismatch map of the growth rate under 4.5% sucrose (SK) conditions for (SWSWSW)x(SWSWSW) + (WSWSWS)x(WSWSWS) distal constructs (bases 7-12).

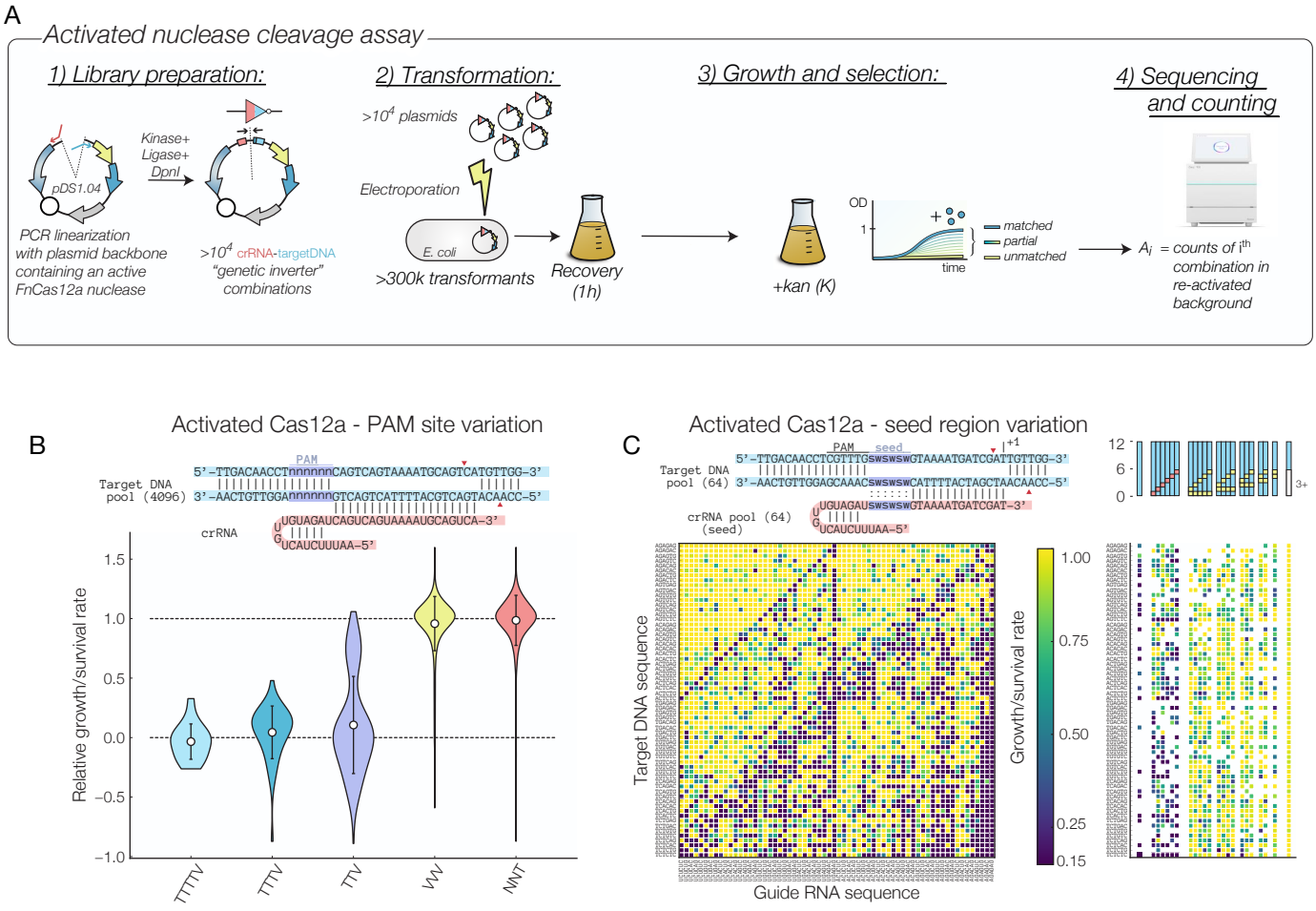

Fig. S7. A) Experimental workflow for experiment with activated FnCas12a nuclease sites. B) Relative growth/survival rates for active FnCas12a nuclease that targets a DNA sequence with a 5'-NNNNNN-3' PAM site located at the promoter's -19 position. C) Cross-talk map target DNA dependent mismatch map of the relative density of (SWSWSW+WSWSWS) seed constructs for FnCas12a with active nuclease sites.

SI Dataset S1 (Supplementary-Table-1.xlsx)
DNA primer and plasmid sequences used in this study.

SI Dataset S2 (https://github.com/lambert-lab/Massively-parallel-dCas12a-assays)
This archive contains all the scripts needed to process the data presented in the manuscript.

SI Dataset S3 (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA549693)
Repository for the raw fastq files from sequencing and data generated during this study.

SI References

[1] Weinert, F. M., Brewster, R. C., Rydenfelt, M., Phillips, R. & Kegel, W. K. Scaling of Gene Expression with Transcription-Factor Fugacity. *Physical Review Letters* **113**, 258101 (2014).

[2] Brewster, R. *et al.* The Transcription Factor Titration Effect Dictates Level of Gene Expression. *Cell* **156**, 1312–1323 (2014).

[3] Jeon, Y. *et al.* Direct observation of DNA target searching and cleavage by CRISPR-Cas12a. *Nature Communications* **9**, 2777 (2018).

[4] Jones, D. L. *et al.* Kinetics of dCas9 target search in Escherichia coli. *Science* **357**, 1420–1424 (2017).

[5] Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* **74**, 829–836 (1979).

[6] Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15**, R29 (2014).