Development and Experimental Validation of a Mechanistic Model of a Recombinase-Based Temporal Logic Gate*

Jack Bowyer[1], Victoria Hsiao[2] and Declan G. Bates[1]

***Abstract*— DNA recombination provides an ideal mechanism for constructing stable and reversible synthetic biological switches. Recent advances in recombinase-based circuitry that account for more than one protein input have been shown to enable the construction of circuits with temporal Boolean logic operations *in vivo*. Associated mathematical models have to date only captured the qualitative dynamical features of such systems and are thus of limited utility as tools to aid in the design of such circuitry. Here we develop a detailed mechanistic model of a two-input temporal logic gate circuit based on unidirectional DNA recombination with bacteriophage integrases to detect and encode sequences of input events. The model is validated against *in vivo* experimental data and is shown to quantitatively replicate and predict key dynamical features of the logic gate.**

I. INTRODUCTION

Site-specific recombinases (SSRs) are capable of highly efficient integration, excision and inversion of genetic sequences [1]. Manipulating DNA in this manner has the potential to enable inducible expression of any gene of interest. Hence, fully characterized genetic switches based on DNA recombination represent powerful biological parts for the design of novel synthetic devices [2]. Serine integrases, one of the main SSR families, bind attB and attP attachment sites on the target DNA sequence, causing double stranded breaks that are reconnected by opposite ends of the intermediate genetic sequence. The sequence is thereby inverted and flanked by newly formed attL and attR attachment sites. A recombination directionality factor (RDF), often referred to as excisionase, works in conjunction with integrase to mediate the inverse process; binding to attL and attR to re-establish attB and attP. An alternative orientation of attB and attP sites will result in the complete integration or excision of the entire genetic sequence into or from the host, depending on the initial DNA state [3]–[6].

Single-input recombination offers highly efficient integration since integrase alone is sufficient to mediate the reaction. However, the efficiency of subsequent excision is highly sensitive to the stoichiometry of integrase to excisionase [7]. As a result, synthetic biologists looking to exploit this particular mechanism may be limited to designs that can either support constant protein expression or that do not rely on temporal induction of the desired outputs. For the latter, a single-input system would largely amount to a genetic switch that can be turned 'on' and hold its state, but cannot be turned 'off'; providing useful albeit limited functionality. Boolean logic gates have been developed successfully in a host of biological contexts, enabling a digital response to multiple inputs [9]–[11]. In realising the true potential of this circuitry, many synthetic circuits would therefore benefit greatly from the ability to provide temporal response dynamics, especially those pertaining to *in vivo* logic operations. In [12], a temporal logic gate with "a then b" logic was demonstrated in *E. coli* with a system of two orthogonal integrases (integrases A and B). A simple stochastic model of this circuit was created to help develop better intuition for overall circuit behavior, but this model does not account for any specific molecular interactions between the integrases and the DNA, instead representing integrase activity as probabilities based on concentration. This model was shown to be effective for predicting overall final population fractions as well as forward experimental design of the system. However, the inherent limitations of the model design mean that the circuit cannot be simulated on a molecular scale, and timescales with regards to specific molecular interactions cannot be incorporated.

Here, we adapt a recently developed mechanistic model of DNA recombination [7], [8] to model the two-integrase temporal logic gate circuit developed in [12]. We demonstrate that the mechanistic model successfully captures key dynamical features of circuit time course trajectories derived from *in vivo* experimental data, thus improving our capability to perform model-aided integrase circuit design.

II. DEVELOPING A MECHANISTIC MODEL OF A TEMPORAL LOGIC GATE

The model of [12] describes the transitioning of the system from the original DNA state (S_0) to each of the three end states (S_a, S_b and S_{ab}) via three corresponding rate constants. As shown in Fig. 1, we replace these all-encompassing parameters with a mechanistic integration reaction structure, in which integration is initiated by the binding of one integrase dimer at each of the associated attachment sites and is strictly unidirectional [13], [14]. We account for dimerization of both monomeric SSRs, allowing for both monomeric and dimeric integrase binding to DNA attachment sites. This process is widely supported in the experimental literature on DNA recombination [3]–[6]. Thus, four distinct intermediate DNA:protein complexes can potentially be formed in facilitating recombination via this combination of monomeric and dimeric integrase binding. We also include the formation of a dysfunctional dimer by both integrases, which is subject to

*Research supported by EPSRC via a DTA studentship to J. Bowyer

[1]J. Bowyer and D. Bates are with the Warwick Integrative Synthetic Biology Centre and School of Engineering, University of Warwick J.E.Bowyer@warwick.ac.uk, D.Bates@warwick.ac.uk

[2]V. Hsiao is with Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125 vhsiao@caltech.edu

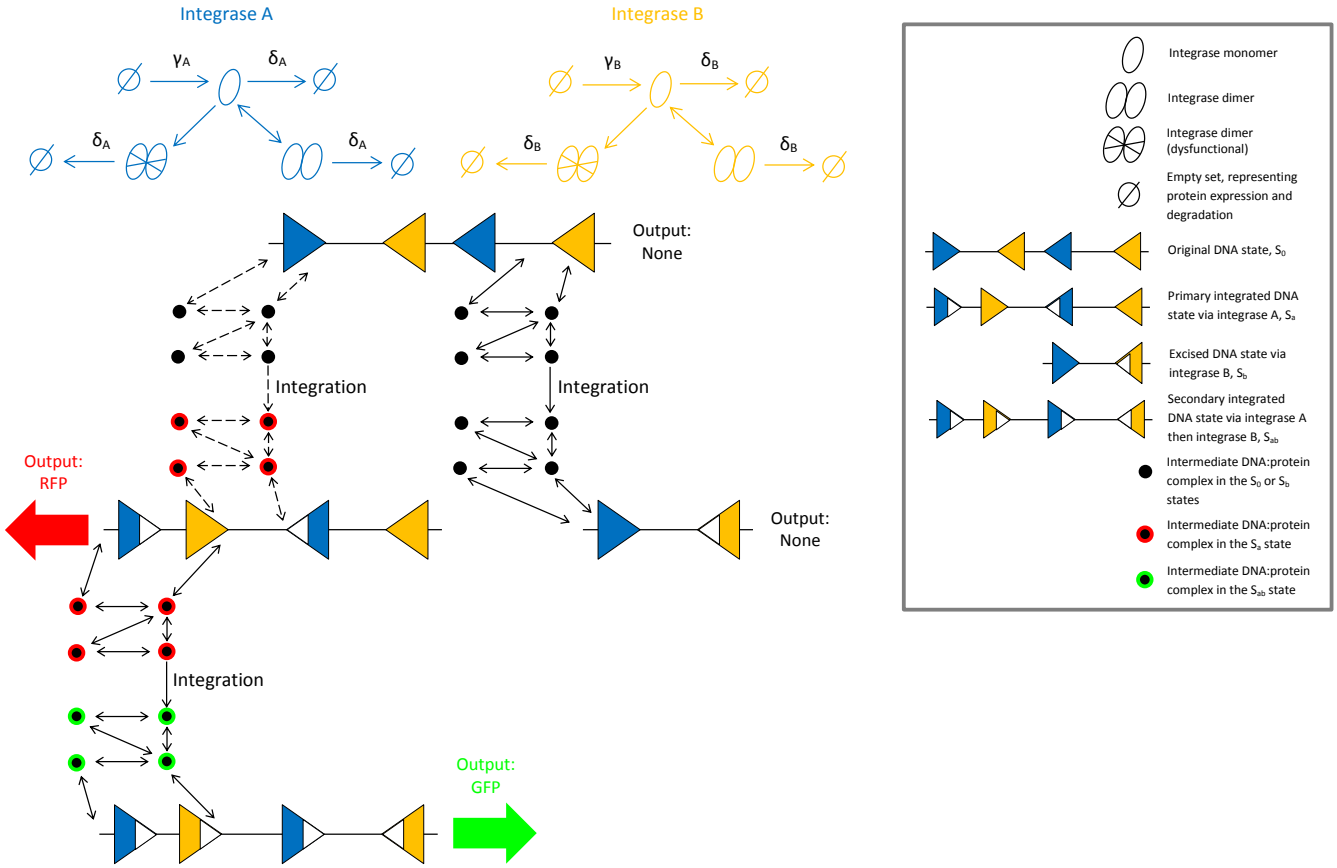

Fig. 1. Diagram of the mechanistic two-input logic gate reaction network. The sequence of DNA:protein interactions facilitating integration, taken from [7], enables the system to transition from the original DNA state (S_0) to the three genetically-differentiated DNA states (S_a, S_b, S_{ab}). Expression and degradation of integrase A and B are denoted by γ_A, δ_A and γ_B, δ_B respectively. Intermediate DNA:protein complexes with red and green outlines are associated with S_a and S_{ab} respectively; the summation of ODEs corresponding to these complexes give rise to (2) and (3) respectively. Dashed and solid black arrows depict reactions mediated by integrase A and B respectively. Adapted from [12].

the same degradation as its functional counterpart; this was shown to significantly improve model fits to experimental data in our previous study [7], [8].

Model validation against experimental data for an *in vivo* recombinase-based system presents a number of issues that need to be taken into account. Cellular recombination *in vivo* is dependent on the expression and degradation of recombinase proteins over time, thus contributing four additional parameters ($\gamma_A, \delta_A, \gamma_B, \delta_B$) to the model. Furthermore, each individual phase of a typical bacterial cell growth cycle influences the rate at which certain cellular processes occur, including protein expression ($\gamma_{A,B}$). Hence, for optimal results, we must account for time-dependent variation of protein expression rather than the constant expression exhibited in previous models [2], [12]. In order to capture the effect of cell growth on protein expression we implement the logistic function, which has been shown to provide accurate representations of bacterial growth in the literature [15], [16]:

$$f_{A,B}(t) = \frac{\gamma_{A,B}}{1+\exp\left(\frac{(t^{\star}_{A,B}-t)}{D_{A,B}}\right)} \tag{1}$$

where the time point $t^{\star}$ and the damping constant D specify the mid point and gradient of the function respectively. We do not model any effects of bacterial cell growth on the rate of protein degradation [17] and hence the corresponding parameters ($\delta_{A,B}$) remain as constants in the model. In the absence of induction, *in vivo* systems exhibit regular expression of SSRs resulting in basal recombination efficiency. Thus we include model parameters describing both basal ($\gamma_{a,b}$) and induced ($\gamma_{A,B}$) expression of the two integrases, allowing for non-zero model output when simulating experiments void of integrase induction.

Our mechanistic model is constructed through the application of mass action kinetics to the biochemical equations arising from the reaction network in Fig 1. This produces a system of ordinary differential equations (ODEs) that are solved numerically to provide a deterministic model output. The formation of intermediate DNA:protein complexes, due to monomeric and dimeric integrase binding, in our mechanistic model gives rise to multiple state variables associated with the two DNA states of interest, S_a and S_{ab}. Summing all the ODEs describing the dynamics of state variables associated with the same DNA state of interest provides the total register of the system in those states (S_{aT} and S_{abT}).

Hence, model outputs are determined through the numerical solutions to the following ODEs:

$$\frac{dS_{aT}}{dt} = k_{RA}S_0I_{4A} - k_{RB}S_aI_{4B}, \quad (2)$$

$$\frac{dS_{abT}}{dt} = k_{RB}S_aI_{4B}. \quad (3)$$

Our mechanistic model consists of thirty-four ODEs with forty-three model parameters, which are solved in determining (2) and (3).

III. MODEL SIMULATION AND PARAMETER OPTIMIZATION

Induction of integrase B prior to integrase A causes transition to the unwanted excised DNA state, S_b, and therefore we optimize our model against data regarding induction of integrase A prior to integrase B only. Our data is comprized of both red and green fluorescent protein (RFP and GFP) levels (S_a and S_{ab}) under eight distinct experimental conditions (See Materials and Methods). Firstly, fluorescence is recorded for no induction of either integrase and, secondly, fluorescence is recorded for induction of integrase A only. Six experimental procedures record fluorescence for induction of integrase B at increasing time intervals, δT, such that $\delta T = 0,\ldots,5$ hours following induction of integrase A. Since the observed fluorescence has no physical dimension, the data for S_a and S_{ab} is normalized with respect to their respective initial (no inducer) case experiment. This establishes the fold change in fluorescence output between each of the eight experiments and provides the numerical comparisons required to fit the parameters in our model.

Given that we are using a deterministic model to simulate recombination efficiencies within a single cell, we overcome uncertainty regarding physical quantities of DNA by choosing an initial DNA concentration (S_0) of 1, hence all model outputs are bounded within $[0,1]$. Once a numerical output has been computed, it is multiplied by 2×10^4 since this is the upper bound of the observed fluorescence. Model outputs are subject to the same normalization applied to the experimental data, establishing fold changes in observed output for varying time intervals between the induction of integrases A and B. All data is normalized with respect to the no-inducer case of that particular experiment. Replicating this process *in silico* involves normalizing all model outputs with respect to the model output corresponding to the no inducer case for the particular experiment being simulated. Simulating the case whereby no induction of integrases A and B occurs relies on the model parameters describing basal expression of the two integrases to replicate the non-zero outputs observed experimentally for no-inducer cases.

IV. RESULTS - VALIDATION AGINST *IN VIVO* EXPERIMENTAL DATA

A comparison of model outputs against experimental data is shown in Fig. 2. The optimized mechanistic model is able to capture the observed system dynamics (Fig. 2A, 2B) using a set of optimized parameter values that are within biologically plausible ranges. For example, all parameters associated with DNA:protein interactions take values within $[1,100]$, exhibiting a maximum variation of two orders of magnitude.

After training our model using datasets from Fig. 2A and 2B, we used our optimized mechanistic model to predict a further set of experimental data that was excluded from the training set. Fig. 2C shows experimental data for increasing integrase induction delays ($\delta T = 1,\ldots,5$) along with optimized model predictions. Additionally, we validated our model by predicting endpoint GFP fractions relating to both A then B temporal response data (Fig. 2C) and an entirely separate dataset regarding B then A temporal responses. The endpoint response of the system as a function of the integrase induction separation interval δT is shown in Fig. 3. Optimal fold change endpoint model outputs as a function of δT align closely with that of the experimental data, providing further evidence of the model's predictive capability. Fig. 3 also reveals that the efficiency of recombination induction via integrase B must be superior to that of integrase A since identical efficiencies would result in a 50:50 split for $\delta T = 0$. This inequality in integrase flipping was previously observed in [12], but no mechanistic comparisons had been performed at that time. Consequently, it remains to examine functional differences between distinct integrases as these properties may allow for specific logic operations dependent on the pair of integrases selected and the arrangement of the associated attachment sites.

V. MATERIALS AND METHODS

A. Experimental procedure

Experimental measurements were performed using an engineered *E. coli* strain with chromosomally-integrated temporal logic gate as previously described in [12]. Precise control of inducer separation times ($\delta T = 0,\ldots,5$) was achieved using a Hamilton STARlet liquid handling robot (Hamilton Robotics, Reno, NV, USA) integrated with an incubating plate reader. The inducers anhydrous tetracycline (aTc, 200 ng/ml) and arabinose (0.01%/vol) were used to induce expression of integrases A and B, respectively. Time course data for GFP and RFP fluorescence were measuring using a BioTek Synergy H1MF plate reader (BioTek Instruments, Inc. Winooski, VT, USA). Cell cultures were grown inside the plate reader at 37°C in minimal M9CA media.

B. Global optimization

We employ a parallelized coding of a Genetic Algorithm (GA) on a high-performance computing cluster to perform global optimization of the mechanistic model against our large experimental dataset. The GA converges to the global minimum within the allocated parameter space by evolving an initial population of randomly generated solutions over a large number of generations. Best results are achieved for relatively large population sizes and generations compared to the number of parameters subject to inference. However, this also significantly increases the computational workload and hence a reasonable compromise is required for viable development times.

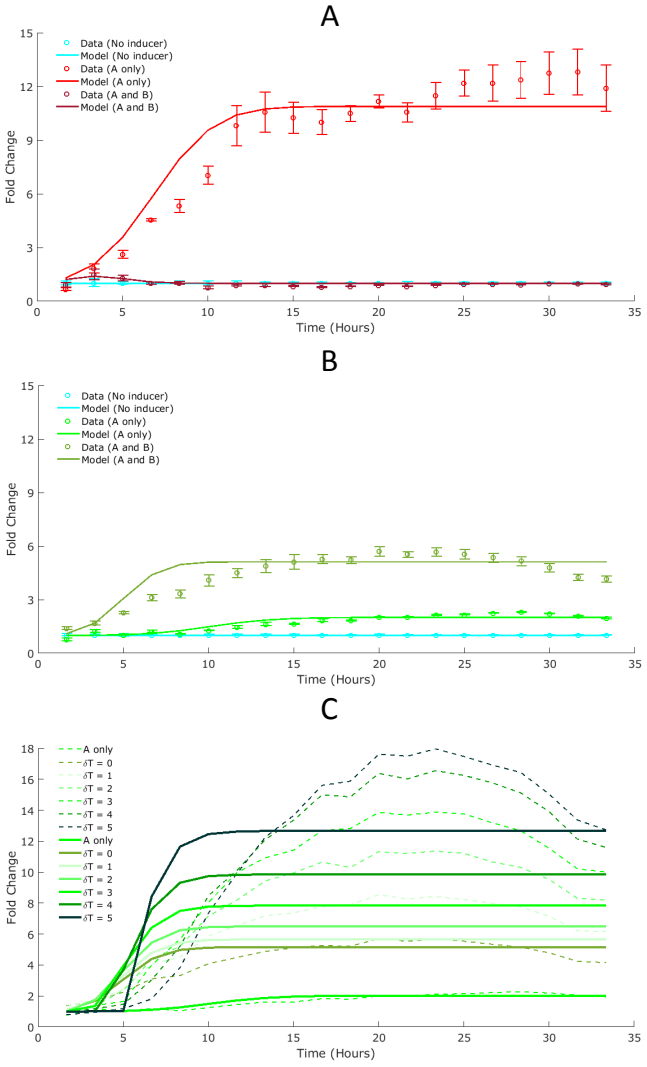

Fig. 2. Model matching and prediction results. Optimized responses of our mechanistic model against the two corresponding S_a (A) and S_{ab} (B) time course datasets. (C) Predictions of the mechanistic model of outputs in response to different induction separation intervals. Dashed lines depict experimental data; solid lines depict optimal model outputs (A only and δT$= 0$) and model predictions (δT$= 1,\ldots,5$).

VI. CONCLUSIONS

We have developed a detailed mechanistic model of a synthetic two-input temporal logic gate that incorporates experimentally well-supported mechanisms underlying DNA recombination *in vivo*. The predictive power of the model was validated against an experimental dataset on time course dynamics of the logic gate in the presence of none, one or both integrases, as well as data on the response of the gate to inputs separated by five different induction separation intervals. Temporal logic operations have the potential to expand the range of outputs of standard Boolean systems and hence this validated model could be a potentially valuable design tool for synthetic biologists working on the construction of more complex recombinase-based genetic circuitry. Future work will extend our modelling of fluorescent protein fold change output as a function of the separation

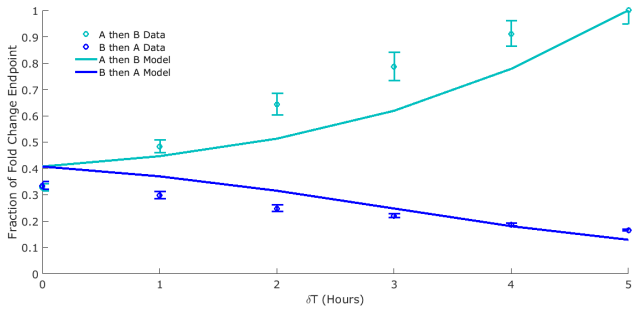

Fig. 3. A then B GFP fold change predictions from Fig. 2C converted to a fraction of the maximum fold change endpoint and plotted as a function of δT (δT$= 0,\ldots,5$; light blue plot line). Equivalent predictions of B then A temporal responses are shown by the dark blue plot line.

time between integrase induction events in order to identify potential functional distinctions between different integrases. The overlapping arrangement of integrase attachment sites may emerge as an important selection process in the event that functional distinctions are identified; presenting an array of potential input permutations and, in turn, expanding the overall performance specifications of the logic gate.

VII. SUPPLEMENTARY MATERIAL

A complete list of the model ODEs and the relevant MATLAB code is available on request from the authors.

REFERENCES

[1] M. Smith and H. Thorpe. Mol Microbiol. 2002 Apr;44(2):299-307.
[2] J. Bonnet, P. Subsoontorn and D. Endy. Proc Natl Acad Sci USA. 2012 Jun 5;109(23):8884-9. doi: 10.1073/pnas.1202344109.
[3] P. Fogg, S. Colloms, S. Rosser, M. Stark and M. Smith. J Mol Biol. 2014 Jul 29;426(15):2703-16. doi: 10.1016/j.jmb.2014.05.014. Epub 2014 May 22.
[4] P. Ghosh, N. Pannunzio and G. Hatfull. J Mol Biol. 2005 Jun 3;349(2):331-48. Epub 2005 Apr 7.
[5] P. Ghosh, L. Bibb and G. Hatfull. Proc Natl Acad Sci U S A. 2008 Mar 4;105(9):3238-43. doi: 10.1073/pnas.0711649105.
[6] T. Khaleel, E. Younger, A. McEwan, A. Varghese and M. Smith. Mol. Microbiol. 2011, 80, 14501463.
[7] J. Bowyer, J. Zhao, S. Rosser, S. Colloms and D.G. Bates, Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Milano, Italy, 2015, pp. 945-948.
[8] J. Bowyer, J. Zhao, P. Subsoontorn, W. Wong, S. Rosser and D. Bates. IEEE Transactions on Biomedical Circuits and Systems. doi: 10.1109/TBCAS.2016.2526668
[9] J. Bonnet, P. Yin, M. Ortiz, P. Subsoontorn and D. Endy. Science. 2013 May 3;340(6132):599-603. doi: 10.1126/science.1232758. Epub 2013 Mar 28.
[10] T. Moon, C. Lou, A. Tamsir, B. Stanton and C. Voigt. Nature. 2012 Nov 8;491(7423):249-53. doi: 10.1038/nature11516. Epub 2012 Oct 7.
[11] D. Shis, F. Hussain, S. Meinhardt, L. Swint-Kruse and M. Bennett. ACS Synth Biol. 2014 Sep 19;3(9):645-51. doi: 10.1021/sb500262f. Epub 2014 Jul 28.
[12] V. Hsiao, Y. Hori, P. Rothemund and R. Murray. Mol Syst Biol. 2016 May 17;12(5):869. doi: 10.15252/msb.20156663.
[13] A. Groth and M. Calos. J Mol Biol. 2004 Jan 16;335(3):667-78.
[14] F. Olorunniji, D. Buck, S. Colloms, A. McEwan, M. Smith, M. Stark and S. Rosser. Proc. Natl. Acad. Sci. U. S. A. 2012, 109, 1966119666.
[15] H. Fujikawa, A. Kai and S. Morozumi. Shokuhin Eiseigaku Zasshi. 2003 Jun;44(3):155-60.
[16] M. Zwietering, I. Jongenburger, F. Rombouts and K. v Riet. Appl Environ Microbiol. 1990 Jun;56(6):1875-81.
[17] G. Baxter and C. Stanners. J Cell Physiol. 1978 Aug;96(2):139-45.