Controlling Genes is hard – An approach to a biological bit counter

Moritz Fuller, Manuel Walss

Semester project, Biological Computers, 2020/21

Part 1: Specification requirements

Biological systems use elaborate computing mechanisms to keep up function. Some approaches can be compared to digital computing. For a long time, there has been an urge to combine the two areas. Many experiments have already shown how we can manipulate biological systems to replicate computing mechanisms. As a proof of concept, we simulated a binary counter.

In their seminal work, Friedland et al. created a DNA invertase cascade counter (DIC) that can count several events. Counting units are set into forward orientation by recombinases. Up to three units are chained together. Invertases are recombinases that are capable to flip a DNA sequence in between two recognition sites. These sites can be concordant or vary slightly. Consequently, the process can be reversible or not. Gene expression can be turned on and off by arranging promoter regions and genes in recombination sections.

A deterministic finite state machine (DFA) processes a string of 0s and 1s. A Mealy automaton is a DFA that allows exactly one transition for each current state and input. Our molecular machine is based on this mechanism.

Firstly, we describe a single bit counter. The entire sequence represents one biological bit (BB). The module can be used as a base for multi-digit machines. The machine is built on a plasmid in an E. coli strain. We are using an invertase to control the expression of a fluorescent marker protein. That way, we create a visual interface. The state change is mediated by the same inducer molecule in both directions.

Secondly, we propose a two-bit counter. We can count four pulses of the inducer molecule. Again, the states are marked by fluorescent molecules. The system switches from showing no emission [0 0] to green [1 0] to red [0 1] to yellow [1 1]. Our method scales exponentially in its counting ability by adding more components. Like a digital binary counter, it has an overflow mechanism. After reaching its maximum capacity, it resets to the initial state and the counting starts from zero. The carriage return bit can be used as the input for the next counting unit. This way, we can chain together independent bit counters.

Part 2: High-level system description

One-bit Flp_e counter

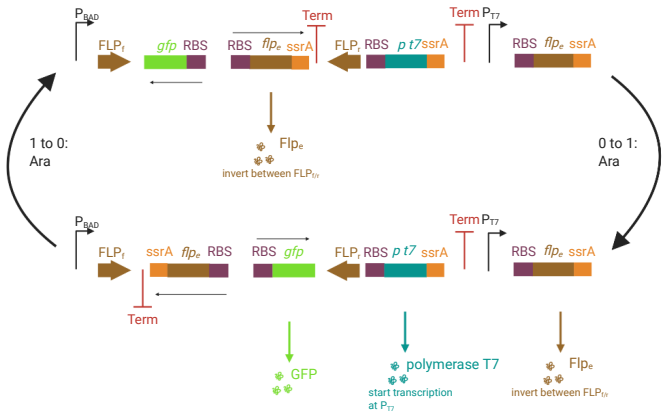


Figure 1: 1-bit counter model. State transition (0 to 1, resp. 1 to 0) is induced by addition of Arabinose (Ara). P_{BAD}: Ara-dependent promoter region. flp_e: Coding sequence of Flp_e. Flp_e: Translated recombinase. FLP_{fr}: Recognition sites of Flp_e. gfp: Coding sequence of GFP. GFP: Green fluorescent protein. p t7: Coding sequence of polymerase T7. Polymerase T7: Translated polymerase. P_{T7}: Promoter for polymerase T7. RBS: Ribosome binding site. ssrA: Degradation tag. Term: Transcription termination site.

The Flp_e invertase can invert DNA between two oppositely oriented cognate recognition sites. The same sequence can be used to both turn the expression of the fluorescence marker on and off. State transition is induced by the addition of Arabinose (Ara) to the strain. A pulse is delivered for 8 hours. This allows the RNA polymerase (RNAP) to bind at the Ara transactivated P_{BAD} promoter and start transcription. In state 0, only Flp_e is produced. The invertase inverts the sequence between the recognition sites FRT_f and FRT_r. This transition represents the step of counting from 0 to 1. After a recovery period of 8 hours, the second pulse of Ara can be delivered. Starting at the same promoter, the RNAP now transcribes the sequence for GFP and the polymerase T7. GFP acts as the marker for state 1, meaning that prior to Ara induction, the DNA counted 1 event already. The polymerase T7 binds at the promoter P_{T7} and transcribes the second sequence for Flp_e. The invertase now flips the pivotal sequence back to state 0. This step represents the overflow of our 1-bit counter.

The mechanism relies on the tempo of the inversion process. The Flp_e invertase takes around 4 hours to invert a sequence. During this time, the DNA sequence can't be transcribed. Therefore, it should only be inverted once per pulse.

Two-bit counter

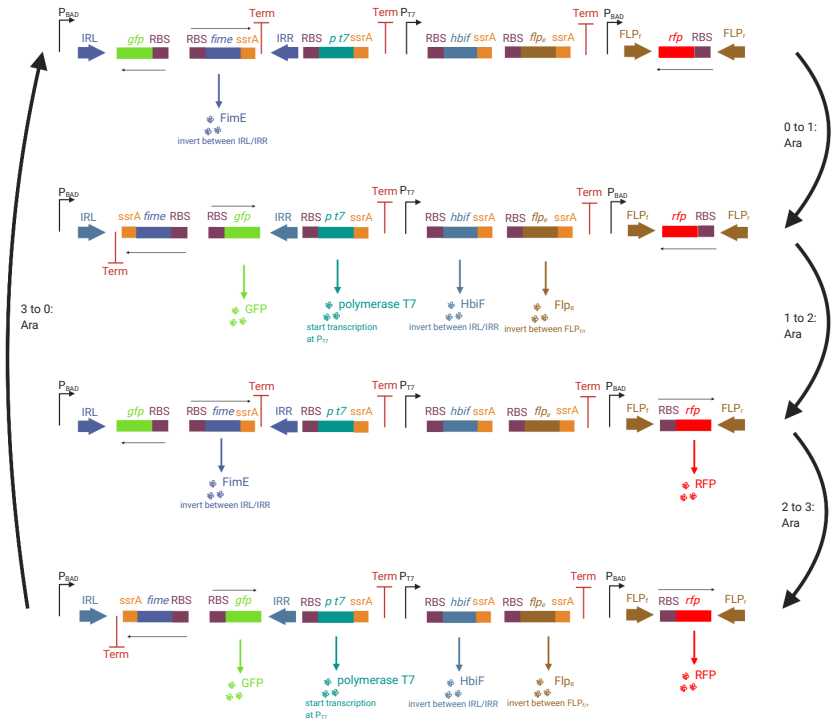


Figure 2: 2-bit counter model. State transition between states 0 to 3 is induced by the addition of Arabinose (Ara). P_{BAD}: Ara-dependent promoter region. fime: Coding region of FimE. FimE: Translated recombinase. hbif: Coding region of HbiF. HbiF: Translated recombinase. IRL/IRR: Recognition sites of FimE and HbiF. flp_e: Coding sequence of Flp_e. Flp_e: Translated recombinase. $FLP_{f/r}$: Recognition sites of Flp_e. gfp: Coding sequence of GFP. GFP: Green fluorescent protein. rfp: Coding sequence of RFP. RFP: Red fluorescent protein. p t7: Coding sequence of polymerase T7. Polymerase T7: Translated polymerase. P_{T7}: Promoter for polymerase T7. RBS: Ribosome binding site. ssrA: Degradation tag. Term: Transcription termination site.

Two BBs are chained together. The sequence starts with the P_{BAD} promoter. The first BB is made up of a switch containing the sequences for GFP (reverse) and FimE (forward) followed by a transcription termination site. The recombination sites are IRL and IRR. The sequence is inverted by FimE. Outside the SIMM, the sequence for the T7 polymerase is transcribed once the termination site inside the switch is inactive. The sequence that follows is the affiliated promoter molecule P_{T7}. Downstream are the genes of HbiF and Flp_e and a terminator site. HbiF flips the first BB back. Flp_e flips the second BB. The inversion sites are FRT_f and FRT_r. They

frame the RFP sequence. All the genes are preceded by a ribosome binding site RBS. The invertase enzymes and the T_7 polymerase are followed by ssrA tags. This sequence leads to rapid degradation, allowing closer control (Kim et al., 2000; Singh et al., 2000).

In state 0, no marker molecule is expressed [0 0]. State 1 is indicated by the green fluorescence of GFP [1 0], state 2 is indicated by RFP [0 1] and state 3 by the combined proteins [1 1]. The trigger for state transition is the addition of Ara to the system.

The first pulse triggers the expression of FimE. This recombinase inverts the first unit in between the IRL/IRR sites. The recombinase acts unidirectional. After the unit is flipped, it stays in its inverted state. The half-life for the inversion is 4 hours. Any further transcription is not to be expected. GFP is expressed with the next pulse of Ara. Transcription keeps running and the polymerase T7 is expressed, which initiates the transcription of HbiF and Flp_e. Through these molecules, the first unit is set back to the initial orientation (no GFP) and the second unit is prepared for RFP expression, with about the same half-life. With the next pulse of Ara, RFP is expressed. As in the initial state, FimE is expressed by the first sequence. The first BB is flipped again. The next pulse of Ara triggers expression of both GFP and RFP. The polymerase T7 is expressed also and consequently are HbiF and Flp_e. This way, all switches are returned to the initial state with no GFP/RFP expression. The next pulse of Ara shows no fluorescence. This shows that the counter is reset as in an overflow mechanism.

Part 3: Network diagram and model

We simulated the one bit counter approach based on the Gillespie algorithm (Gillespie, 1977). Further simulations could not be produced due to time restrictions and the fact that Gillespie does not provide a way to integrate the duration of a reaction. The full source code can be found in Fuller. As there were different opinions (Roussel, n.d.) on how to implement the propensity function, we decided to stick to the method outlined in the examples section of Gillespie algorithm. To prove the concept, we simulated the reactions for a single DNA molecule. In the extent of this report, we focussed on the key reactions. Stoichiometrically insignificant reactions were left out of the simulation for simplification, but they are still documented in the code. State transitions were simulated separately, which leaves us with four different simulations. The first simulation corresponds to the counter in its initial state, receiving the first pulse of Ara. The second simulation corresponds to the 8-hour absence of Ara after the first conduction. Those to steps are repeated for the second counting step, setting the counter to its initial state. The simulated reactions can be found in Fuller.

Kommentiert [MW1]: Brauchts das?

Kommentiert [MF2R1]: Ich würds der vollständigkeit halber drin lassen weil es im code auch nicht weiter erklärt wird

Kommentiert [MW3R1]: Aber wer hatte denn ‘different opinions’?

The initial single cell concentration was a vector containing one molecule of DNA and 10^4 molecules of Ara which corresponds to a concentration of roughly $10^{-4}\%$ after the following calculation:

$$Arabinose \rightarrow 150.13 \frac{g}{mol}$$
$$1\ mol \rightarrow 6{,}022 * 10^{23}$$
$$10^{-4}\%\ of\ 1l \rightarrow 0.001\ ml \sim 0.001\ g$$
$$150{,}13g = 6{,}022 * 10^{23}\ molecules \mid : 150130$$
$$0{,}001g = 4{,}011 * 10^{18}$$
$$V_{E.coli} = 5\mu m^3$$
$$1l\ of 10^{-4}\%\ Ara \rightarrow 4{,}011 * 10^{18}\ Ara\ molecules \mid :\ 10^{15} \mid * 5$$
$$5\mu m = 2 * 10^4$$

Equation 1: Calculation of the amount of Ara molecules per cell

The concentration vector after the simulation of the DNA in state 0 and state 1 with an Ara pulse was used as input for the respective no pulse simulations.

There are a few things that should be noted regarding the simplification of the simulation of our model:

a) We simplified the recombination process as described by Ringrose et al. to one single reaction with a rate constant that is derived from the half-life as described in Table 1 for all of our recombination.
b) We simplified the Ara binding process as described by Guzman et al. to one single reaction with a rate constant that is derived from the formula given in Table 1.
c) We assume that we can use a ssrA tag on the T7 polymerase and Flp to degrade it at the same rate as GFPssrA is degraded (Kim et al., 2000; Singh et al., 2000)
d) We are essentially not modelling the possibility of the DNA having flipped – thus being in the next state already – and Ara still being present. This would be as if the DNA were in the next state and received a fresh pulse. Ara would activate P_{BAD} promotion and thus the downstream genes would be expressed. We decided to do this as the Gillespie algorithm doesn't provide a way of modelling how long a process takes. In reality, the DNA fragment that is being inverted wouldn't be able to be transcribed anymore. As the process has a half-life of 4 hours, this has a large impact on the outcome of the simulation. For an in-vitro implementation we would need to experimentally determine the perfect timing and duration of the Ara pulse being given, just as Friedland et al. did

in their paper. This way we could wash out the Ara right after the inversion took place to stop the above-mentioned process.

Part 4: Quantitative simulations

Step 1:

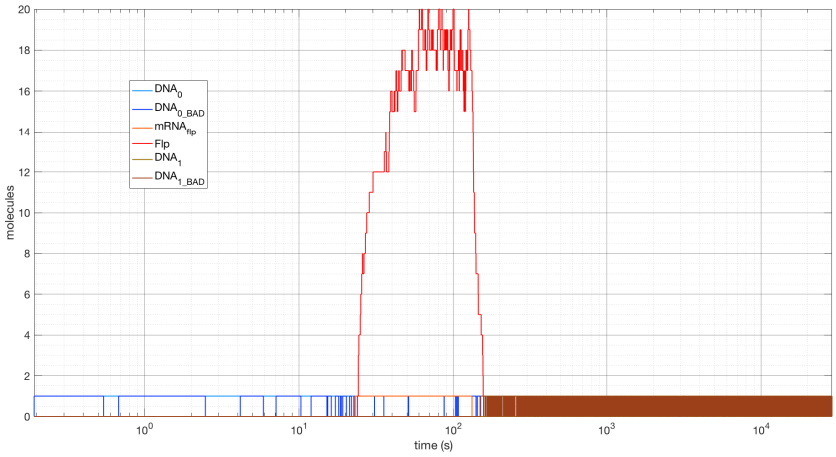
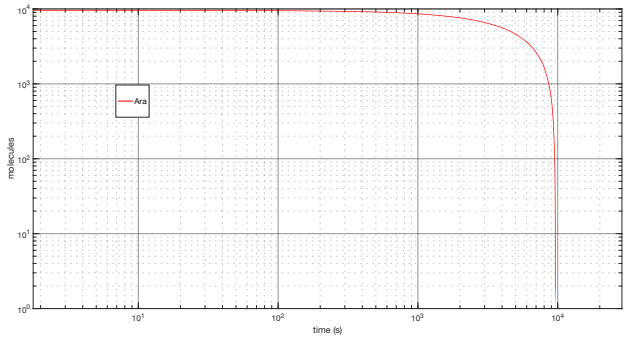


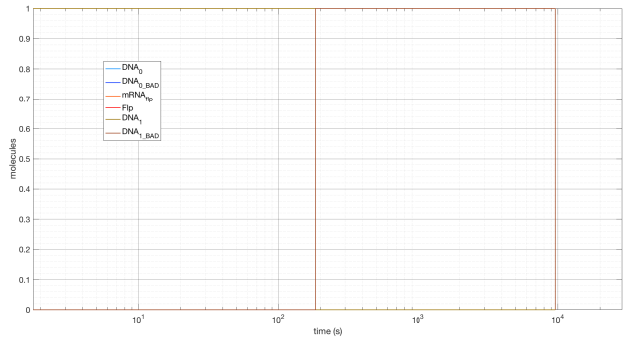
Figure 3: Simulation result of the initial cell receiving an Ara pulse. DNA can occur in state 0 (DNA_0) or 1 (DNA_1). In both states, Ara can bind to the promoter region (DNA_{x_BAD}). $mRNA_{flp}$ is transcribed only from DNA_{0_BAD} and codes for Flp.

The result of the simulation agrees with our expectation. After roughly 10^2 seconds, our system achieves stability. The DNA stays in state 1 but alternates between having the P_{BAD} activated and not activated. We assumed a steady state concentration of 100 molecules per cell but only hit about 20 in our simulation. This is due to the fact that the Flp is ssrA tagged which makes it degrade quicker.

Step 2:



Figures 4 and 6: Ara degradation and DNA behavior after ending the pulse. Fluctuation between states DNA_1 and DNA_{1_BAD} happens until the end. The final state is DNA_1.



Within 8 hours of absence all of the Ara degrades. Also, the state of the DNA doesn't change anymore besides that fact that the P_{BAD} fluctuates between being active and inactive.

Step 3:

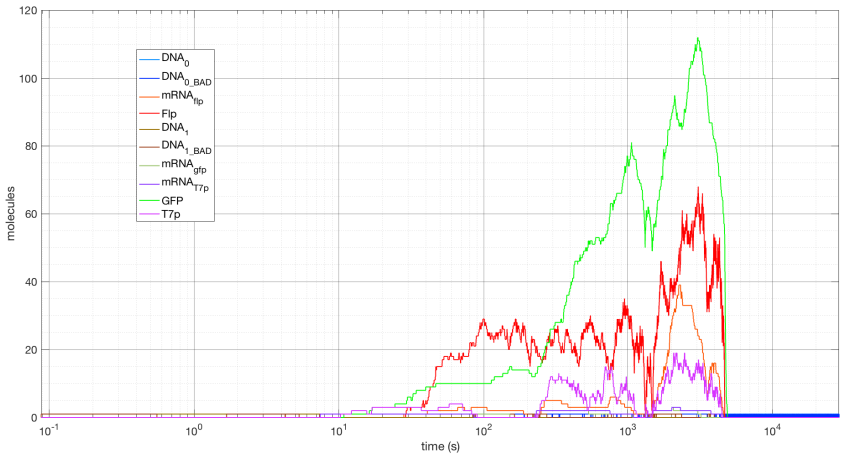
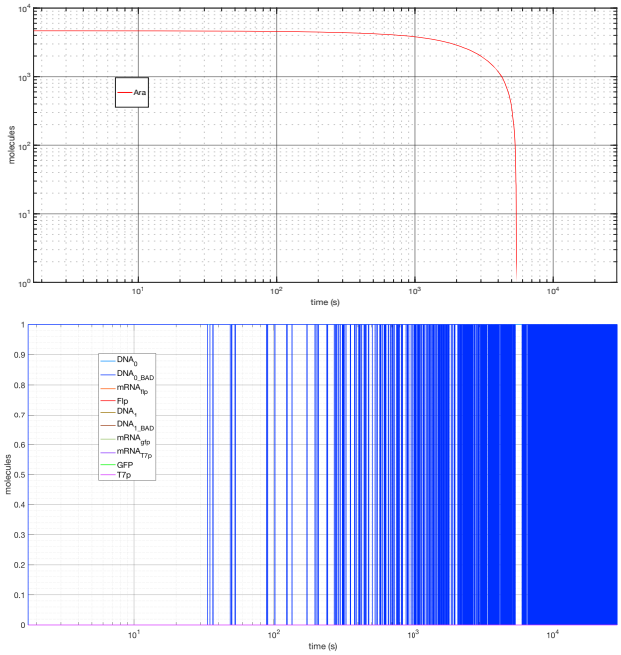


Figure 5: Simulation of the cell in state 1 and receiving an Ara pulse. DNA can occur in state 1 (DNA_1) or 0 (DNA_0). In both states, Ara can bind to the promoter region (DNA_{x_BAD}). $mRNA_{gfp/T7p}$ is transcribed only from DNA_{1_BAD} and codes for GFP and T7p. T7p transcribes $mRNA_{flp}$ independently of Ara. $mRNA_{flp}$ codes for Flp.

Again, our expectations are being met by the simulation of the second pulse of Ara. After the translation of T7 polymerase and GFP, the polymerase starts to translate Flp, which in turn inverts the DNA as shown in **Fehler! Verweisquelle konnte nicht gefunden werden.**. As expected, GFP levels meet our assumed steady-state concentration of roughly 100 molecules per cell. Flp and T7 polymerase concentrations are lower, as they are ssrA tagged and degrade more quickly.

Step 4:



Figures 8 and 9: Ara degradation and DNA behavior after ending the second pulse. Fluctuation between induced and uninduced state (DNA_0/DNA_{0_BAD}) keeps happening.

The behaviour of the system in the absence of Ara shows the same result as in step 2. The state of the DNA doesn't change anymore besides that fact that the P_{BAD} fluctuates between being active and inactive. The only difference is that now the DNA is in state 0 instead of state 1.

Part 5: Molecular implementation and feasibility assessment

The machine is constructed as a high-copy plasmid inside an *E. coli* strain, i.e., the K-12pro strain (Friedland et al., 2009). All components have been extensively studied. The genes we use are in order of appearance:

- P_{BAD}: This promoter allows RNA polymerase binding and gene transcription if Ara levels are high and glucose levels are low (Guzman et al., 1995).
- FimE/HbiF and IRL/IRR system: Fernandez-Rodriguez et al. described this switch in 2015. It allows exact control over gene expression by unidirectional inversion. The switch is stable over 14 state changes. In the initial state, the recombinase FimE inverts the sequence in between the recombination sites IRL and IRR. The sequence can only be flipped back by HbiF. The IRL/IRR sequences are conserved and allow further inversions. The half-life for the inversion is approximately 4 hours.

- GFP: The most commonly used fluorescence marker.
- RNA Polymerase T7: This is a single-subunit DNA-dependent RNA polymerase (ssRNAP). It is specific to the P_{T7} promoter. Other RNAPs of the family include RNAP T3 and SP6 (Arnold et al., 2001; Rong et al., 1998; Tabor, 1990).
- P_{T7}: The promoter region for the RNAP T7 (Arnold et al., 2001; Rong et al., 1998; Tabor, 1990).
- Flp_e system: The Flp_e recombinase inverts a sequence between the FRT_f and FRT_r sites. The inversion is possible in both directions. The half-life for the inversion is approximately 4 hours (Whiteson & Rice, 2008).
- RBS: Standard ribosome binding site (Friedland et al., 2009)
- ssrA: This tag causes rapid degradation of the mRNA (Kim et al., 2000; Singh et al., 2000).
- Ara: Induction molecule for RNAP binding at P_{BAD}. Transcription is initiated when Ara levels are high and glucose levels are low. (Guzman et al., 1995).

Bidirectional flipping could pose a problem, as the sequence could be flipped right back to the initial position within a single pulse. This problem could be solved by using sets of unidirectional recombinases for all BBs. To our knowledge, no other set than the FimE/HbiF set has been studied sufficiently though. The implementation of the mechanism would also make the molecule more complex again.

Part 6: General assessment

The system is proof that bit-counters can be realised on DNA level. They can be scaled up by adding further independent switches with similar function. There are ways to reset the system in the end and reuse it.

A medical implementation for inducible protein production can be seen in interval drug therapy. Therapeutic molecules can be delivered by the control of time.

For many industrial processes, this system is simply too unstable and slow. Unsurprisingly, with each step some leakage occurs. If a DNA molecule does not end up in the correct state, the mechanism is broken. Presumably, accuracy drops with each bit added. The systems require much further studying to the control the process. Furthermore, it is a slow process compared to what we are used to from mechanical computers.

References

Arnold, S., Siemann, M., Scharnweber, K., Werner, M., Baumann, S. & Reuss, M. (2001). Kinetic modeling and simulation of in vitro transcription by phage T7 RNA polymerase. *Biotechnology and Bioengineering*, *72* (5), 548–561. https://doi.org/https://doi.org/10.1002/1097-0290(20010305)72:5<548::AID-BIT1019>3.0.CO;2-2

Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S. & Cohen, S. N. (2002). Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences*, *99* (15), 9697–9702. National Academy of Sciences. https://doi.org/10.1073/pnas.112318199

Fernandez-Rodriguez, J., Yang, L., Gorochowski, T. E., Gordon, D. B. & Voigt, C. A. (2015). Memory and Combinatorial Logic Based on DNA Inversions: Dynamics and Evolutionary Stability. *ACS Synthetic Biology*, *4* (12), 1361–1372. https://doi.org/10.1021/acssynbio.5b00170

Friedland, A. E., Lu, T. K., Wang, X., Shi, D., Church, G. & Collins, J. J. (2009). Synthetic Gene Networks That Count. *Science*, *324* (5931), 1199–1202. American Association for the Advancement of Science. https://doi.org/10.1126/science.1172005

Fuller, M. (2021). *letmejustputthishere/dna_binary_counter*. MATLAB, . retrieved 28.1.2021. Available at: https://github.com/letmejustputthishere/dna_binary_counter

Gibson, B., Wilson, D. J., Feil, E. & Eyre-Walker, A. (2018). The distribution of bacterial doubling times in the wild. *Proceedings of the Royal Society B: Biological Sciences*, *285* (1880). https://doi.org/10.1098/rspb.2018.0789

Gillespie algorithm. (2021, January 10). *Wikipedia*. retrieved 28.1.2021. Available at: https://en.wikipedia.org/w/index.php?title=Gillespie_algorithm&oldid=999553916

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The*

Journal of Physical Chemistry, *81* (25), 2340–2361. https://doi.org/10.1021/j100540a008

Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology*, *177* (14), 4121–4130. https://doi.org/10.1128/jb.177.14.4121-4130.1995

Ham, T. S., Lee, S. K., Keasling, J. D. & Arkin, A. P. (2006). A tightly regulated inducible expression system utilizing the fim inversion recombination switch. *Biotechnology and Bioengineering*, *94* (1), 1–4. https://doi.org/https://doi.org/10.1002/bit.20916

Kim, Y.-I., Burton, R. E., Burton, B. M., Sauer, R. T. & Baker, T. A. (2000). Dynamics of Substrate Denaturation and Translocation by the ClpXP Degradation Machine. *Molecular Cell*, *5* (4), 639–648. https://doi.org/10.1016/S1097-2765(00)80243-9

Laalami, S., Zig, L. & Putzer, H. (2014). Initiation of mRNA decay in bacteria. *Cellular and Molecular Life Sciences*, *71* (10), 1799–1828. https://doi.org/10.1007/s00018-013-1472-4

Ringrose, L., Lounnas, V., Ehrlich, L., Buchholz, F., Wade, R. & Stewart, A. F. (1998). Comparative kinetic analysis of FLP and cre recombinases: mathematical models for DNA binding and recombination. *Journal of Molecular Biology*, *284* (2), 363–384. https://doi.org/10.1006/jmbi.1998.2149

Rong, M., He, B., McAllister, W. T. & Durbin, R. K. (1998). Promoter specificity determinants of T7 RNA polymerase. *Proceedings of the National Academy of Sciences*, *95* (2), 515–519. National Academy of Sciences. https://doi.org/10.1073/pnas.95.2.515

Roussel, M. R. (n.d.). Modelling Biochemical Reaction Networks - Lecture 14: Stochastic theory of reaction kinetics, 12.

Singh, S. K., Grimaud, R., Hoskins, J. R., Wickner, S. & Maurizi, M. R. (2000). Unfolding and internalization of proteins by the ATP-dependent proteases ClpXP and ClpAP. *Proceedings of the National Academy of Sciences*, *97* (16), 8898–8903. https://doi.org/10.1073/pnas.97.16.8898

Tabor, S. (1990). Expression Using the T7 RNA Polymerase/Promoter System. *Current Protocols in Molecular Biology*, *11* (1), 16.2.1-16.2.11. https://doi.org/https://doi.org/10.1002/0471142727.mb1602s11

Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J. et al. (2010). Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science*, *329* (5991), 533–538. American Association for the Advancement of Science. https://doi.org/10.1126/science.1188308

Whiteson, K. L. & Rice, P. A. (2008). Binding and Catalytic Contributions to Site Recognition by Flp Recombinase. *The Journal of Biological Chemistry*, *283* (17), 11414–11423. https://doi.org/10.1074/jbc.M800106200

Supplementary Information

The following table gives an overview of the parameters used in the model and how they were obtained.

Parameter	Value (s^{-1})	Description	Calculation
k_BAD_on	0.0002	TF activation rate	Derived from k_BAD_off $Hill\ equation: \frac{TF}{IC_{50} + TF}$ $IC_{50} = k_d = \frac{k_{off}}{k_{on}}$ $50 = \frac{0.01}{k_{on}}$
k_BAD_off	0.01	TF deactivation rate	Given in the lecture
kM_BAD_flp	0.00231	Flp transcription	mRNA steady-state concentration 1 (Taniguchi et al., 2010), half-life 5 min (Bernstein et al., 2002; Laalami et al., 2014): $k_{deg} = \frac{\ln(2)}{5 * 60}$ $1 = \frac{k_{syn}}{k_{deg}}$ $k_{syn} = k_{deg}$
kP_Flp	0.0577	Flp translation	Protein steady-state concentration 100 (Taniguchi et al., 2010), half-life 20 min (Gibson et al., 2018) : $k_{deg} = \frac{\ln(2)}{20 * 60}$ $100 = \frac{k_{syn}}{k_{deg}}$ $k_{syn} = k_{deg} * 100$
kR_Flp	4.81*10^-5	Flp recombination	4 hour half-life derived from Ham et al. : $k_{syn} = \frac{\ln(2)}{60 * 60 * 4}$
gM_flp	0.00231	flp mRNA degradation	$k_{deg} = \frac{\ln(2)}{5 * 60}$
gP_Flp	0.011	Flp degradation	60 second ssrA tagged protein half-life derived from Kim et al. and Singh et al.: $k_{deg} = \frac{\ln(2)}{60}$
gP_Ara	0.0003 or 0.1201	Ara degradation whilst Ara pulse and whilst Ara absence	Degradation rates derived from Friedland et al.
kM_BAD_gfp_T7p	0.00231	T7 promoter transcription	See kM_BAD_flp
kP_GFP	0.0577	GFP translation	See kP_Flp
kP_T7p	0.0577	T7 polymerase translation	See kP_Flp
kM_T7p_flp	5*0.00231	T7 polymerase transcription rate of flp mRNA	Derived from Tabor
gM_gfp	0.00231	gfp mRNA degradation	See kM_BAD_flp

gM_T7p	0.00231	T7 polymerase mRNA degradation	See kM_BAD_flp
gP_GFP	0.000577	GFP degradation	See kP_Flp
gP_T7p	0.011	T7 polymerase degradation	See gP_Flp

Table 1:Model parameters and how they were obtained.