

## Supporting Information

### **Rewritable Digital Data Storage in Live Cells via Engineered Control of Recombination Directionality**

Jerome Bonnet, Pakpoom Subsoontorn and Drew Endy

Department of Bioengineering, Stanford University, Stanford, CA 94305-USA

\*To whom correspondence should be addressed.

Phone: +1 (650) 799-9851

Email: [endy@stanford.edu](mailto:endy@stanford.edu)

### **Supplementary Methods**

#### **Molecular Biology**

Plasmids were constructed using standard BioBrick™ (1) or Gibson assembly (2). Coding sequences for BioBrick™ versions of Bxb1 integrase and excisionase and for LR and BP DNA register were synthesized by DNA 2.0 (Menlo Park, CA, USA). Plasmids and parts encoding PBAD/AraC (ref. 33 of main text) (BBa\_I0500), Superfolder GFP (3) (BBa\_I746916), and the PL<sub>tet</sub>-O1 (ref. 32 of main text) (BBa\_R0040), were obtained from the MIT Registry of Standard Biological Parts (<http://partsregistry.org>). For the decoupled reset circuit, Bxb1 integrase was cloned on pSB4A5 plasmid (pSC101 origin, 5-10 copies (4) while the excisionase was cloned on J64100 plasmid (regulated ColE1; 50-70 copies). Sequences are available via GENBANK accession numbers JQ929581 to JQ929585 and via the MIT Registry of Standard Biological Parts.

#### **Site specific chromosomal integration**

The DNA registers in BP and LR states were integrated into *E. coli* DH5alphaZ1 chromosome (Ref. 32 of main text) using a modified version of the CRIM system (5; St-Pierre F., Cui L., Endy D., and Shearwin K., manuscript in preparation), using phages

HK022 (Fig. 2B, Fig. S3) or Phi80 integrases integration sites (Fig. 2A, E, F, G; Fig. 3, Fig. 4, Fig S2, S9 and S11). Both integration sites are around the same chromosomal location (25 minutes (5))

### **Architecture of the DNA data register**

The DNA data register in BP and LR states consist of a constitutive promoter (BBa\_J23119) flanked by BP or LR recombination sites positioned in opposite orientation, resulting in DNA inversion when recombined (Figs 1 & 2 and Figs S1E and S1F). A Rnp T1 terminator (BBa\_J61048) was added in reverse orientation upstream of the promoter to prevent transcriptional read-through in the opposite orientation, so that in each state, only one fluorescent protein is visibly expressed. On each side of the recombination target, we cloned superfolder GFP and mKate2 (6) under translational control of measured strong RBSs (BIOFAB pilot C-dog project <http://biofab.org/data>). DNA data registers were first cloned on pSB4A5 low copy plasmid then integrated on the chromosome as described in the material and methods section.

### **Details on the set and reset circuits and alternate reset architectures**

To build the set circuit presented in Fig 2A, Bxb1 integrase was cloned downstream of the PBAD/AraC promoter (BBa\_I0500) on pSB3K1 plasmid bearing a p15A origin of replication (15-20 copies). This version of the integrase has a 6-His-tag which we found to stabilize the protein. Therefore, a weak RBS (BBa\_B0031) and a LAA ssrA tag was added to reduce the basal expression of the enzyme. For the reset circuit, PBAD controls expression of a polycistron encoding excisionase with a strong RBS designed using the RBS Calculator (7) to have a target Translation Initiation Rate (TIR) of 50000, followed by Bxb1 integrase with a GTG start codon to decrease its TIR (8) and increase the excisionase to integrase ratio. This construct was cloned in a plasmid with a regulated ColE1 origin of replication (J64100, 50-70 copies). We found that another way to match the correct excisionase to integrase ratio was to have a different copy number of each gene, by cloning them on different plasmids with different compatible origins of replication (Fig. S2A). We cloned PBAD-Bxb1 integrase (with no 6-His tag and no SsrA tag) on a pSC101 plasmid pSB4A5 (5-10 copies), and transformed it in cells containing the DNA data register in LR state along with pBad-RBS50000-Xis construct cloned in J64100 (excisionase only). We found that cells were able to flip from LR to BP with approximately 85% efficiency (Fig. S2B).

### **Cell culture and circuits operating conditions**

Plasmids were transformed in chemically competent *E. Coli* DH5alphaZ1 and plated on LB agar plates containing the appropriate antibiotics. The set circuit presented in Fig 2B was transformed in cells containing HK022 integrated BP DNA register bearing

a chloramphenicol resistance cassette. The polycistronic reset circuit, the decoupled reset circuit, and all S/R RAD modules were transformed in cells containing Phi80-integrated BP or LR DNA register bearing a kanamycin resistance cassette. Cells containing the chromosomal DNA data register were grown with an additional 5µg/ml of kanamycin or chloramphenicol depending on the integrated cassette.

### **Data collection and analysis**

Flow cytometry analysis was performed at the Stanford Shared Facs Facility (SSFF). The use of 2 channels increased the resolution of the measurements and was important to analyze the intermediate state resulting from bidirectional flipping. For figures 4B and 4C of the main text, data were gated by forward and side scatter, and forward scatter vs GFP was plotted.

### **Generation of a degenerated library of *ssrA* tagged reset integrases and library screening**

We built a screening vector (Fig. S1C) containing the excisionase fused to an AAK *Ssra* tag under the control of the arabinose promoter. A  $P_{LtetO-1}$  promoter followed by HindIII and NsiI sites allows for cloning and screening of Atc controlled set circuits. The excisionase gene is followed by AscI and the BioBrick™ suffix SpeI and PstI sites, allowing for cloning of a reset integrase.

The last residues of the *Ssra* tag of the reset integrase were randomized using a reduced 12 amino-acids alphabet (9) (ndt codons: Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, Gly), therefore reducing the library size to 144 variants with no stop codons while conserving an equal representation of each type of amino-acids. The library was built by amplifying the Bxb1 integrase gene with forward primers containing B0031 RBS and a GTG start codon and reverse primers containing the randomized *ssrA* tag, and cloning the PCR library into the screening vector in between the AscI and SpeI sites. Ligations were transformed into DH5alphaZ1 cells containing the chromosomal LR DNA register. 384 clones, corresponding to almost 95% coverage of the library (10) were inoculated in supplemented M9 in 96 deep-well plates and grown for 18H at 37C. Cells were then diluted 1:200 in M9 + 0.5% arabinose and grown for another 18H, then analyzed by flow cytometry on a LSRII cytometer equipped with a plate reader. Clones that were showing higher flipping toward the BP state were diluted 1:2000 in M9 without inducer and grown again for 18H, after what they were analyzed by flow cytometry. One of the working reset circuits isolated, termed “C1”, was chosen and combined with “G8” set generator to obtain the G8-C1 RAD module.

### **Computational design of RBSs using the RBS calculator and screening**

We designed 14 different RBSs, using the RBS calculator (7) (<https://salis.psu.edu/software/>), with Translation Initiation Rate (TIR) values ranging from 1, 5, 10, 20, 50, 100 and 200. We designed 2 different RBSs for each TIR values. Cells were grown without or with Anhydrotetracycline to test for leakiness and capacity to be induced, respectively. The results are presented in Figure S4. DNA inversion was monitored by cloning Gemini (11) a bifunctional reporter containing the alpha-fragment and GFP, downstream of the invertible promoter. Therefore, recombinase activity can be monitored by beta-galactosidase activity. For beta-gal. plate assay, X-Gal was used at a final concentration of 70 µg/mL concentration and IPTG at a 80µM final concentration.

### **RBS randomization and set circuits screening**

The Bxb1 integrase gene was PCR amplified using primers containing a RBS with a randomized Shine-Delgarno sequence and a GTG start codon (cggtttcacacNNNNNngctagcGTG) (11), cloned into the screening vector downstream of the PL<sub>tetO-1</sub> promoter (HindIII/NsiI). Clones were inoculated in supplemented M9 in 96 deep well plates and grown for 18H at 37°C, diluted 1:200 in SM9 containing 20ng/ml Atc, and analyzed by flow cytometry. Screening 288 clones was sufficient to obtain 5 different working set circuits. We picked the circuit termed “G8”, which was a good compromise between low basal expression (and thus low spontaneous flipping) and efficient switching.

### **Mathematical modeling**

We used ordinary differential equation (ODE) to describe the dynamical behaviors of RAD modules (Fig. 1D, 3A, Fig S6, S7, S8) and S/R latches (Fig. S10). ODEs keep track of the levels of recombinases and total DNA register in each state. ODEs were numerically solved using MATLAB © ODE23 solver. We assume that all binding–unbinding reactions are at quasi-steady-states. The quasi-steady-states are updated every time step of ODEs by using MATLAB © fsolve function. For the stochastic simulation (Fig. 2D), we only keep track of the number of DNA registers in each state and fluorescent reporters. The stochastic simulation is implemented in MATLAB using Gillespie algorithm.

## 4. Kinetic model of RAD modules

### Integrase-excisionase based RAD module model

The model consists of three components: integrase (I), excisionase (X) and DNA register (D). BP to LR recombination is catalyzed by an integrase tetramer (a complex of integrase dimer binding to attB and attP sites). LR to BP recombination is catalyzed by an integrase-excisionase complex. integrase-excisionase stoichiometry in the complex for Bxb1 is unknown but is assumed to be 1:1. As shown schematically in Fig 1C, the dynamics of total register in an LR state,  $D_{LRtot}$ , can be written as:

$$\frac{d[D_{LRtot}]}{dt} = -k_c[D_{LR}I_4X_4] + k_c[D_{BP}I_4],$$

when  $k_c$  is an inversion rate constant.

Previous biochemical studies also indicated that: 1) Bxb1 integrase can form a dimer in solution, 2) integrase-excisionase cannot form complex in solution, 3) integrase dimer and integrase excisionase complex can bind to all four possible recombination sites (attB, attP, attL and attR). We assume that all complexes are at quasi-steady-state relative to integrase-excisionase production, degradation and DNA recombination. The quasi-steady-state concentration of each complex can be written as:

$$[DI_2X_m] = \frac{[D][I]^2[X]^m}{K_{di}K_iK_{dix}^m}, m = 0, 1 \text{ or } 2,$$

$$[DI_4X_m] = \frac{[D][I]^4[X]^m}{K_{di}^2K_i^2K_{dix}^m}, m = 0, 1, 2, 3 \text{ or } 4$$

when  $K_i$ ,  $K_{di}$  and  $K_{dix}$  are dissociation equilibrium constants of integrase-integrase dimer, of integrase dimer-recombination site complex and of integrase-excision complex on a recombination site, respectively.

Note that in the default model we assume that the  $[DI_4X]$ ,  $[DI_4X_2]$  and  $[DI_4X_3]$  are inactive and that integrase-excisionase cannot form a complex without an att site. For the model used for generating Figure S8A-vi, we assume that  $[DI_4X]$ ,  $[DI_4X_2]$  and  $[DI_4X_3]$  can undergo bidirectional inversion. In other words,

$$\frac{d[D_{LRtot}]}{dt} = -k_c[D_{LR}I_4X_4] + k_c[D_{BP}I_4] - k_c[D_{LR}I_4X_m] + k_c[D_{BP}I_4X_m], \quad m = 1, 2 \text{ or } 3.$$

For the model used for generating Fig. S8A-vii, we assume that there also exist the following integrase-excisionase complexes in cytosol,

$$[I_2 X_m] = \frac{[I]^2 [X]^m}{K_i K_{dix}^m}, m = 1 \text{ or } 2.$$

### Dual recombinase RAD module model

The dual recombinase RAD module model (Fig. S10A) consists of three components: recombinase-1 (R1), recombinase-2 (R2) and DNA register (D). BP to LR and LR to BP recombination are catalyzed by R1 tetramer and R2 tetramer, respectively. We assume that R1 and R2 bind to an att site as dimer; only one dimer can bind to each att site at a time; R1<sub>2</sub> and R2<sub>2</sub> can bind to both states of the register. Putting together, the system is governed by the following system of equations.

$$\frac{d[D_{LRtot}]}{dt} = -k_c [D_{LR}(R1)_4] + k_c [D_{BP}(R2)_4],$$

$$[D(R1)_2] = \frac{[D][R1]^2}{K_{di} K_i}, \quad [D(R2)_2] = \frac{[D][R2]^2}{K_{di} K_i},$$

$$[D(R1)_4] = \frac{[D][R1]^4}{K_{di}^2 K_i^2}, \quad [D(R2)_4] = \frac{[D][R2]^4}{K_{di}^2 K_i^2},$$

$$[D(R1)_2(R2)_2] = \frac{[D][R1]^2 [R2]^2}{K_{di}^2 K_i^2},$$

where  $k_c$  is an inversion rate constant and  $K_i$ , and  $K_{di}$  are dissociation equilibrium constants of recombinase dimer and recombinase dimer-recombination site, respectively.

### Mutual-inhibition S/R latch model

Binary state could be stored epigenetically using a bistable gene regulator network, for example, a system of two mutually repressing genes (Ref. 2 of the main text). The network can be set and reset simply by adding external inducers (IPTG, aTc, heat shock, etc.) that can inactivate one of the two repressors, allowing the other repressor to express.

In order to make this system modular with respect to inputs, i.e., capable of storing arbitrary transcriptional input signal, one could have an extra copy of each repressor gene driven by external transcriptional input signal (Fig S10B). This way, a pulse of generic transcriptional signal can switch the system state by temporarily increasing the production of one repressor over the other. Such input modularization has been demonstrated, albeit for only a set circuit (12).

The mutual inhibition S/R latch model presented here ((Fig S10B) consists of repressor R1 and R2 mutually repressing each other expression and an extra copy of R1 and R2 driven by set input and reset input, respectively. Assuming that repressors bind to their cognate operator sites as tetramer (as we assume for recombinases), the dynamics of R1 and R2 concentrations can be written as:

$$\frac{d[R1]}{dt} = \alpha_{set} + \frac{\alpha K_d^4}{K_d^4 + [R2]^4} - \gamma [R1],$$

$$\frac{d[R2]}{dt} = \alpha_{reset} + \frac{\alpha K_d^4}{K_d^4 + [R1]^4} - \gamma [R2],$$

where  $\alpha_{set}$  and  $\alpha_{reset}$  are production rate of R1 and R2 from input set and input reset, respectively;  $\alpha$  is the maximal production rate of R1 and R2 from their repressible promoters.  $K_d$  is a dissociation equilibrium constant between each repressor and its cognate repressible promoter.  $\gamma$  is a degradation rate of each repressor.

### Parameter non-dimensionalization and efficiency measurements

While exact kinetic parameters are unknown, we can still understand the general features of the latch behaviors by nondimensionalizing all concentration and time units, here, in term of  $K_i$  and  $k_c^{-1}$  respectively. Our default parameter set has  $K_{di} = K_{dix} = 1 K_i$ ; degradation rate constants of integrase and excisionase,  $\gamma_i$  and  $\gamma_x$ , equal to  $1 k_c$ ; the default expression scaling for integrase and excisionase,  $\beta_i$  and  $\beta_e$ , are equal to  $1 K_i k_c$ . Basal integrase or excisionase production rates are  $0.1 \beta_i$  and  $0.1 \beta_e$ , respectively; induced excisionase production rate during reset and induced integrase production rate during set or reset are  $10 \beta_e$  and  $10 \beta_i$ , respectively (i.e., fold change,  $F_c$ , from basal production = 100). Total DNA register concentration is 1 (in  $K_i$  unit).

For a mutual inhibition S/R latch, we have repressor binding dissociation equilibrium constant,  $K_d = 1$ ; basal and induced production of each repressor from set and reset promoter =  $0.1 K_i k_c$  and  $10 K_i k_c$ , respectively. The maximal production of each repressor from the repressible promoter is  $10 K_i k_c$ . “Set (reset) efficiency” is defined as the fraction of the register that can be switched from BP to LR (LR to BP) state after inducing with a set (reset) pulse for  $200 k_c^{-1}$  and also can hold state for at least  $200 k_c^{-1}$  after the pulse ends. “set-reset efficiency” is defined as the product of set and reset efficiency. For a mutual inhibition SR-latch, we use the ratio between  $[R1]$  and  $[R1]+[R2]$  instead of  $[BP]$  fraction.

We studied how the efficiency of the latch changes as we tune basal integrase and excisionase (or R1 and R2) concentrations while keeping the fold change constant. In Fig

3, we scale steady-state concentrations of Int and Xis by tuning  $Y_i$  and  $Y_x$ . For our default modeling parameters and conditions, scaling proportionally scaling basal and induced production also gives similar results.

### **General features of RAD module operable range and modes of failure**

One could think of a RAD module as a device with two elements: state storage and input interface elements. The state storage element allows the latch to maintain the state; this includes a DNA register of our DNA inversion RAD module or a bistable mutual inhibition circuit of a mutual inhibition S/R latch. A DNA register encodes a state in a DNA sequence which is naturally maintained and replicated inside living cells; a mutual inhibition circuit encodes a state as repressor concentration which can be maintained through a feedback loop. The input interface element allows external inputs (transcriptional signals for examples presented here) to perturb and change the state of the state storage element. This includes the integrase-excisionase genes (or dual recombinases) for our DNA inversion RAD module or extra copies of repressor genes driven by inputs, for a mutual inhibition S/R latch.

A challenge in implementing a RAD module is to properly “map” the dynamic ranges of the external input of interest, via the input interface element, to the state phase of the state storage element. In this study, such mapping is represented as the expression scaling parameter  $\beta$ . Physically,  $\beta$  is proportional to translation rate and inverse proportional to protein degradation rate. When the scaling parameter for integrase,  $\beta_i$ , is large, even low transcriptional set or reset signal is mapped to high integrase concentration (the upper part of Fig. 1D). If  $\beta_i$  is too large, even basal transcriptional signal set or reset signal results in too much integrase for the RAD module to hold the state. On the contrary, if  $\beta_i$  is too small, transcriptional signal during a set or reset will not give enough integrase for the RAD module to set or reset. The same idea holds true for a mutual inhibition S/R latch: scaling parameter for the repressor from input interface element must be low enough to allow the state storage element to remain bistable in the absence of an input pulse and high enough to cause the loss of bistability in the presence of an input pulse.

Another challenge for implementing a RAD module is to optimize two antagonizing mechanisms, the set and the reset mechanisms, within the same chassis. Optimal conditions for resetting of the integrase-excisionase based S/R latch, for example, having a stable and efficiently translating excisionase would be likely to have so high excisionase basal expression that can interfere with the setting mechanism.

In general, the size of the operable range with respect to the scaling parameter will be proportional to the fold changes between the basal input level and the input pulse. If the fold change is small, one needs to precisely match the basal input level to state



storage regime and the induced input level to the state switching regime. Thus, the operable range will be narrow (Fig. S8A, iv).

### Features of integrase-excisionase based RAD modules operable range

Let's first consider switching efficiencies from BP to LR state during a long set pulse and from LR to BP state during a long reset pulse (Fig 3A, top row). Both set and reset operable ranges have a lower bound of integrase expression level corresponding to the induced integrase level that is "enough" for efficient recombination. For a set, at the limit of low excisionase expression, the rate of BP to LR recombination is governed by the level of  $[DI_4]$ , which can be switched to LR state, relative to  $[D]$  and  $[DI_2]$  which cannot:

$$\frac{[DI_4]}{[D] + [DI_2]} = \frac{\frac{[I]^4}{K_{di}^2}}{1 + \frac{[I]^2}{K_{di}}}$$

$$[I_{tot}] - 4 * [D_{tot}] < [I] < [I_{tot}],$$

\*all concentrations are normalized with respect to  $K_i$ .

If  $[I_{tot}]$  during set is small compared to the square root of  $K_{di}$  then the ratio between  $[DI_4]$  and  $[D] + [DI_2]$  become small and thus setting is inefficient. Therefore, we would expect that the integrase expression lower bound for set to scale with a square root of  $K_{di}$ . For a reset, interestingly, the lower bound of integrase expression is somewhat lower than that for set. This is because excisionase binding could help pushing the complex equilibrium toward  $[DI_4X_4]$  complex.

Excisionase determines directionality of state switching. Too high basal excisionase expression will break a set. On the contrary, too low induce excisionase expression will break a reset. Let's consider how much excisionase expression scaling will allow both efficient set and reset. The net BP to LR and LR to BP recombination rates depend on the relative amount of the active recombination complex,  $[DI_4]$  and  $[DI_4X_4]$ , which, in turn, depends on excisionase level:

$$\frac{[DI_4]}{[DI_4X_4]} = \frac{K_{dix}^4}{[X]^4}$$

$$[X_{tot}] - 4 * [D_{tot}] < [X] < [X_{tot}]$$

If basal  $[X]$  during set is significantly larger than  $K_{dix}$ ,  $[DI_4]/[DI_4X_4]$  will become very small and set will not work. In Fig. 3A, basal excisionase level is  $0.1\beta_e$  so we would expect to see the upper bound of excisionase level for set operable range at  $0.1\beta_e \sim K_{dix}$ .

On the contrary, if induced  $[X]$  during reset is significantly smaller than,  $[DI_4]/[DI_4X_4]$  will become very large and reset will not work. In Fig. 3A, induced excisionase level is  $10\beta_e$  so we would expect to see the lower bound of excisionase level for reset operable range at  $10\beta_e \sim K_{dix}$ . Note that since excisionase can only bind to integrase on the recombination site but not free integrase, no matter how much integrase we express, the lower limit of  $[X]$  will not go below  $[X_{tot}] - 4*[D_{tot}]$ . Therefore, the upper bound of excisionase expression for a set operable range and the lower bound of excisionase expression for reset operable range will be almost independent of integrase level. However if we have large amount of  $[D_{tot}]$  or allow integrase-excisionase binding in cytosol, the  $[X]$  will also depend on how much integrase we have and this lower bound will increase as integrase level increases (Fig. S8A, v and vii).

Now, consider switching efficiency after the input pulse is gone (Fig 3A, 2<sup>nd</sup> to 4<sup>th</sup> rows). Reset operable range has an additional bound: the upper bound of integrase production which is not enough for causing spontaneous BP to LR recombination at the end of the reset pulse (Fig. S6). Reset becomes inefficient if  $[I_{tot}] > [X_{tot}]$  during the reset pulse because, at the end of the pulse, excisionase will disappear first and thus left over integrase will drive BP state register back to LR state again. Note that this upper bound is not simply a straight line  $[I_{tot}] = [X_{tot}]$ ; once there is enough basal excisionase to bind to all integrases on att sites, BP to LR switching become very inefficient, regardless of how much integrase we have in the system. Therefore, this upper bound sharply rises asymptotically toward the excisionase upper bound of the set operable range described above. When  $[D_{tot}]$  is large or Int-Xis can form complex in the cytosol, this the upper bound becomes a diagonal line  $[I_{tot}]=[X_{tot}]$ .

The set operable range after an input pulse remains almost the same as during an input pulse. The only difference is a small notch locating approximately where  $[I_{tot}]=[X_{tot}]$  line cross the excisionase upper bound of the set operable range (Fig. S8). At this point, there is enough basal integrase expression and basal excisionase expression to cause LR to BP switching in the absence of a pulse.

The sharp boundary between operable and non-operable regions of the set or the reset arises from the fact that four integrase monomers act cooperatively to catalyze BP to LR recombination and four excisionase monomers act cooperatively to enable LR to BP recombination. Cooperative effect is particularly strong because we assume inactive intermediate complex (a DNA register with integrase tetramer and 1-3 excisionase monomers can neither undergo BP to LR nor LR to BP recombination). If we instead assume that these intermediate complexes can undergo both BP to LR and LR to BP recombination, then the boundary of the operable range become smoother (Fig. S8A, vi).

### **Operable range features of S/R latches with alternative mechanisms**

Similar to integrase-excisionase RAD module based S/R latch, operable ranges of dual recombinase RAD module based S/R latch or mutual inhibition S/R latch with respect to input element expression scaling are constrained by (Fig. S10): 1) the expression scaling lower bound that is large enough to allow state change during a pulse, and 2) the expression scaling upper bound that is small enough to not allow spontaneous state switching in the absence of an input pulse. Operable range size, i.e., the distance between the lower and the upper expression scaling bounds is approximately the fold changes between the induced and the basal input levels.

Rectangular operable range shape results from the fact that the set and the reset mechanism for the dual recombinase RAD module (or for the mutual inhibition S/R latch) do not directly interact with each other and that there is no loss of operable regime due to stoichiometry mismatch. Note that for mutual inhibition S/R latch, there is a sharp transition between efficient and inefficient set or reset operable range due to bistability of the system.

### **Simulated operable range with respect to DNA register and integrase-excisionase gene copy numbers**

Simulated RAD operable range can recapitulate experimentally observed dependence between the copy number of DNA register, the copy number of integrase-excisionase genes and S/R latch efficiency (Fig S3A and S3B). Specifically, increasing the copy number of DNA register relative to the copy number of integrase excisionase genes decreases resetting efficiency. Parameter setting used in the simulation shown here is the same as that of the default parameter setting except for that excisionase production rate during reset is reduced to only half of integrase production rate. At the limit of high DNA register copy number, the amount of integrase-DNA register complexes approaches the total amount of integrase. If total integrase outnumbers total excisionase, there are too many integrase-DNA register complexes for excisionase to bind to and thus BP to LR recombination cannot be suppressed completely. At the limit of low DNA register copy number, the amount of integrase-DNA register is limited by the number of total DNA register. Thus, although total excisionase is less than total integrase, there could still be enough excisionase to bind to all integrase-DNA register complexes, allowing for complete suppression of BP to LR recombination and thus efficient resetting. If we change the model assumption to allow for integrase-excisionase complex formation in the absence of DNA register, reset efficiency is no longer sensitive to the copy number of DNA register (Fig S3C). Therefore, the fact that reset efficiency is sensitive to DNA register copy number supports a prior finding that excisionase only binds to integrase-recombination site complex but not to free integrase (Ref. 27 of main text).

Note that varying DNA register and integrase-excisionase gene copy number also has other side effects. Lowering integrase-excisionase copy number relative to DNA register copy number means that we reduce the amount of enzyme (integrase or integrase-excisionase complex) relative to the amount substrate (DNA register). Thus, state switching, both set and reset, will need more time to complete which could explain why both set and reset become inefficient (Fig S3B and S3C, lower right corner). On the contrary, increasing integrase-excisionase gene copy number results in higher basal expression levels of integrase and excisionase, leading to spontaneous state switching. Thus, state storage becomes inefficient (Fig S3B and S3C, bottom row, upper left corner).

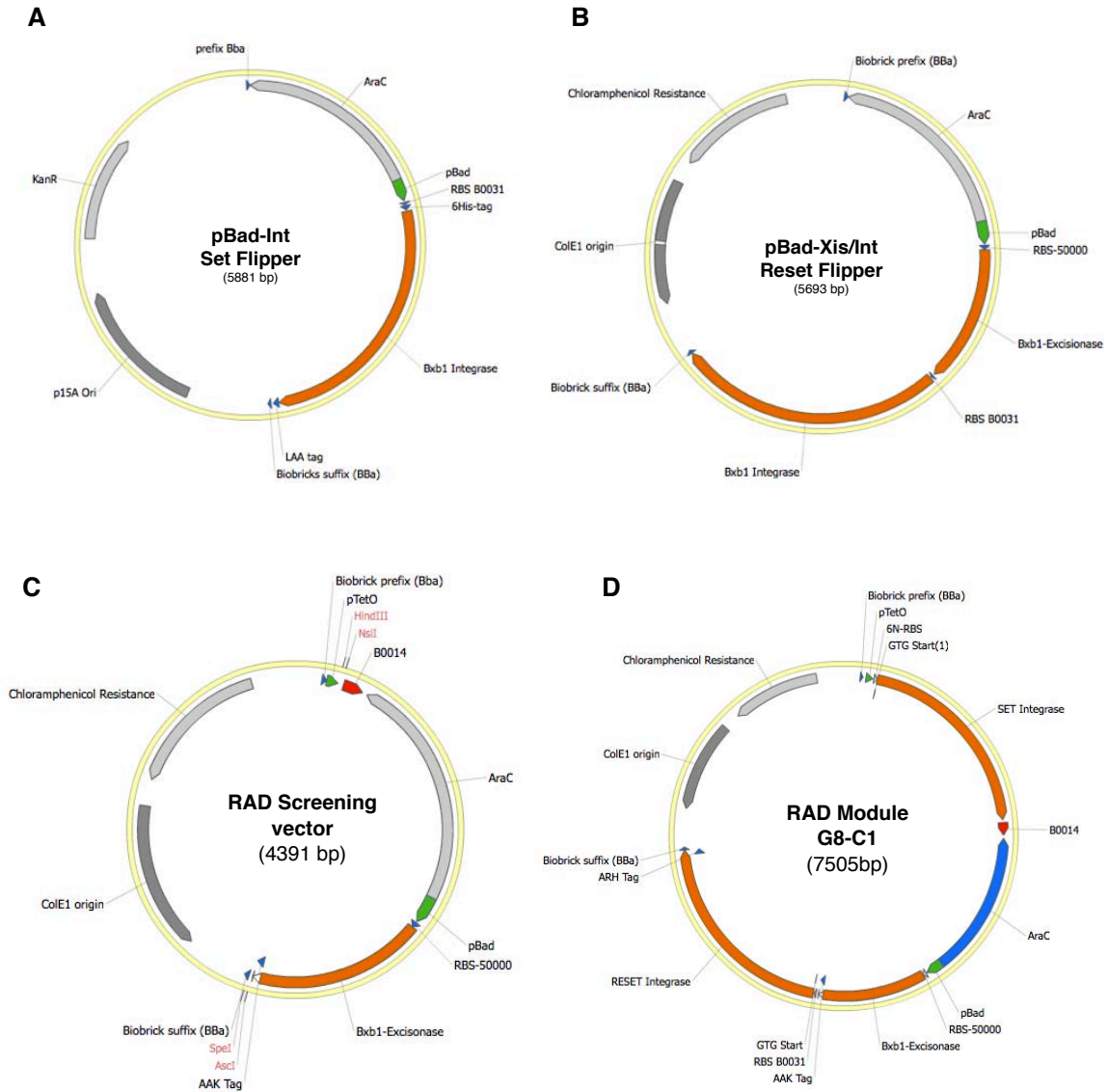
### **Stochastic simulation of bidirectional DNA inversion**

In Fig. 3B of the main text, we present the failure mode of a RAD module in which each cell in the population expressing both GFP and RFP during an input pulse and then splitting into two populations of cells expressing either GFP or RFP after the pulse. We could explain this observation using a stochastic simulation of bidirectional DNA inversion. The model consists of a single copy DNA register which can be in either state 0, expressing GFP, or state 1, expressing RFP. We simulate the scenario in which the net propensity for inverting from state 0 to state 1 and from state 1 to state 0 are equal. This scenario corresponds to the region between the set and the reset regime in Fig. 1D. We also assume that the degradation propensity of the both reporters is ten times slower than the inversion propensity (We expected that in our experimental system reporter kinetics is slow because GFP and RFP are stable). During an input pulse, bidirectional inversion occurs so fast relative reporter kinetics that cells appear to have both reporters expressing although its DNA register can be in either state 0 or state 1 at any given moment. After an input pulse, DNA inversion stops and half of the population will stochastically end up in state 0 and the other half in state 1.

## SI References

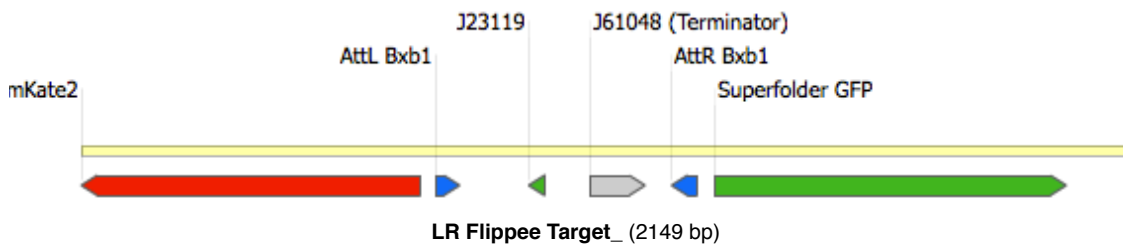
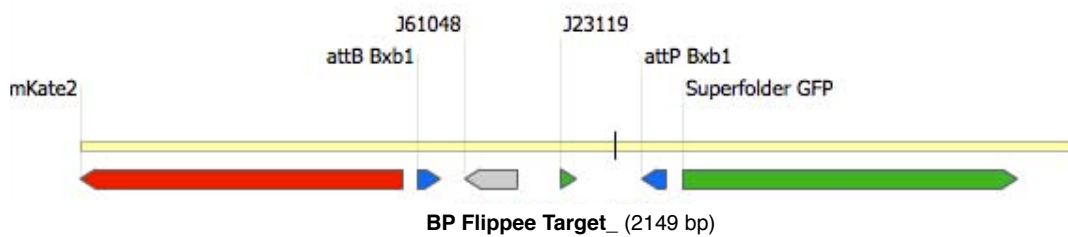
1. Knight T (2003) Idempotent Vector Design for Standard Assembly of Biobricks.
2. Gibson DG et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 6:343–345.
3. Pédelacq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2005) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* 24:79–88.
4. Shetty RP, Endy D, Knight TF (2008) Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* 2:5.
5. Haldimann A (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *Journal of Bacteriology*.
6. Shcherbo, D. et al. Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J.* 418, 567 (2009).
7. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27:946–950.
8. Barrick, D., Villanueva, K., Childs, J. & Kalil, R. Quantitative analysis of ribosome binding sites in *E. coli*. *Nucleic acids Research* 22, 1287-1295 (1994).
9. Reetz, M.T. & Wu, S. Greatly reduced amino acid alphabets in directed evolution: making the right choice for saturation mutagenesis at homologous enzyme positions. *Chem. Commun.* 5499 (2008).
10. Reetz, M.T., Kahakeaw, D. & Lohmer, R. Addressing the numbers problem in directed evolution. *ChemBioChem* 9, 1797–1804 (2008).
11. Martin, L., Che, A. & Endy, D. Gemini, a bifunctional enzymatic and fluorescent reporter of gene expression. *PLoS ONE* 4, e7569 (2009).
12. Tashiro, Y., Fukutomi, H., Terakubo, K., Saito, K. & Umeno, D. A nucleoside kinase as a dual selector for genetic switches and circuits. *Nucleic Acids Research* 39, e12 (2011).
13. Kobayashi, H. Programmable cells: interfacing natural and engineered gene networks. *Proceedings of the National Academy of Sciences* 101, 8414–8419 (2004).

## Supporting Figures

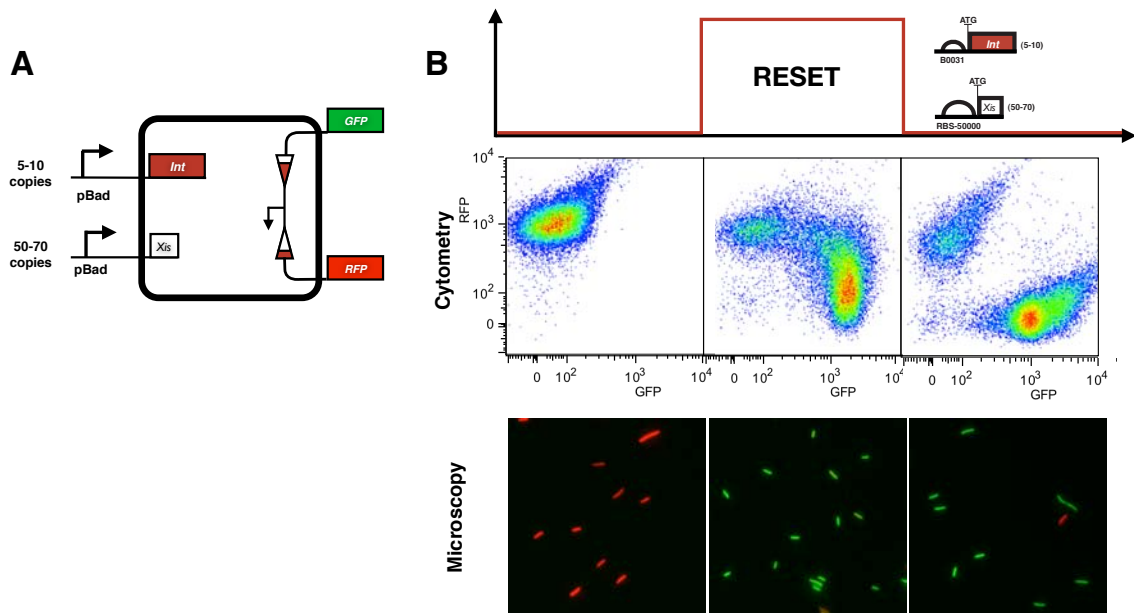


**Figure S1:** Maps of the principal constructs used in this study.

**A**, The  $P_{BAD}$ -Int set flipper where Bxb1 integrase was cloned downstream of the  $P_{BAD}$ /AraC promoter (Bba\_I0500) on pSB3K1 plasmid bearing a p15A origin of replication (15-20 copies). **B**, The  $P_{BAD}$ -Xis/Int reset flipper circuit. **C**, The screening vector and **D**, The RAD module depicted in Fig. 4A.

**E****F**

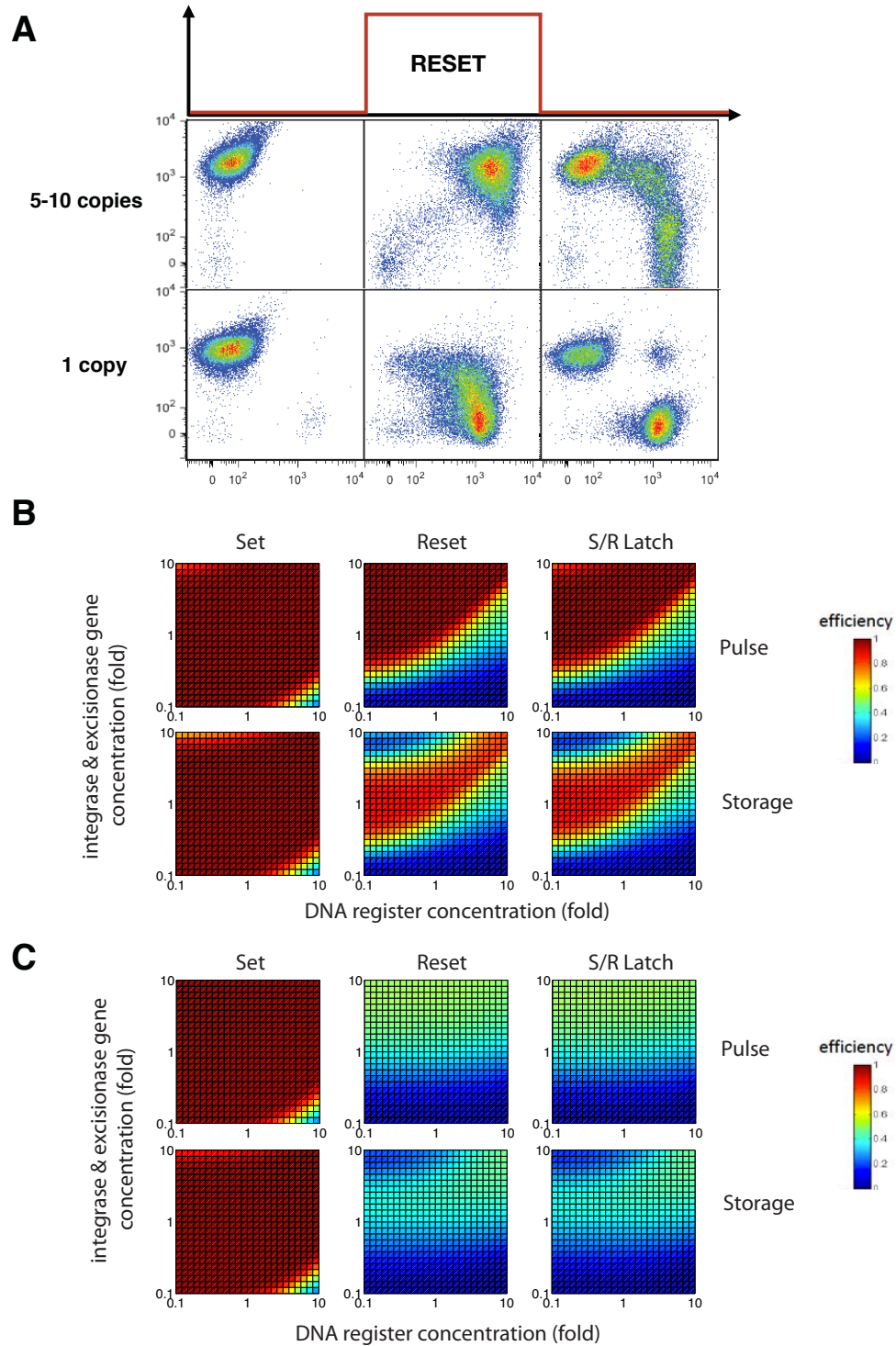
**Figure S1 (continued): E and F:** Detailed architecture of the DNA data register in state 0 (BP) and state 1 (LR) (see SI text). Maps were generated using the JBEI (Joint Bio-Energy Institute, Emeryville, CA) online vector editor (<http://j5.jbei.org/VectorEditor/VectorEditor.html>).



**Figure S2:** Alternate architecture for a reset circuit.

**A**, Schematic diagram of the decoupled reset circuit where integrase is expressed from a low-copy plasmid while excisionase is expressed from a medium-copy plasmid. **B**, Cells bearing the chromosomal LR DNA register were transformed with both plasmids encoding integrase and excisionase, pulsed with arabinose and analyzed by flow cytometry. Cells relaxed to the BP state after induction with approximately 85% efficiency.





**Figure S3:** Influence of register copy number on recombination efficiency and consequences for integrase-excisionase mechanism. **A**, Influence of copy number of the DNA register on the efficiency of integrase-excisionase mediated recombination. The bi-directional reset generator form Fig2C was transformed in cells containing the DNA data register in the LR state on a pSC101 plasmid

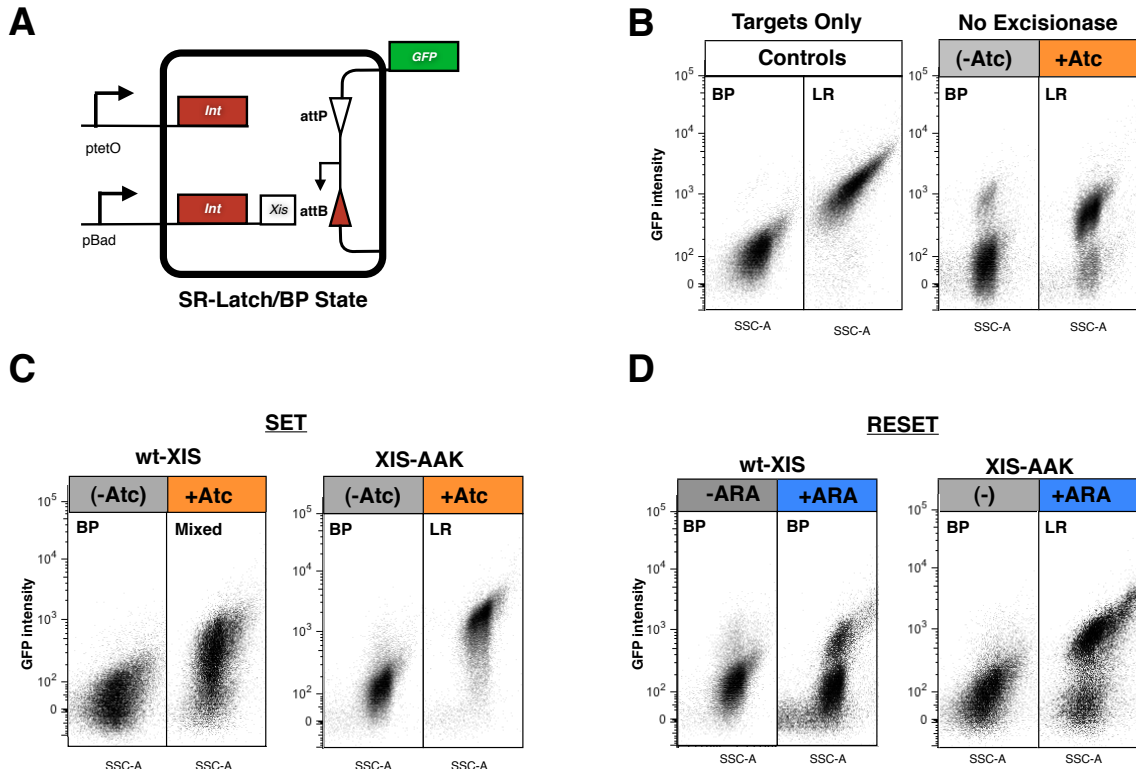
(upper panel) or integrated in the chromosome (lower panel), and cells were pulsed with arabinose. When the register was on the chromosome, cells were driven toward the BP state more efficiently during induction, and the recombination efficiency of the Int/Xis reaction for LR to BP reaction was higher after inducer removal. Part of the cells with a DNA register on a plasmid stayed in the intermediate state after inducer removal, due to the presence of various copies of the DNA register in different states within the same cell. These data support the biochemical studies suggesting that excisionase interacts preferentially with DNA-bound integrase (Ref. 27 of main text) as we assumed in our kinetic model. **B**, **C**, Simulated operable range of set, reset, and set/reset function, during and after a pulse, with respect to variations in register copy number and flipper copy number, using 2 different mechanistic assumptions for integrase-excisionase recombination. in **B**, we assume that excisionase interacts with DNA bound integrase only. in **C**, we assume that integrase and excisionase can interact in the cytosol (see SI text for details).

TIR Atc	1.1	<u>1.2</u>	5.1	<u>5.2</u>	10.1	10.2	20.1	20.2
-								
+								

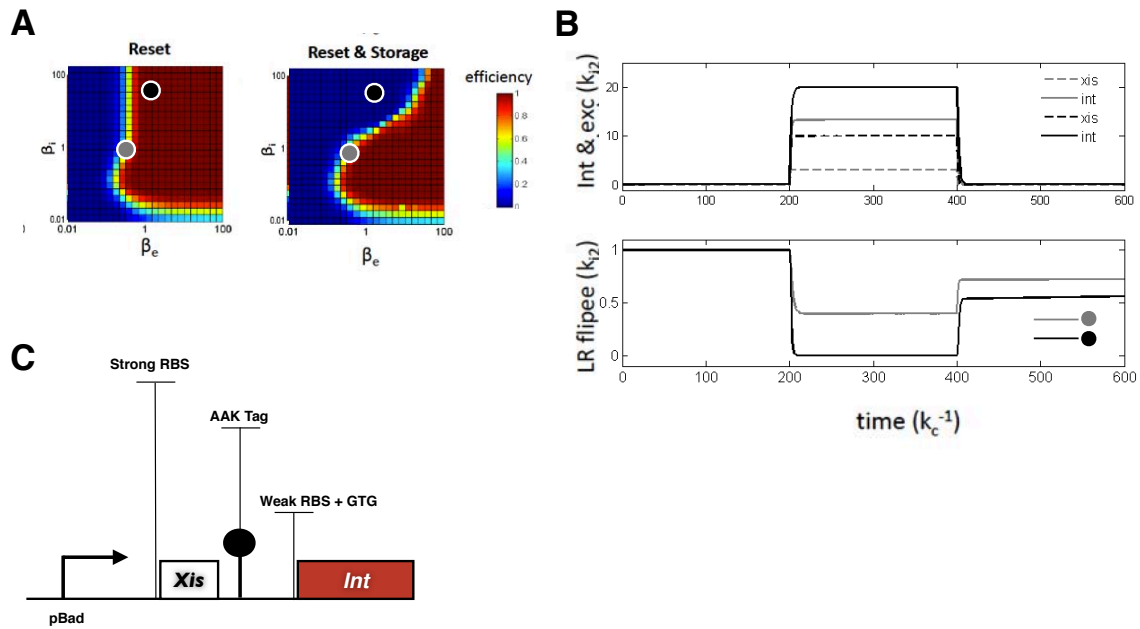
  

TIR Atc	<u>20.3</u>	50.1	50.2	<u>100.1</u>	<u>100.2</u>	200.1	<u>200.2</u>	280
-								
+								

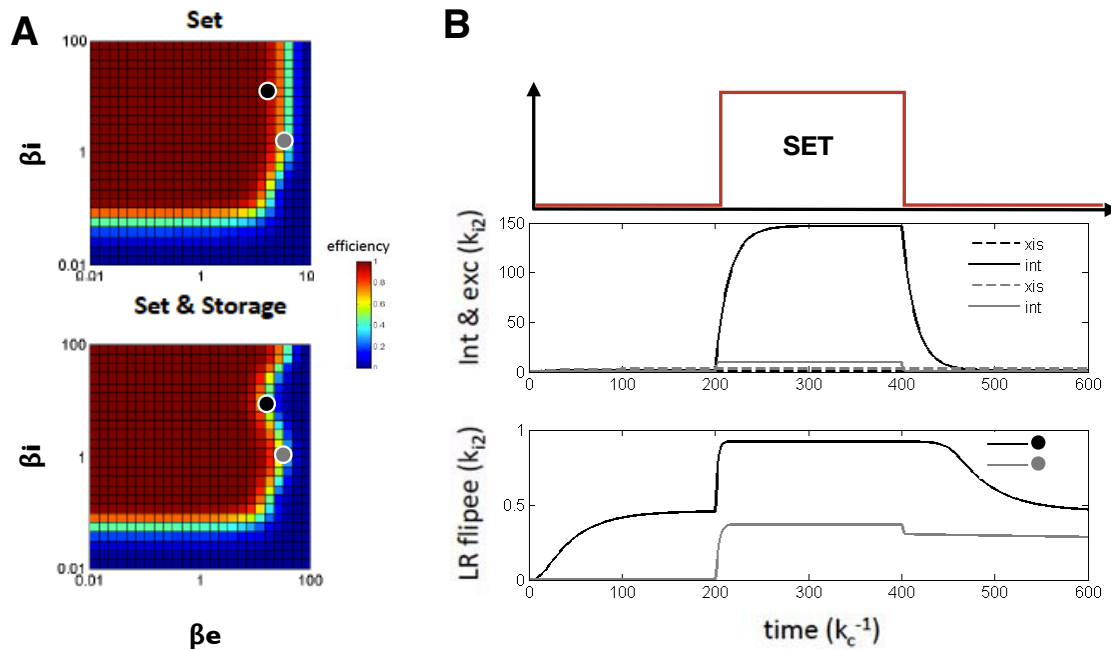
**Figure S4:** Screening of different RBS designed using the RBS calculator (7). Constructs were tested using a DNA register that expresses gemini (10) when flipped. Cells were co-transformed with the target and the different RBS variants constructs. Cells were grown with or without Anhydrotetracycline (Atc) to test for leakiness and capacity to be induced, respectively. Cultures were spotted after induction on plates supplemented with X-Gal plus IPTG and beta-galactosidase activity was observed. Target TIR (Translation Initiation Rate) values entered in the forward engineering mode of the RBS calculator are indicated on top rows. Constructs displaying the required characteristics (non-leaky and inducible) are highlighted in red. Many designs expected to exhibit the same quantitative performance produced qualitatively different results. Such differences are likely due to three coupled factors: i) because the recombinase enzyme is very active, a small level of spontaneous basal expression can switch the target DNA, ii) we used the RBS calculator to target extremely low TIRs, likely operating near a limit of the algorithm in terms of its design range, iii) regardless, as published, the calculator is expected to only have a ~47% chance of producing a specific designed RBS that produces a TIR to within a factor-of-two target range.



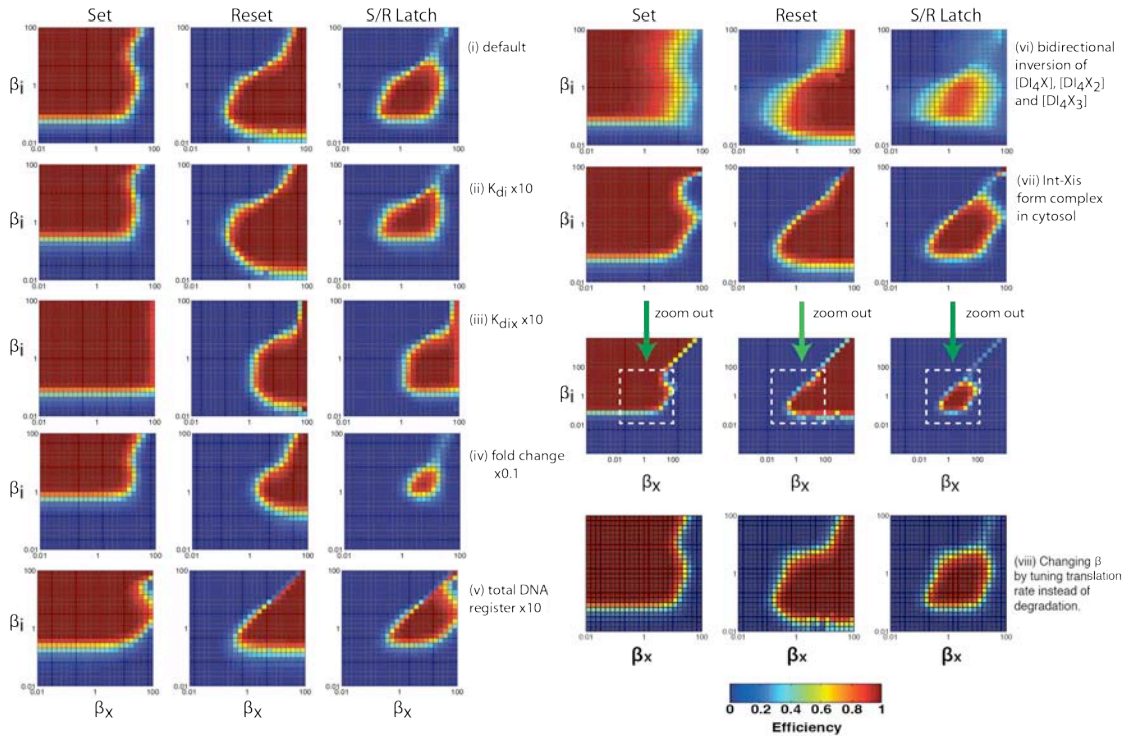
**Figure S5:** Effect of the down-regulation of excisionase on set and reset functions. **A**, schematic representation of the RAD module used for this particular experiment. Here the DNA data register has only one output, GFP, in the LR state. The  $P_{Ltet-O-1}$  promoter controls the set integrase while the  $P_{BAD}$  promoter controls a polycistron expressing integrase and excisionase. Both set and reset circuits are cloned on pSB3K1 plasmid (p15A origin, 15-20 copies). **B**, control experiments. *Left panel:* BP and LR Target constructs on pSB4A5 low-copy plasmid (5-10 copies), showing the 2 states of the system (low or high GFP for BP and LR, respectively). *Right panel:* a RAD module with no copy of the excisionase was transformed in cells containing the DNA data register in the BP state on pSB4A5 and the set generator was induced with Atc. Cells flipped to the LR state as monitored by GFP expression. **C**, reduction of interference by down regulation of excisionase basal levels. While inducing the set led to entry into an intermediate state in presence of the wild type excisionase (wt-Xis, left panel), addition of a AAK *ssrA* tag to the excisionase restored the set function (Xis-AAK, right panel). **D**, down-regulation of excisionase breaks the reset circuit. Arabinose induction of a reset generator containing wt-excisionase (wt-Xis, left panel) shows little flipping from BP to LR, due to excisionase mediated inhibition of integrase activity toward the BP DNA register. However, a reset generator in which the excisionase is down-regulated with an AAK *ssrA* tag (Xis-AAK, right panel) acts as a set generator and can flip from BP to LR due to stoichiometry mismatch and higher integrase levels.



**Figure S6:** Example of an efficient reset generator setting-back after the end of a pulse. **A**, Kinetic model based simulation of the reset efficiencies during and after the pulse for different expression scaling. A gray dot marks an integrase–excisionase expression scales that keeps the latch in an intermediate state both during and after a RESET pulse; a black dot marks an expression scale that causes the latch to revert back to LR state after a reset pulse. **B**, Time-course simulation showing resetting failures during and after a RESET pulse. Gray and black lines are simulated using integrase and excisionase expression scaling parameters as marked with gray and black dots in **(A)**. **C**, Schematic representation of the particular construct used in Fig 3C of the main text, in which excisionase has a strong RBS and AAK degradation tag while integrase has a very weak RBS (BBa\_B0033) followed by a GTG start codon. Therefore, even if the stoichiometry between the two proteins is correct during the pulse, the degradation rate of excisionase is higher than for integrase, resulting in entry into the set regime after the pulse.

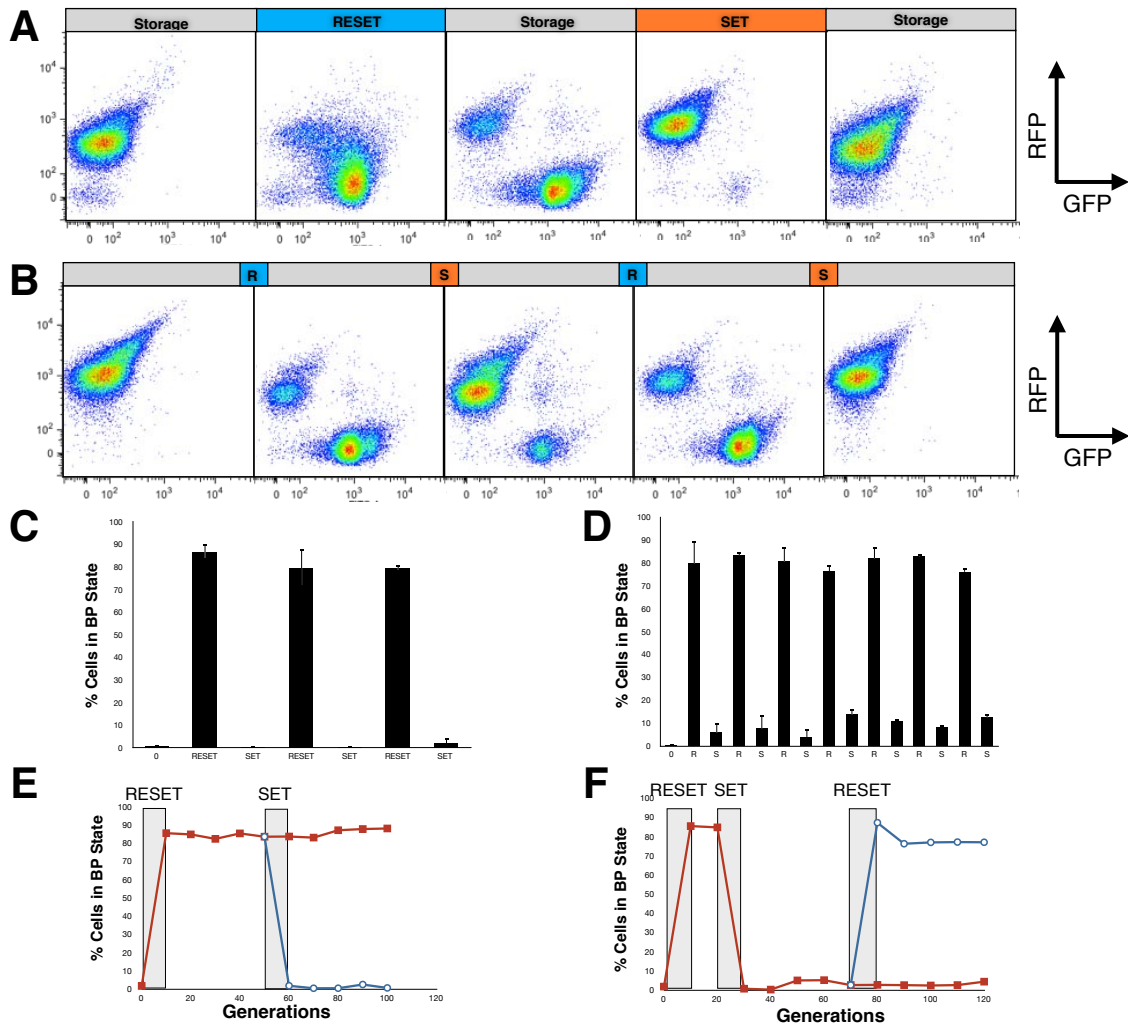


**Figure S7:** Example of a set circuit resetting back after the end of a pulse. **A**, Kinetic model based simulation of the reset efficiencies during and after the pulse for different expression scaling. A gray dot marks an integrase–excisionase expression scales that keep the latch in an intermediate state both during and after a set pulse; a black dot marks an expression scales that cause the latch to revert back to BP state after a SET pulse. **B**, Time-course simulation showing setting failures during and after a RESET pulse. Gray and black lines are simulated using integrase and excisionase expression scaling parameters as marked with gray and black dots in **(A)**. Note that for the black line simulated condition, the basal expression levels of integrase and excisionase are both high enough so that the DNA register quickly reaches an intermediate BP-LR state even before an input pulse arrives.



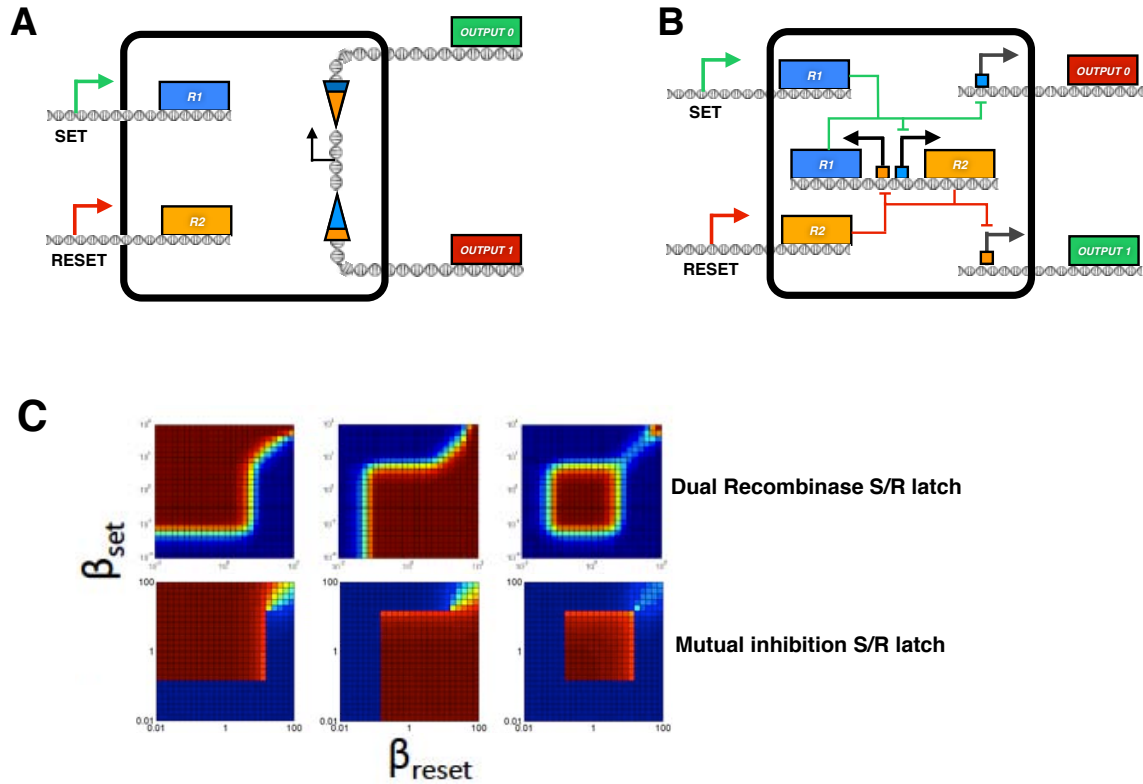
**Figure S8:** Detailed parameter sensitivity analysis of DNA inversion RAD module operable range. **(i)**, Kinetic model based simulation of the set, reset and SR-latch efficiencies after the pulse for different expression scaling and kinetic parameters. The integrase lower bound of the operable range scales with integrase-flipper dissociation constant **(ii)**, the excisionase upper bound of the set operable range and the excisionase lower bound of the reset operable range scale with integrase–excisionase dissociation constant **(iii)**, the excisionase lower bound of the reset operable range and the integrase lower bound of the set and the reset operable range scale with the fold change between induced and basal expression **(iv)** increasing the amount of DNA register changes the excisionase upper bound of set operable range and the excisionase lower bound of reset operable range. With higher DNA register concentration, these bounds become more sensitive to integrase level even at high excisionase level **(v)**, allowing incomplete integrase–excisionase complex to catalyze bidirectional recombination broadens the parameter ranges for partially switching latch **(vi)**, allowing integrase–excisionase to form complexes in the cytosol changes the set and the reset circuits operable range in a similar way to increasing DNA register concentration **(vii)**, under this assumption, the set operable range can be extended to arbitrarily high excisionase level by increasing integrase level. **(viii)**, Changing the scaling factor  $\beta$  by altering proteins translation rates instead of proteins degradation rates does not give a qualitatively different operable range in our default scenario.



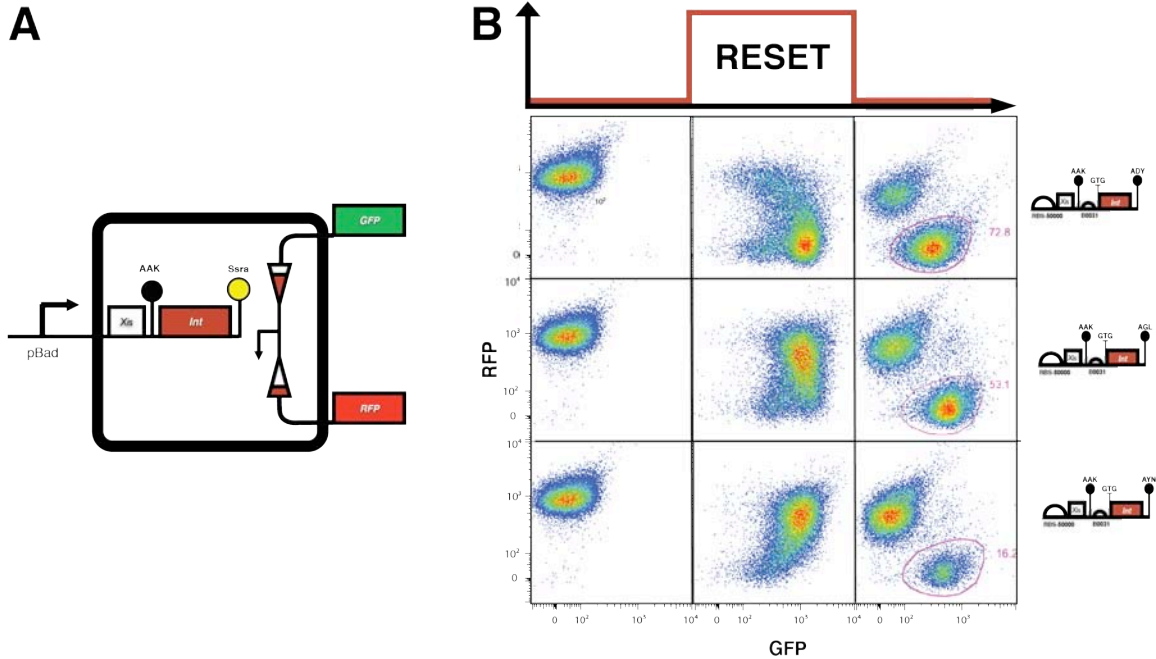


**Figure S9:** Detailed data for the RAD module operation cycles. **A** and **B**: Typical RFP/GFP plots for switching and storage cycle for long and short pulses, respectively. **C**, Quantification of switching plus storage efficiency of the RAD module for long input cycles. **D**, Quantification of switching plus storage efficiency of the RAD module for short input cycles. (**C**) and (**D**): error bars are the SD of 2 experiments performed in triplicates **E**, long term storage of the BP state in the context of the RAD module. Cells in the LR state were RESET with arabinose, switched and stored the BP state. After 40 generation, cells were SET back with Atc (blue line) or grown without inducer up to 100 generations of storage (red line). **F**, long term storage of the LR state in the context of the RAD module. After one cycle of RESET with arabinose and SET with Atc, cells were grown for 40 generations, RESET with arabinose or grown without inducer up to 100 generations. In Both (**E**) and (**F**), switching plus storage efficiency was comparable with the initial efficiencies.





**Figure S10:** Comparison of operable ranges of S/R latches using alternative mechanisms. **A**, Schematic diagram of a hypothetical S/R latch based on a DNA inversion RAD module whose DNA register can be inverted and reverted by two different recombinase, Rec1 and Rec2 respectively. **B**, Schematic diagram of a mutual inhibition S/R-latch. A pair of mutually repressed genes functions as a bistable switch; an additional copy of each gene, driven by SET or RESET inputs, is necessary to couple arbitrary transcriptional signals to the state of the bistable switch. **C**, Kinetic model based simulation of the set, reset and SR-latch. Note that the operable range for set and reset circuits are symmetric and there is no efficiency loss due to stoichiometry mis-match as in the case of integrase-excisionase based DNA inversion latch (see SI text).



**Figure S11:** Tuning the reset-specific integrase degradation rate allows for different bias in the outcome of integrase-excisionase mediated recombination. **A**, Schematic representation of the system used in this experiment. Only the integrase *ssrA* degradation tag changes. These constructs are alternate resets elements obtained during the screen for a functional reset using a destabilized excisionase (see SI text). **B**, Three resets elements using different *ssrA* sequences for the integrase display different proportions of cells in the BP state after pulse (pink: gate used to measure the percentage of cells in BP along with the actual value). Note also that: i) the intermediate state is observable during induction even at single copy number of the DNA register, as simulated by the stochastic model in Fig 2D. ii) differences in bias toward BP during induction are also observed. In front of each row is depicted a schematic representation of the actual reset element with relevant sequence information.

BP-Target

	10	20	30	40	50	60	70	80	90	100	
1	gaattcgcggccgcttctagagttatcaacgatgtcctaatttcgacggaagggtcacaataacgggctacggctacttcatggtgttcacgtaggttc										100
	*	*	*	*	*	*	*	*	*	*	
											
	prefix Bba										
	110	120	130	140	150	160	170	180	190	200	
	*	*	*	*	*	*	*	*	*	*	
101	tttatcagcttctttaatcagttctaagcgacgatcaacatagataactccaggcatTTTtaaggtTTTtagcgggTTTTTggaacgataggtgttttc										200
											
	210	220	230	240	250	260	270	280	290	300	
	*	*	*	*	*	*	*	*	*	*	
201	aggttgcagattaataatgaccgcctccaactaatttcagtgccatgtctgcgctccttctaagccaccgtccgcaggatataaggtttctgttgatgctt										300
											
	310	320	330	340	350	360	370	380	390	400	
	*	*	*	*	*	*	*	*	*	*	
301	cccaaccaagagTTTTTgcatctacggggtccgttactcgggaaatttacgccacggatTTtaacattgtaaatcagacagccatcttgtaaacctgct										400
											
	410	420	430	440	450	460	470	480	490	500	
	*	*	*	*	*	*	*	*	*	*	
401	atcttgagttgctgtaagtacaccgcatcttcataagttggttacacggtcccaggtaaagccttcaggaaaggactgtttaaaaagttaggaatgcct										500
											
	510	520	530	540	550	560	570	580	590	600	
	*	*	*	*	*	*	*	*	*	*	
501	tgggtgtgattgataaaagTTTTtagaccgcctacataaagctcgttgcgagaatgtcaaacgcaaaggcagggtccaccttcaactgcttgatgcgca										600
											
	610	620	630	640	650	660	670	680	690	700	
	*	*	*	*	*	*	*	*	*	*	
601	ttgTTTgTgTgccttcataTgTTTtTccttcaccttccggatgtgcatttGaaatgatgattgttgacagtaccttccatatataatttcatgtgcataT										700
											
	710	720	730	740	750	760	770	780	790	800	
	*	*	*	*	*	*	*	*	*	*	
701	ttctttaattaattctgacatTAATATATACCTCTTAATTTTactagTCGGCCGGCTTGTCGACGACGGCGGTCTCCGTGTCGTCAGGATCATCCGGGCC										800
											
											
	810	820	830	840	850	860	870	880	890	900	
	*	*	*	*	*	*	*	*	*	*	
801	CTCCTGAGGTGCCACCCGTGCGGCACGTTGAAGCCGATTAGTACAATAGATTTTTTTggtatagcgtcgtggacagtcattcattctttctgccctc										900
											
	910	920	930	940	950	960	970	980	990	1000	
	*	*	*	*	*	*	*	*	*	*	
901	caaaagcaaaaaccgcccgaagcgggTTTTtacgtaaatcaggtgaaactgaccgataagccggTTATTCAACCCCAAAGGTCTACTACCGTGCTAACGC										1000
											
	1010	1020	1030	1040	1050	1060	1070	1080	1090	1100	























\* \* \* \* \*  
5701 gtgaagaagctcgaccgtcttggccgcgacaccgccgacatgatccaactgataaaagagtttgatgctcaggggtgtagcggttcggtttattgacgacg 5800

5810 5820 5830 5840 5850 5860 5870 5880

\* \* \* \* \*  
5801 ggatcagtaccgacggtgatatggggcaaattggtggtcaccatcctgtcggctgtggcacaggctgaacgccggaggatcc 5881

























ColE1 origin

ColE1 origin

6310 \* 6320 \* 6330 \* 6340 \* 6350 \* 6360 \* 6370 \* 6380 \* 6390 \* 6400 \*
agccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaccaccgctggtagcgggtggtttttttggttgaagcagcagattaccgcg 6400

ColE1 origin

ColE1 origin

6410 \* 6420 \* 6430 \* 6440 \* 6450 \* 6460 \* 6470 \* 6480 \* 6490 \* 6500 \*
cagaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaactcacgtaagggttttggtc atgacttgt 6500

ColE1 origin

ColE1 origin

6510 \* 6520 \* 6530 \* 6540 \* 6550 \* 6560 \* 6570 \* 6580 \* 6590 \* 6600 \*
gcttgattctcaccaataaaaaacgccggcggaaccgagcgttctgaacaaatccagatggagttctgagtcattactggatctatc aacaggagt 6600

6610 \* 6620 \* 6630 \* 6640 \* 6650 \* 6660 \* 6670 \* 6680 \* 6690 \* 6700 \*
ccaagcgagctcgatatca aattacgccccgcctgccactcatcgcagactgttgaattcattaagcattctgccacatggaagccatcacagacg 6700

Chloramphenicol Resistance

6710 \* 6720 \* 6730 \* 6740 \* 6750 \* 6760 \* 6770 \* 6780 \* 6790 \* 6800 \*
gcatgatgaacctgaatcgccagcggcatcagcaccttgcgccttgcgtataaatatttgccatggtgaaaacggggcgaagaagtgtccatattgg 6800

Chloramphenicol Resistance

6810 \* 6820 \* 6830 \* 6840 \* 6850 \* 6860 \* 6870 \* 6880 \* 6890 \* 6900 \*
ccacgtttaaatcaaaactggtgaaactcaccagggattggctgagacgaaaaacatatttcaataaacctttagggaatagccaggttttcacc 6900

Chloramphenicol Resistance

6910 \* 6920 \* 6930 \* 6940 \* 6950 \* 6960 \* 6970 \* 6980 \* 6990 \* 7000 \*
gtaacacgccacatcttgcaatataatgttagaaaactgccggaatcgtcgtggtattcactccagagcggatgaaaacgtttcagttgtcatgga 7000

Chloramphenicol Resistance

7010 \* 7020 \* 7030 \* 7040 \* 7050 \* 7060 \* 7070 \* 7080 \* 7090 \* 7100 \*
acggtgtaacaagggtgaacactatcccatatcaccagctcaccgtctttcattgccatacgaattccggatgagcattcatcagcgggcaagaatgt 7100

Chloramphenicol Resistance

7110 \* 7120 \* 7130 \* 7140 \* 7150 \* 7160 \* 7170 \* 7180 \* 7190 \* 7200 \*
gaataaaggccgataaaaacttgtgcttattttctttacggcttttaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaac 7200

Chloramphenicol Resistance

7210 \* 7220 \* 7230 \* 7240 \* 7250 \* 7260 \* 7270 \* 7280 \* 7290 \* 7300 \*
tgactgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacgggtgtatatccagtgattttttctccat tttagcttccttagctcct 7300

Chloramphenicol Resistance

7310 \* 7320 \* 7330 \* 7340 \* 7350 \* 7360 \* 7370 \* 7380 \* 7390 \* 7400 \*
gaaaatctcgataactcaaaaaatagcccgtagtgatcttatttcattatggtgaaagtggaacctcttacgtgccgatcaacgtctcattttcgcc 7400

7410 7420 7430 7440 7450 7460 7470 7480

\* \* \* \* \*

7401 agatatcgacgtctaagaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtcttcacc 7483