Submitted, 2015 Winter q-bio Conference (5 Nov 2014)
http://www.cds.caltech.edu/~murray/papers/hhm15-wqbio.html

Characterization of minimum inducer separation time for a two-input integrase-based event detector

Victoria Hsiao,[*,†] Yutaka Hori,[‡] and Richard M Murray[†,¶]

Biology and Biological Engineering , California Institute of Technology, Pasadena CA, Computing and Mathematical Sciences, California Institute of Technology, Pasadena CA, and Control and Dynamical Systems, California Institute of Technology, Pasadena CA

E-mail: vhsiao@caltech.edu

Abstract

In this work, we present modeling and experimental characterization of the minimum time needed for flipping of a DNA substrate by a two-integrase event detector. The event detector logic differentiates the temporal order of two chemical inducers. We find that bundling biological rate parameters (transcription, translation, DNA searching, DNA flipping) into only a few rate constants in a stochastic model is sufficient to accurately predict final DNA states. We show, through time course data in *E.coli*, that these modeling predictions are reproduced *in vivo*. We believe this model validation is critical for using integrase-based systems in larger circuits.

[*]To whom correspondence should be addressed
[†]Biology and Biological Engineering , California Institute of Technology, Pasadena CA
[‡]Computing and Mathematical Sciences, California Institute of Technology, Pasadena CA
[¶]Control and Dynamical Systems, California Institute of Technology, Pasadena CA

Keywords

integrase, DNA memory, event detectors, stochastic biomolecular models

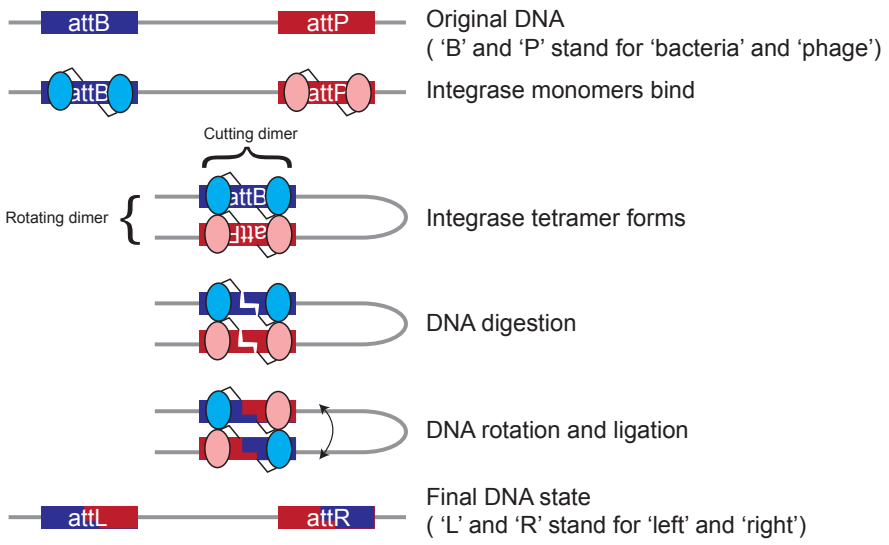

Figure 1: Serine integrase mechanism of action. The integrase monomers bind to the attB and attP recognition sites, forming two cutting dimers. These dimers then form a tetramer and the DNA is subsequently digested, flipped and rotated.

Introduction

The large serine-type phage integrases (Figure 1) are a class of tetrametric enzymes used by phages to integrate viral DNA into host bacterial chromosomes (*1*). They are closely related to tyrosine integrases such as Cre and Flp, but have the additional advantage of being strongly directional. Serine integrases recognize flanking DNA binding domains (attB, attP) and subsequently digest, rotate, and re-ligate the DNA between the recognition sites. Their directionality and relatively short recognition sites (40-50bp) make them attractive options for genome editing and synthetic circuits.

In 2009, Friedland *et al.* pioneered the use of serine integrases for sequential counting of single inducer pulses (*2*). They showed that the circuit responded to 8 hr pulses of arabinose (with 8 hrs of no inducer), and the final output, GFP, was only expressed upon detection of

three sequential pulses. In 2012, Bonnet *et al.* demonstrated a reversible integrase memory module using the integrase Bxb1 and its cofactor Xis that binds onto the enzyme and reverses the directionality (*3*). Similar cofactors have also been found for ϕC31 (*4*). Following that work, Siuti *et al.* (*5*) and Bonnet *et al.* (*6*) independently then both used systems of two integrases to compose multiple logic gates. Siuti *et al.* used Bxb1 and ϕC31 to compose 16 logic gates, while Bonnet *et al.* used Bxb1 and TP901-1 to demonstrate a set of 8 logic gates.

In the same 2009 work, Friedland *et al.* (*2*) also demonstrated the used of their integrase event counter system for the detection of three sequential pulses of three different inducers. The output, GFP, would only be expressed if the inducers were introduced in the correct order.

Recently, Yang *et al.* identified a set of 11 previously uncharacterized orthogonal integrases (*7*). Their fastest integrase, int8, had a two-hour activation time. This new library of integrases opens up the possibilities of integrase-based circuits.

We would like to build upon this previous work by incorporating the integrase-based memory module into a larger event detector system. This work discusses the initial characterization and modeling of the system. Specifically, we have designed an *A then B* logic gate and analyzed the response time as well as required separation time of A and B for maximum response.

Stochastic model of integrase flipping

First we implemented our event detector logic design into a stochastic model (Figure 2). The model has six species - the four possible DNA states, and the two integrases, intA and intB. The four possible DNA states are: the original state, the intB excision state, the intA single flip state, and the *A then B* double flip state. Once a DNA cassette has flipped into any of the states other than the original state, there is no reverse process. The logic gate

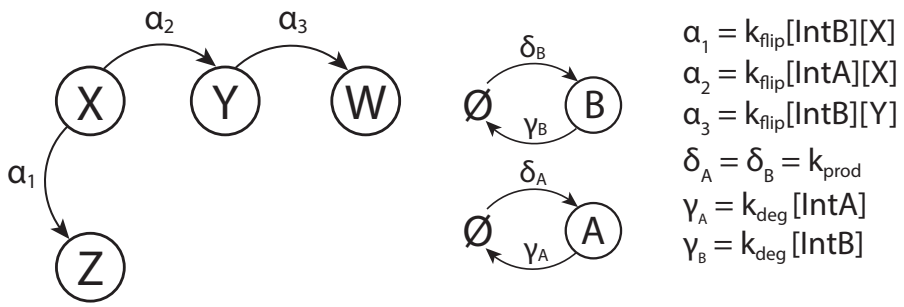

Figure 2: Stochastic model of integrase mediated DNA flipping. The four possible DNA states, illustrated as a Markov chain. All the DNA starts in state [X].Note that none of the processes are reversible. The propensity functions, $\alpha_1, \alpha_2, and \alpha_3$, are dependent on the concentration of the two integrases and the rate constant for DNA flipping, k_{flip}.

is designed such that if integrase B is expressed prior to integrase A, the DNA cassette is excised and the chain reaches a dead end (state *B first*). The desired result is for the DNA to first go to state A, then state *A then B*.

Since the target DNA cassette is chromosomally integrated into the cell, each cell only has one copy. In order to replicate this *in silico*, each simulation of the system only has one copy of DNA that can be flipped ($[X]_o = 1$).

This model combines the rates transcription and translation of the integrases into a single production rate constant, k_{prod}, and set to 1 $(\mu m^3 \cdot hr)^{-1}$, where the estimated volume of a single cell is $1\mu m^3$. The integrase degradation rate constant, k_{deg}, is set to 0.1 hr^{-1}. Any individual integrase, once produced, will need to search for the DNA binding site, bind to the DNA, tetramerize, then digest, rotate, and ligate the DNA. We have combined all of those rates into a single rate constant, k_{flip}, and that is also set at 1 hr^{-1}. The production, degradation, and flipping rates are the same for integrases A and B. The propensity functions are dependent on the concentration of the two integrases and the rate constant for DNA flipping, k_{flip}.

Using the Gillespie stochastic simulation algorithm(*8*) to simulate individual cells in a

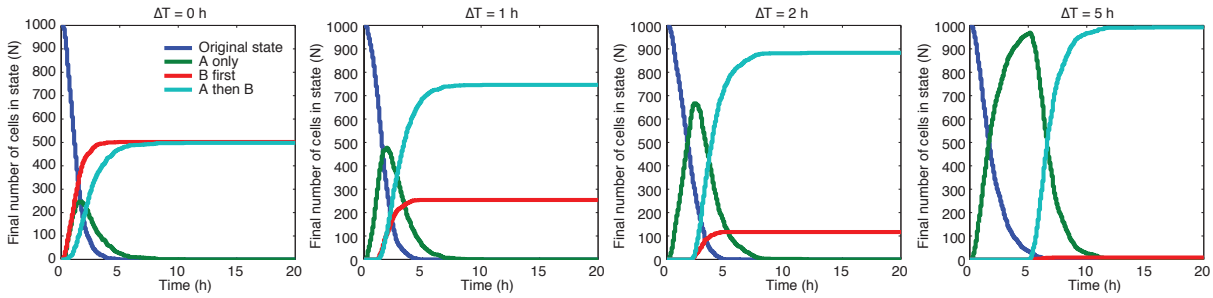

Figure 3: Effect of inducer separation time and final cell population states. Simulations for separation times of 0, 1, 2, and 5 hours between inducer A ($t = 0$h) and inducer B ($t = \Delta T$h) are shown.

population (n = 1000), we can look at population states over time (Figure 3) as a function of 0, 1, 2, and 5 hour separation time between the addition of inducer A and inducer B. We assume that each cell only has one DNA cassette that can be flipped. Inducer A is always added at $t = 0$h, and inducer B is added at time $t = \Delta T$ h. We see that when A and B are added simultaneously the number of cells in the B excision state is equal to the number of cells in the *A then B* state. With a separation time of 1 hour, we see a 25/75% split between the *B first* and *A then B* states, respectively. With a ΔT of two hours, that split widens to 10/90%, and with ΔT of five hours, nearly 100% of all the cells are in the desired *A then B* state.

We can then simulate an experiment in which we test the effect of ΔT from 0 to 10 hours. We would like to focus on the final fraction of cells in the *A then B* state, the rate at which the circuit responds to inducer input, and the minimum time required to get 90 % of population in one state. Figure 4A,B shows the superimposed timecourse results of just the cells in the *A then B* state over a range of $\Delta T = 0 - 10$ hours. In Figure 4A, we see cells that have been exposed to the A then B sequence of inducers - Just as we observed before, the simultaneous addition of both inducers results in only 50% of the final population in the *A then B* state. As ΔT increases, so does the fraction of cells.

Figure 4B shows cells that are exposed to a B then A sequence of events. In this case inducer B is added at t = 0h, and inducer A is added at $t = 0 + \Delta T$. Here we see the

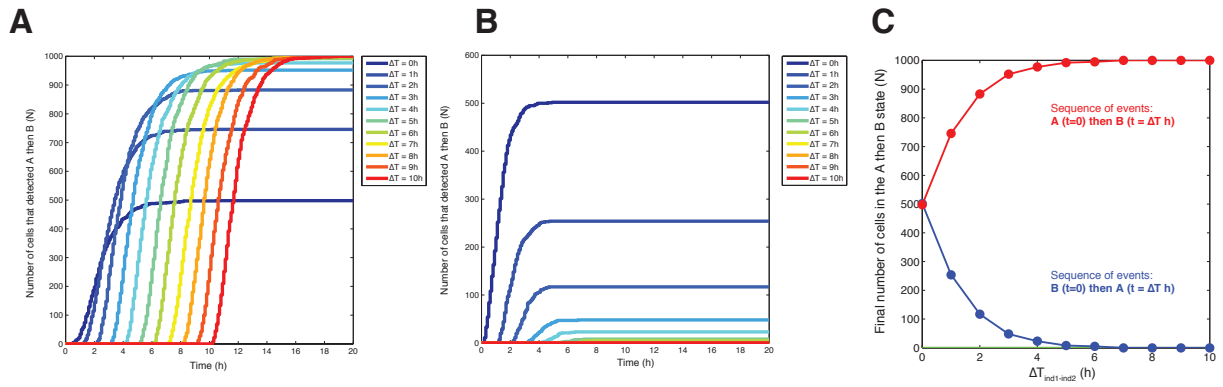

Figure 4: Staggered inducer separation times. A) Cells that see the inducer order *A then B* see increasing output as separation time increases. B) Cells that see the inducer order *B then A* see decreasing output as separation time increases. C) A summary of final number of cells in the *A then B* state as a function of separation time.

opposite trend as Figure 4A, with the number of cells in the A then B state decreasing as inducer A is added later.

Figure 4C shows the final number of cells from panels Fig. 4A and Fig.4B. We see that a minimum four hour separation time is needed to have at least 90% of the population detecting the same result.

Experimental results

The inducer separation experiment was conducted for $\Delta T = 0 - 9h$. GFP expression represents successful DNA flipping into the *A then B* state. In Figure 5 we see a distinct curves arise with the addition of inducer B at $t = \Delta T$. The curves have been normalized to the background fluorescence of a non-induced control. Notably, we see some amount of leaky GFP expression in the *A only* and *B only* cases - This indicates some amount of background leaky expression of both of the integrases. Figure 5A shows the increase in GFP expression with increased separation time - $\Delta T \geq 4h$ is sufficient for maximum expression. Additionally, the circuit continued to respond well past the final induction time of 9 hours, reaching the same steady state after 30 hours.

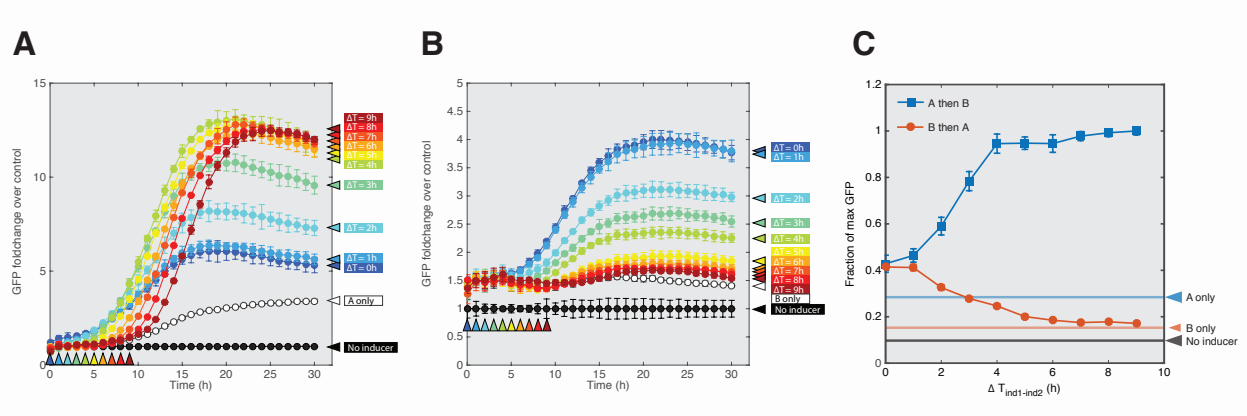

Figure 5: Experimental results for staggered inducer separation times. A) Cells that see the inducer order *A then B* see increasing output as separation time increases. B) Cells that see the inducer order *B then A* see decreasing output as separation time increases. C) A summary of final number of cells in the *A then B* state as a function of separation time.

In Figure 5B, we see cells that have been exposed to a B then A sequence of events. We see that maximum GFP expression occurs when both inducers are added at the same time, then, as cells are exposed to only B for longer periods of time, steady state GFP expression drops. Figure 5C summarizes the final steady state GFP expression based on the event order that the cells are exposed. For complete population switching, four hours is needed, but even within two hours, the population reaches 60% of the maximum expression.

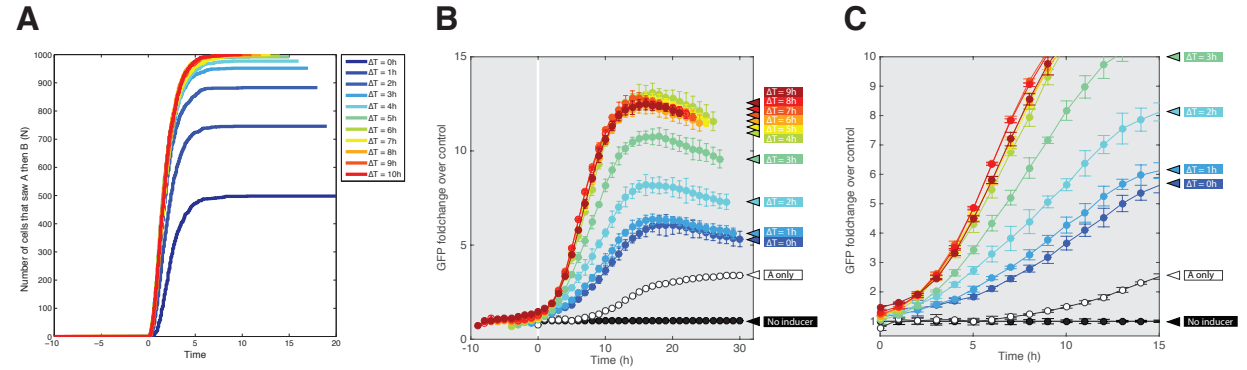

Figure 6: Alignment of output curves by time of second inducer. A) Aligned simulation results show that the slopes do not align for the lower separation times. B) Experimental data reproduces these trends. C) A zoomed in view of the experimental data shows the clear separation based on ΔT as well as the overlap once the ΔT has reached the minimum separation time of 4 hours.

In Figure 6, the curves for the A then B event sequence have been aligned to the time at which inducer B is added. This will allow for slope comparison, which is a proxy of how quickly cells are transitioning into the *A then B* state. In Figure 6A, the simulation shows that we should not expect the slopes to be the same. In the case where the two inducers are added simultaneously, it takes longer for cells to transition to the *A then B* state because they first need to go to the *A only* state. As ΔT increases, there is a greater buildup of cells in the *A only* state, so that they can immediately be converted into the *A then B* state when inducer B is added. Figure 6B shows the same *in vivo* data with all of the curves aligned to the time of inducer B induction, and indeed we see that only the slopes of $\Delta T \geq 4$ hours line up - separation times less than four hours result in lesser GFP production slopes. Figure 6C shows a zoomed-in view of Figure 6B, and we can see that all of the cultures start to produce GFP within 1 hour of B induction, but produce it at varying rates.

Conclusion

In this brief technical report, we have investigated the kinetics of an integrase-mediated two input event detector. We have determined that the minimum separation time for maximum population output is 4 hours and that GFP production (as a product of integrase transcription/translation/flipping) begins immediately after induction. These initial tests measure bulk fluorescence, but flow cytometry will determine the fraction of the population that is actually expressing GFP. Additionally, further studies will be done to determine minimum pulse duration - previous studies have suggested 15 minutes is sufficient (*6*).

Materials and Methods

The DNA logic cassette was chromosomally integrated into a strain of DH5α-Z1 *E.coli* cells (*9*) and grown in M9CA media. Arabinose and aTc were used as inducers A and B, respectively.

The Gillespie SSA was implemented in MATLAB and simulated 1000 times.

Acknowledgments

The authors would like to thank J. Bonnet for initial plasmids used in this work.

VH is supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. Y.H. is supported by JSPS Fellowship for Research Abroad. Research supported in part by the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

1. Yuan, P.; Gupta, K.; Van Duyne, G. D. Tetrameric structure of a serine integrase catalytic domain. *Structure* **2008**, *16*, 1275–1286.

2. Friedland, A. E.; Lu, T. K.; Wang, X.; Shi, D.; Church, G.; Collins, J. J. Synthetic gene networks that count. *Science* **2009**, *324*, 1199–1202.

3. Bonnet, J.; Subsoontorn, P.; Drew, E. Rewritable digital data storage in live cells via engineered control of recombination directionality. *PNAS* **2012**, *109*, 8884–8889.

4. Khaleel, T.; Younger, E.; McEwan, A. R.; Varghese, A. S.; Smith, M. C. M. A phage protein that binds φC31 integrase to switch its directionality. *Molecular Microbiology* **2011**, *80*, 1450–1463.

5. Siuti, P.; Yazbek, J.; Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology* **2013**, 1–6.

6. Bonnet, J.; Yin, P.; Ortiz, M. E.; Subsoontorn, P.; Drew, E. Amplifying genetic logic gates. *Science* **2013**, *340*, 599–603.

7. Yang, L.; Nielsen, A. A. K.; Fernandez-Rodriguez, J.; McClune, C. J.; Laub, M. T.; Lu, T. K.; Voigt, C. A. Permanent genetic memory with >1-byte capacity. *Nature Methods* **2014**,

8. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* **1977**, *81*, 2340–2361.

9. Haldimann, A.; Wanner, B. L. Conditional-Replication, Integration, Excision, and Retrieval Plasmid-Host Systems for Gene Structure-Function Studies of Bacteria. *Journal of Bacteriology* **2001**, *183*, 6384–6393.