Mechanistic Modeling of a Rewritable Recombinase Addressable Data Module

Jack Bowyer, Jia Zhao, Pakpoom Subsoontorn, Wilson Wong, Susan Rosser, and Declan Bates

***Abstract*—Many of the most important applications predicted to arise from Synthetic Biology will require engineered cellular memory with the capability to store data in a rewritable and reversible manner upon induction by transient stimuli. DNA recombination provides an ideal platform for cellular data storage and has allowed the development of a rewritable recombinase addressable data (RAD) module, capable of efficient data storage within a chromosome. Here, we develop the first detailed mechanistic model of DNA recombination, and validate it against a new set of *in vitro* data on recombination efficiencies across a range of different concentrations of integrase and gp3. Investigation of *in vivo* recombination dynamics using our model reveals the importance of fully accounting for all mechanistic features of DNA recombination in order to accurately predict the effect of different switching strategies on RAD module performance, and highlights its usefulness as a design tool for building future synthetic circuitry.**

***Index Terms*—DNA recombination, mathematical modeling, synthetic biology.**

I. Introduction

SYNTHETIC Biology is a relatively young field with immense potential in numerous applications at the interface of engineering and biology [1]–[5]. Several prototype synthetic circuits and systems have now been successfully designed, among the most important of which is the genetic switch, a key component in engineering cellular memory and providing a lasting response to transient stimuli in both *in vitro* and *in vivo* contexts. The genetic toggle switch in *E. coli* [6] was the first to successfully demonstrate the potential of synthetic memory devices through circuit design and the assembly of biological parts to realise desired transcriptional responses. The toggle switch exhibits bistability through a mutually inhibitory arrangement of promoters and repressors. Despite showing promise, regulating gene expression in this manner comes with limitations. These systems are volatile, having to continuously consume resources, in this case for the production and degradation of repressor, to maintain states. Difficulties also arise when integrating devices into a variety of organisms given that gene regulation networks vary greatly between distinct cellular environments. Furthermore, the highly inducible and stable switching that these devices demand can be compromised by spontaneous switching events caused by the inherent stochasticity of gene regulation. As a result, research into cellular memory has become increasingly focused towards site-specific recombinases (SSRs), capable of precise DNA manipulation both *in vitro* and *in vivo* [7]. DNA-based systems are favourable due to the fact that they exploit a natural data storage material and have the added advantage of eliminating cell specificity requirements. SSRs are classified as belonging to two groups, the tyrosine recombinases and the serine recombinases. The former have been shown to provide effective genetic switch mechanisms however, their functionality is often dependent upon cell-specific cofactors as in the case of λ integrase [8]. This is problematic in a similar vein to that of the transcriptional systems when looking to deploy modules across multiple organisms. Tyrosine recombinase systems that are not dependent on host cofactors are bidirectional and are therefore incapable of highly efficient switching [9]. In contrast, the serine recombinases do not require such cofactors and have been used effectively to perform highly efficient gene assembly and modification [10]. This has led to the construction of a rewritable RAD module exhibiting passive information storage within a chromosome [11]. Switching the RAD module "on" requires only the presence of integrase whereas the "off" switch requires integrase in conjunction with a recombination directionality factor (RDF), also referred to as excisionase. Our modeling investigation is primarily concerned with the dynamic behavior of the serine integrase ϕC31 and its associated RDF, gp3, although our literature review spanned a variety of SSRs.

There exists two distinct DNA recombination mechanisms, comprised of three distinct recombination events. The first mechanism exploits two of these three events, insertion and deletion (Fig. 1, top). In this case, recombination occurs between DNA attachment sites on a bacterial host chromosome and a bacteriophage. Integrase alone is sufficient to mediate insertion; dimeric integrase bound to a host chromosome attB site and a bacteriophage attP site causes a double stranded break in each. The exposed ends of the phage DNA fragment bind to those of the host, thus inserting the fragment into the host chromosome. The newly formed composite attachment sites flanking the genetic insert are termed attL and attR. Following insertion,

Manuscript received August 19, 2015; revised December 2, 2015; accepted February 1, 2016. Date of publication May 24, 2016; date of current version December 30, 2016. This work was supported by EPSRC via a DTA studentship to J. Bowyer and research grants EP/K034359/1, EP/H019154/1. This paper was recommended by Associate Editor R. Sarpeshkar.

J. Bowyer and D. Bates are with the Warwick Integrative Synthetic Biology Centre and School of Engineering, University of Warwick, Coventry CV4 7AL, U.K. (e-mail: J.E.Bowyer@warwick.ac.uk; D.Bates@warwick.ac.uk).

J. Zhao is with the Institute of Molecular, Cell and Systems Biology, University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail: j.zhao.1@research.gla.ac.uk).

P. Subsoontorn is with the Department of Plant Sciences, University of Cambridge, Cambridge CB2 1TN, U.K. (e-mail: ps690@cam.ac.uk).

W. Wong is with the Department of Biomedical Engineering, Boston University, Boston, MA 02215 USA (e-mail: wilwong@bu.edu).

S. Rosser is with SynthSys, School of Biological Sciences, University of Edinburgh, Edinburgh EH8 9YL, U.K. (e-mail: srosser2@staffmail.ed.ac.uk).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBCAS.2016.2526668

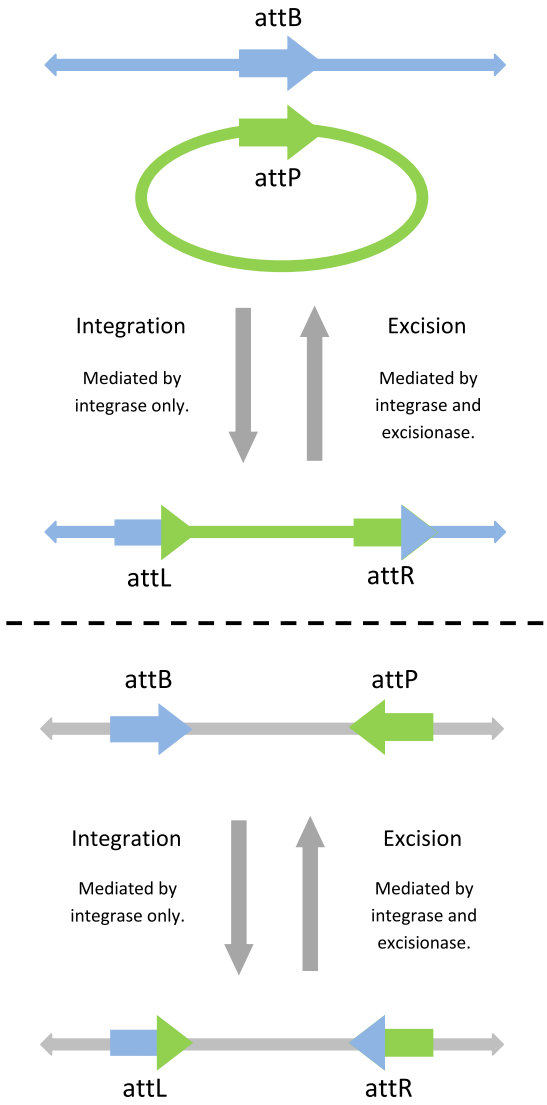

Fig. 1. Schematic diagram of DNA recombination mechanisms. Top: the phage genome attachment site, attP, is integrated into the host chromosome attachment site, attB. Integration gives rise to attL and attR, each formed of half of attB and attP. Excision restores attB and attP, removing the integrated phage genome from the host chromosome. Bottom: genetic material flanked by attB and attP is inverted through integration to form attL and attR. Excision restores attB and attP via a secondary inversion event.

binding of RDF molecules to the attL and attR DNA:integrase synaptic complexes facilitates the deletion event. Here the attL and attR sites are dismantled, thus deleting the genetic insert and allowing the reformation of the attB and attP sites. Insertion and deletion are generally referred to as integration and excision respectively since integrase alone mediates the former and a combination of integrase and excisionase is necessary to mediate the latter. The second mechanism exploits the third event, inversion (Fig. 1, bottom). In this case, attB and attP sites located on the same DNA sequence are subject to binding by dimeric integrase which causes double stranded breaks in each. Exposed ends in the genetic sequence then bind the opposite ends of the intermediate fragment hence resulting in an inverted section of DNA flanked by attL and attR. The binding of RDF molecules to the attL and attR DNA:integrase synaptic complexes facilitates a successive inversion event. The attL and attR sites are dismantled, allowing for the exposed ends of the intermediate fragment to adopt their original orientation and hence the reformation of the attB and attP sites. Inversion events giving rise to attL, attR sites and attB, attP sites are generally referred to as integration and excision respectively for the same reasons discussed earlier. Here we investigate inversion-based DNA recombination and refer to inversion events as integration and excision according to the attachment sites involved. We refer to the presence of attB, attP sites and attL, attR sites in the system as the BP and LR states respectively. The concentrations of integrase and excisionase in the system dictate the efficiency of the switching between these distinct DNA states.

A validated mathematical model is a vital tool in the design of a synthetic biological circuit. The ability to investigate the predicted behavior of systems outside of the laboratory offers the potential for greatly reduced development times, since computational results can often be achieved in a fraction of the time of their experimental counterparts. In addition, well defined mathematical models are capable of providing indications of biological details that may not be well documented experimentally from a mechanistic perspective. Sections II and III detail the construction and validation of a detailed mathematical model of DNA recombination, that can form the basis of a powerful design tool for the construction of recombinase-based circuitry *in vivo*, as demonstrated in Section IV. A preliminary version of this model that considered only *in vitro* data was presented in [12]. Here we provide a more comprehensive description and validation of this model and further extend it to the *in vivo* setting to investigate switching dynamics of recombinase-based circuitry. Compared with the simple model of *in vivo* recombination presented in [11], our mechanistic model provides a significantly more accurate prediction of switching dynamics in light of its increased level of mechanistic detail.

II. A Mechanistic Model of DNA Recombination

An extensive review of the experimental literature was carried out in order to synthesize the current state of knowledge regarding the mechanistic basis of DNA recombination. The literature review identified several mechanistic properties of the system that are well established; the application of mass action kinetics to each of the associated biochemical equations allows us to derive the system of ordinary differential equations (ODEs) comprising our mechanistic model, as described below.

Monomeric integrase is known to form dimers reversibly in solution [11], [13]–[36], and thus we can write the biochemical equation

$$I + I \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} I_2 \tag{1}$$

where I and I_2 denote monomeric and dimeric integrase respectively and k_i are the relevant reaction rate constants. Reversible reactions are denoted using right and left arrows with the corresponding forward and backward reaction rates written above and below respectively.

One integrase dimer bound to attB and attP is necessary to mediate the integration reaction [11], [13]–[21], [23]–[26], [28]–[54], which is unidirectional (irreversible) [11], [13], [14],

[17], [18], [20], [24], [26], [28]–[30], [32], [35], [37], [38], [42], [43], [45], [48], [50]–[57]. This gives the following additional dynamics:

$$D_{\mathrm{BP}} + I_2 \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} D_{\mathrm{BP}}I_2 \tag{2}$$

$$D_{\mathrm{BP}}I_2 + I_2 \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} D_{\mathrm{BP}}I_4 \tag{3}$$

$$D_{\mathrm{BP}}I_4 \xrightarrow{k_R} D_{LR}I_4 \tag{4}$$

$$D_{\mathrm{BP}}I_2R + I_2 \underset{k_{-8}}{\overset{k_8}{\rightleftharpoons}} D_{\mathrm{BP}}I_4R \tag{5}$$

where D_{BP} denotes free DNA in the BP state; $D_{\mathrm{BP}}I_2$, $D_{\mathrm{BP}}I_4$ denote DNA:protein complexes with one and two integrase dimers bound respectively in the BP state; $D_{\mathrm{LR}}I_4$ denotes the DNA:protein complex with two integrase dimers bound in the LR state. Irreversible reactions are denoted by a single right arrow with the corresponding reaction rate written above. We include (5) here to demonstrate that dimeric integrase is able to bind to any unoccupied attachment site, however, these particular complexes are not directly involved in the integration reaction.

Gp3 binds to dimeric integrase already bound to DNA attachment sites and it does not bind directly to DNA attachment sites [11], [16], [19], [20], [30], [35], [58]. This gives

$$D_{\mathrm{BP}}I_2 + R \underset{k_{-10}}{\overset{k_{10}}{\rightleftharpoons}} D_{\mathrm{BP}}I_2R \tag{6}$$

$$D_{\mathrm{BP}}I_4 + R \underset{k_{-11}}{\overset{k_{11}}{\rightleftharpoons}} D_{\mathrm{BP}}I_4R \tag{7}$$

$$D_{\mathrm{BP}}I_4R + R \underset{k_{-12}}{\overset{k_{12}}{\rightleftharpoons}} D_{\mathrm{BP}}I_4R_2 \tag{8}$$

where R denotes the RDF, gp3; $D_{\mathrm{BP}}I_2R$ denotes the DNA:protein complex with one integrase dimer and one gp3 monomer bound in the BP state; $D_{\mathrm{BP}}I_4R$, $D_{\mathrm{BP}}I_4R_2$ denote DNA:protein complexes with two integrase dimers and one/two gp3 monomers bound respectively in the BP state. Here we have shown biochemical equations arising from interactions involving DNA in the BP state; there are an equivalent number of biochemical equations and ODEs corresponding to equivalent interactions in the LR state which we omit here for the sake of brevity.

Binding of gp3 to dimeric integrase already bound to both attL and attR is necessary to mediate excision, restoring attB and attP [11], [13], [17], [20]–[23], [25], [26], [28], [30], [36], [40], [42], [44], [45], [47], [49], [54]

$$D_{\mathrm{LR}}I_4 + R \underset{k_{-11}}{\overset{k_{11}}{\rightleftharpoons}} D_{\mathrm{LR}}I_4R \tag{9}$$

$$D_{\mathrm{LR}}I_4R + R \underset{k_{-12}}{\overset{k_{12}}{\rightleftharpoons}} D_{\mathrm{LR}}I_4R_2 \tag{10}$$

$$D_{\mathrm{LR}}I_4R_2 \xrightarrow{k_R} D_{\mathrm{BP}}I_4R_2 \tag{11}$$

where $D_{\mathrm{LR}}I_4R$ and $D_{\mathrm{LR}}I_4R_2$ denote DNA:protein complexes with two integrase dimers and one/two gp3 monomers bound respectively in the LR state.

In contrast to the strong experimental evidence for each of the above mechanisms, we were unable to find a consensus in the literature regarding three further significant biological details, namely:

1) The directionality of the excision reaction.
2) Gp3 dimerization and subsequent tetramerization in solution.
3) Monomeric integrase binding to DNA substrates.

These properties are all potentially valid mechanisms within a model of DNA recombination, each resulting in a mathematical model with distinct features. In the case of excision reaction directionality, bidirectional excision results in the following equation:

$$D_{\mathrm{LR}}I_4R_2 \underset{k_{-R}}{\overset{k_R}{\rightleftharpoons}} D_{\mathrm{BP}}I_4R_2. \tag{12}$$

Implementing gp3 dimerization and subsequent tetramerization requires a significantly greater number of additional equations, since we now have monomeric gp3 forming dimeric gp3, with these dimers binding to dimeric integrase to form a tetramer in solution, and this tetramer being able to bind directly to DNA attachment sites. The resultant equations are as follows:

$$R + R \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} R_2 \tag{13}$$

$$I_2 + R_2 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} I_2R_2 \tag{14}$$

$$D_{\mathrm{BP}} + I_2R_2 \underset{k_{-6}}{\overset{k_6}{\rightleftharpoons}} D_{\mathrm{BP}}I_2R_2 \tag{15}$$

$$D_{\mathrm{BP}}I_2R_2 + I_2R_2 \underset{k_{-7}}{\overset{k_7}{\rightleftharpoons}} D_{\mathrm{BP}}I_4R_4. \tag{16}$$

This mechanism gives rise to dimeric gp3, R_2, the integrase:gp3 tetramer, I_2R_2, and the DNA:protein complexes $D_{\mathrm{BP}}I_2R_2$ and $D_{\mathrm{BP}}I_4R_4$. Again, there are an equivalent number of new equations corresponding to equivalent interactions in the LR state which we omit for the sake of brevity. We also investigated the performance of a model accounting for gp3 dimerization, but not tetramerization in solution.

Monomeric integrase binding to DNA substrates contributes a further six biochemical equations due to the fact that six intermediate complexes arise from monomeric integrase binding compared to simplistic pairwise, dimeric binding

$$D_{\mathrm{BP}} + I \underset{k_{-13}}{\overset{k_{13}}{\rightleftharpoons}} D_{\mathrm{BP}}I \tag{17}$$

$$D_{\mathrm{BP}}I + I \underset{k_{-14}}{\overset{k_{14}}{\rightleftharpoons}} D_{\mathrm{BP}}I_2 \tag{18}$$

$$D_{\mathrm{BP}}I_2 + I \underset{k_{-15}}{\overset{k_{15}}{\rightleftharpoons}} D_{\mathrm{BP}}I_3 \tag{19}$$

$$D_{\mathrm{BP}}I_3 + I \underset{k_{-16}}{\overset{k_{16}}{\rightleftharpoons}} D_{\mathrm{BP}}I_4 \tag{20}$$

$$D_{\mathrm{BP}}I_2R + I \underset{k_{-17}}{\overset{k_{17}}{\rightleftharpoons}} D_{\mathrm{BP}}I_3R \tag{21}$$

$$D_{\mathrm{BP}}I_3R + I \underset{k_{-18}}{\overset{k_{18}}{\rightleftharpoons}} D_{\mathrm{BP}}I_4R. \tag{22}$$

Again, there are an equivalent number of new equations corresponding to equivalent interactions in the LR state which we omit for brevity.

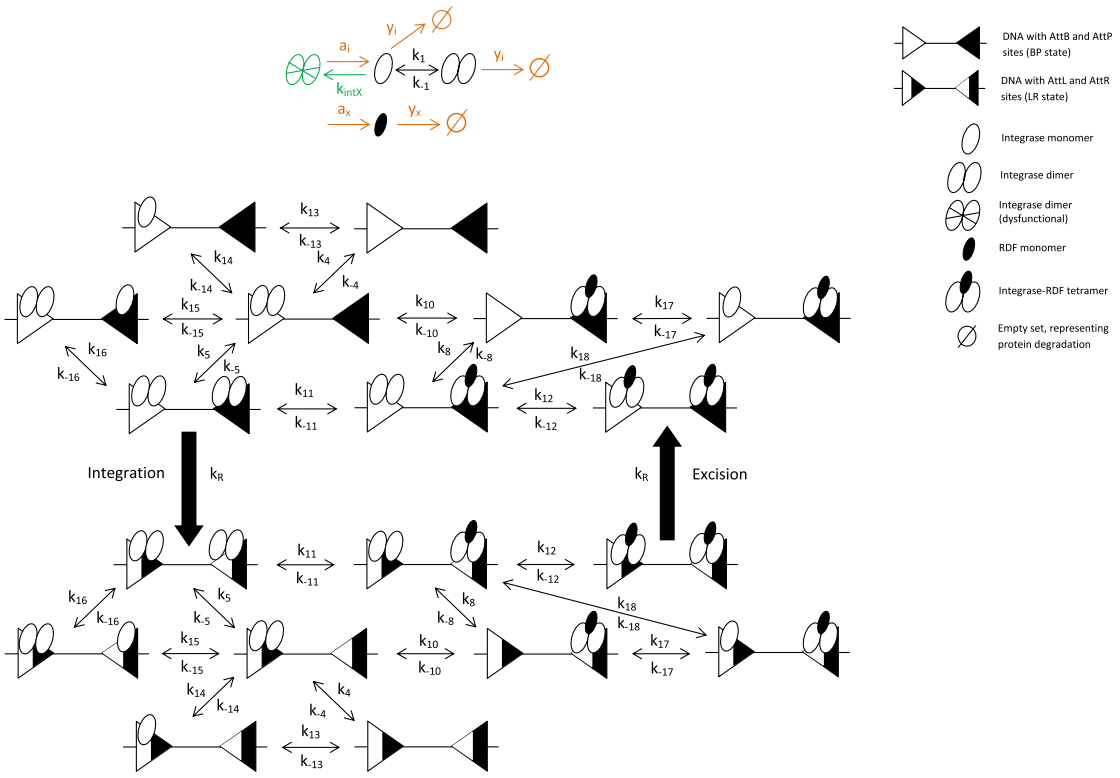

Fig. 2. The DNA recombination reaction network used to derive both our *in vitro* and *in vivo* mechanistic models. Molecular entities, reactions and reaction rate constants common to both models are depicted in black; those depicted in green describe the *in vitro* network and those depicted in orange describe the *in vivo* network. The networks are based on the mechanisms that have been verified in the current experimental literature along with others validated computationally. We model the dynamics of ϕC31 integrase and its RDF, gp3. Reactions and their rate constants are depicted by arrows and their corresponding numbered k. The rate of recombination is k_R. Figure adapted from [12].

We also investigated a variety of other models accounting for combinations of the aforementioned mechanisms as well as alternative gp3 binding mechanisms. Initial testing of all potential models revealed a consistently higher recombination efficiency than that observed in our experimental data. Given that *in vitro* integrase dimerisation can potentially result in the formation of dysfunctional dimers that are unable to bind effectively to DNA att sites, we incorporated a mechanism whereby dysfunctional integrase dimers, I_{2X}, form irreversibly in addition to the reversible formation of functional dimeric integrase

$$I + I \overset{k_{\mathrm{int}X}}{\longrightarrow} I_{2X} \tag{23}$$

where $k_{\mathrm{int}X}$ denotes the rate of dysfunctional integrase dimerization. As expected, this mechanism reduced the concentration of functional dimeric integrase, I_2, and hence overall recombination efficiency, since integrase mediates both recombination reactions.

We tested different models capturing alternative mechanisms implementing each of the above features against our experimental data (see next section and Materials and Methods). The model structure which showed the capability to best match the data is depicted in Fig. 2. Our optimal reaction network consists of a unidirectional excision reaction and monomeric integrase binding, and includes the formation of dysfunctional integrase dimers and 2 : 1 stoichiometry of the synaptic complexes. When versions of the model incorporating gp3 dimerization and subsequent tetramerization in solution were optimized against the experimental data, the minimum error observed between the data and model outputs was larger than that observed for models that do not account for the same mechanisms. We therefore omitted these mechanisms from the model. The network is comprised of twenty-two distinct molecular entities and twenty-eight reactions, thus generating a system of twenty-two ODEs and twenty-eight parameters (the full list of model ODEs can be found in the Supplementary Material).

The efficiency of the RAD module to switch from one DNA state to the other is taken to be the concentration of free DNA and DNA complexes in the final state as a percentage of the concentration of free DNA in the initial state. That is, we analyze the total register of the system in the DNA state of interest. This simply involves summing the ODEs corresponding to all DNA-based molecular entities of the same DNA state. Summing each set of nine ODEs corresponding to each DNA state gives two ODEs describing the dynamics of the total register of the system in BP state, D_{BPtot}, and LR state, D_{LRtot}

$$\frac{dD_{\mathrm{BPtot}}}{dt} = -k_R D_{\mathrm{BP}} I_4 + k_R D_{\mathrm{LR}} I_4 R_2, \tag{24}$$

$$\frac{dD_{\mathrm{LRtot}}}{dt} = k_R D_{\mathrm{BP}} I_4 - k_R D_{\mathrm{LR}} I_4 R_2. \tag{25}$$

Our computational model simulations are the numerical solutions to (24) and (25) and are converted to a percentage of the initial concentration of DNA to demonstrate switching efficiency. The total DNA register is calculated in this fashion

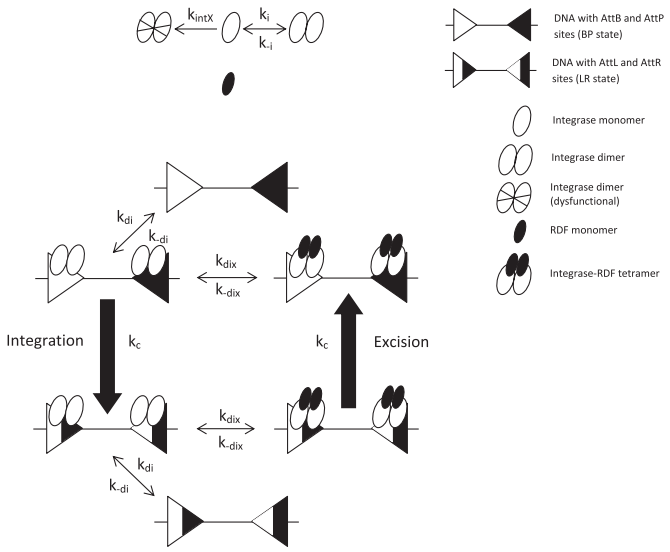

Fig. 3. The DNA recombination reaction network adapted from [11]. The exclusion of SSR expression and degradation coupled with the inclusion of dysfunctional integrase dimerisation accounts for our *in vitro* experimental conditions.

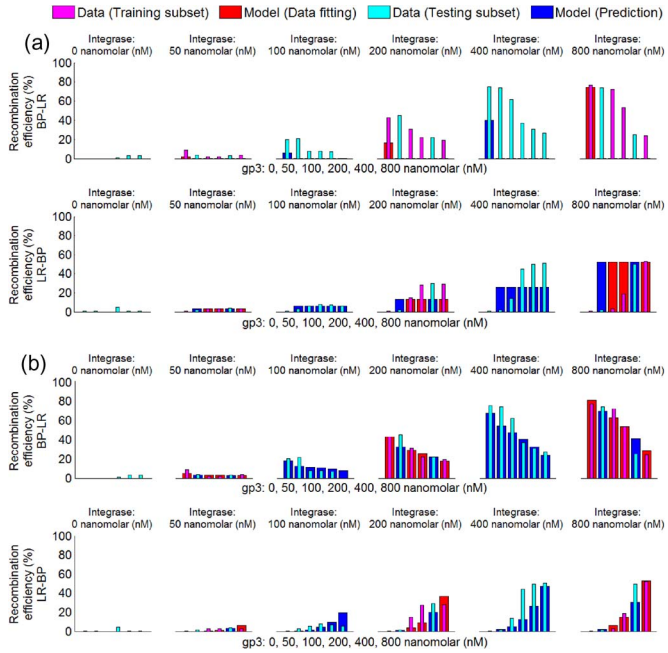

Fig. 4. (a) Data fitting/prediction results for the model of [11]. (b) Data fitting/prediction results for our mechanistic model. In both (a) and (b) the top row of bar graphs represents the integration reaction and the bottom row of bar graphs represents the excision reaction. The wider bars represent model simulations and the thinner bars represent data. Figure adapted from [12].

for all versions of the recombination network tested. In each case the full system of ODEs is solved numerically in order to determine the total register, that is, no attempt is made to reduce complexity as we look to retain mechanistic detail.

III. Model Matching and Validation Against *In Vitro* Data

We compared our model to the other existing model of *in vivo* DNA recombination proposed in [11] for its ability to match and predict a new set of *in vitro* data on dynamic and steady-state recombination efficiency in the presence of different concentrations of integrase and gp3 (see Materials and Methods). The model in [11] is derived from a simple reaction network comprised of nine molecular entities and is void of considerable mechanistic detail such as monomeric integrase binding, the intermediate complexes arising from individual dimeric integrase and monomeric gp3 binding, and a 2 : 1 integrase:gp3 complex stoichiometry.

To ensure the validity of the model comparisons, we adapted the model of [11] to the *in vitro* context and also imposed the same dysfunctional integrase dimerisation mechanism from our optimal model (Fig. 3). Optimal model performance was evaluated by using a Genetic Algorithm to optimize an error function defined to capture the difference between model outputs and our experimental data on *in vitro* steady-state recombination efficiency for both integration and excision reactions (see Materials and Methods).

Six distinct initial concentrations (0, 50, 100, 200, 400 and 800 nM) of integrase and gp3 give the thirty-six pairs of initial concentrations used experimentally to record the steady-state recombination efficiency of the system. This was performed for both the integration and excision reactions, giving an experimental dataset of seventy-two values. The efficiencies are given as a percentage of the initial concentration of free DNA which was set at 10 nM. Time course data was also available whereby the recombination efficiency of both reactions was recorded at ten time points (0, 1, 2, 4, 8, 16, 32, 64, 128 and 180 mins) for two distinct pairs of initial concentrations of integrase and gp3 (800 nM integrase, 0 nM gp3 and 400 nM integrase, 0 nM gp3). Since our dataset was relatively large, we used a subset of the data to define our GA error function with the remaining data used to evaluate the predictive capabilities of our models. The subset used in data fitting was chosen to capture the full spectrum of recombination efficiencies, and all models were optimised against the same subset of experimental data and within the same parameter space.

Fig. 4 shows the data fitting and data prediction capabilities of both models. It is clear that the model of [11] is unable to accurately match the subset of steady-state data used (Fig. 4(a)). In the case of the integration (BP-LR) reaction, simulations appear to be accurate for the relatively low concentration of integrase (50 nM) however, as the concentration of integrase increases, accurate fits can only be found for 800 nM integrase, 0 nM gp3. In fact, the simple model is only capable of simulating negligible recombination efficiencies for non-zero concentrations of gp3, regarding the integration reaction. This may appear to be an intuitive result given that integrase alone mediates integration, however, our data clearly indicates that the system can achieve high integration efficiencies in the presence of both SSRs. Similarly for the excision (LR-BP) reaction, simulations appear to be accurate for 50 nM integrase, but are unable to match the majority of data as integrase concentration increases. In fact, the model is only capable of simulating negligible recombination efficiencies for 0 nM gp3 and uniform efficiencies for all non-zero gp3 concentrations, regarding the

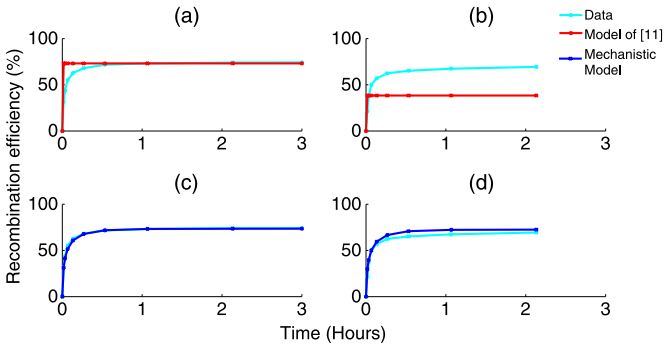

Fig. 5. Data fitting results for the model of [11] ((a), (b), red) and our mechanistic model ((c), (d), blue) against time course data (cyan). Model simulations are plotted against two integration reaction time course data sets, one initiated with 800 nanomolar integrase (a), (c) and the other with 400 nanomolar integrase (b), (d). Figure adapted from [12].

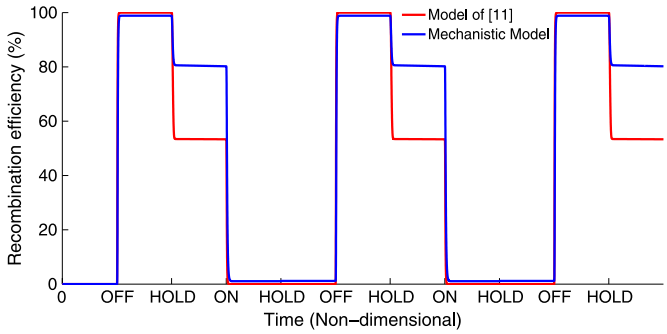

Fig. 6. RAD module *in vivo* switching efficiencies for the model of [11] and our mechanistic model. Both models simulate 2.5 repeated OFF-HOLD-ON-HOLD operative cycles. All non-dimensional parameter values for both models are set equal to 1 to simulate these plots with the exception of a_i and a_x. For the ON operation $a_i = 10$, $a_x = 0.1$; the OFF operation $a_i = a_x = 10$; the HOLD operation $a_i = \mathrm{a}_x = 0.1$. The model of [11] is plotted in red with our mechanistic model plotted in blue.

excision reaction. The former observation is intuitive, since gp3 is required in combination with integrase to mediate excision, but we have no logical reason to justify the latter. With regards to prediction, we observe the same trends for both reactions meaning that the model of [11] is void of the predictive qualities required of a useful design tool.

In contrast, our mechanistic model clearly provides a strong fit to the subset of steady-state data used in the GA global optimization (Fig. 4(b)). In the case of both reactions, we observe accurate replication of the majority of data values. The model also predicts the remaining data values effectively, presenting a clear validation of the mechanistic structure we have developed and the potential power of our model as a design tool. We note that this result may not be unique to a single optimal parameter set; multiple "optimal" parameter sets may exist, capable of matching our data to a similar degree, depending on the constraints imposed upon the parameter space. Future work will examine the prevalence and biological plausibility of optimal parameter sets through Bayesian model selection and parameter inference techniques. Similar trends are observed when we compare time course integration simulations (Fig. 5). For 800 nM integrase, 0 nM gp3 the model of [11] replicates the overall efficiency, but does not perform to the same extent for 400 nM integrase, 0 nM gp3. In both cases, the model exhibits a step-like response which does not match the gradual response recorded experimentally. Again, our mechanistic model shows much improved performance, capturing the final efficiencies as well as the appropriate dynamical response in both cases.

IV. Modeling the *In Vivo* System

Having validated our mechanistic model against *in vitro* data, we sought to analyze model performance within an *in vivo* context. All available experimental data leads us to believe that the DNA:protein binding interactions of the *in vitro* system are retained *in vivo*, with the introduction of protein expression and degradation representing the key adaptations. Thus, the mechanisms removed from the model of [11] in order to analyze *in vitro* dynamics were restored and we adapted our own mechanistic model accordingly (Fig. 2). Increased expression/degradation rates of the SSRs are induced to realise desired RAD module operations and hence we also account for basal expression/degradation rates occurring in the absence of induction. SSR induction is performed practically through chemical stimuli. The formation of dysfunctional dimeric integrase is removed from both models as we have no reason to justify its existence in this context. This reduces the number of ODEs in our mechanistic model to twenty-one since I_{2X} is no longer present. All *in vivo* modeling utilizes the model of [11] and our mechanistic model developed previously in their non-dimensional forms (see Materials and Methods), in order to permit valid numerical simulations and direct mathematical comparisons [11], [59]. The main focus of our *in vivo* investigation is the excision reaction, since the integration reaction is straightforward to elucidate; over-expression of integrase is guaranteed to induce highly efficient integration for relatively low gp3 expression since integration is mediated by integrase only. However, the excision reaction is nuanced by its mediation by both SSRs and directly influences the functionality of the RAD module as a result. A brute-force approach of over-expressing integrase and gp3 simultaneously to achieve highly efficient excision is an intuitive notion before considering that desirable RAD module functionality will often require hold states. That is, once excision has been induced through simultaneous over-expression, it is likely that spontaneous re-integration will occur due to high residual integrase concentration and gp3 dissociation. Therefore it is not conceptually obvious how to induce highly efficient excision and then hold that state equally efficiently in the absence of induction. Fig. 6 depicts the *in vivo* RAD module dynamics for 2.5 repeated OFF-HOLD-ON-HOLD operative cycles for both the simple model and our detailed model. We define RAD module operations as follows: ON is an integration reaction induced through increased integrase levels only, OFF is an excision reaction induced through simultaneously increased integrase and gp3 levels and HOLD is the restoration of basal SSR levels following either ON or OFF operations. Both models exhibit consistent switching efficiencies across each of their own repeated cycles which demonstrates that the module can maintain performance over many identical induction events. That is, the process of inducing desired SSR expression/degradation should permit efficient

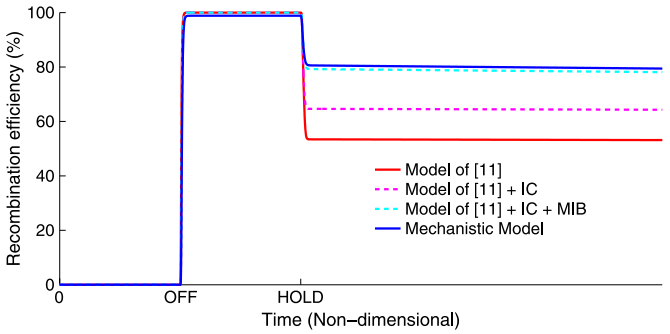

Fig. 7. RAD module *in vivo* HOLD state efficiencies for the model of [11] (red) and our mechanistic model (blue). The remaining colored plots demonstrate the improvements on HOLD state efficiency made by each distinction between the two models. Intermediate complexes and monomeric integrase binding are abbreviated to IC and MIB respectively. In each case an OFF-HOLD operative cycle is simulated, that is, $a_i = a_x = 10$ followed by $a_i = a_x = 0.1$. All remaining non-dimensional parameters are set equal to 1.

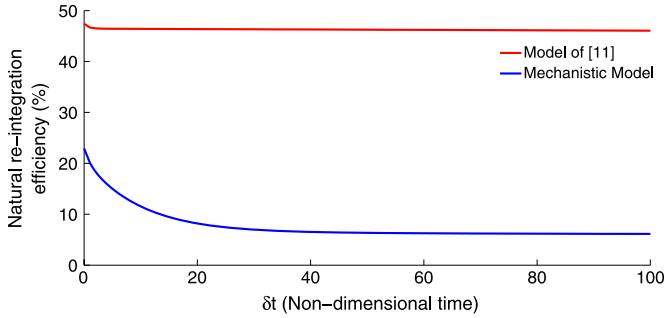

Fig. 8. RAD module *in vivo* natural re-integration efficiencies for the model of [11] and our mechanistic model. Natural re-integration efficiency is plotted against the time interval between integrase cessation and gp3 cessation, δt. In each case an OFF-HOLD operative cycle is simulated, that is, $a_i = a_x = 10$ followed by $a_i = a_x = 0.1$. All remaining non-dimensional parameters are set equal to 1.

module operations whenever required, and regardless of switching the module ON or OFF. Both models hold state efficiently following an integration reaction since basal SSR levels are restored in the absence of induction and hence there is insufficient gp3 to mediate natural re-excision. There is, however, a notable distinction between the performance of the models in the efficiency of the HOLD state following the induction of an excision reaction. We observe natural re-integration efficiencies of $\sim$47% for the model of [11] against $\sim$23% for our mechanistic model. This suggests that the efficiency of RAD module HOLD states following excision is, in fact, greater than expected from the model of [11]. To determine the reasons for this, we analyze the mechanistic distinctions between the two models. Compared to the simple model, the detailed model accounts for the following additional mechanisms:

1) Monomeric integrase binding to free DNA substrates.
2) Formation of synaptic complexes with 2:1 stoichiometry.
3) Formation of intermediate DNA:protein complexes.

Fig. 7 shows the effect of these biological distinctions on HOLD state efficiency following excision. We apply each distinction to the model of [11] cumulatively to observe their influence on HOLD state efficiency. The addition of intermediate complexes is shown to reduce natural re-integration efficiency by $\sim$11% and thus improves HOLD state efficiency. The addition of monomeric integrase binding is shown to further reduce natural re-integration efficiency by $\sim$12%. We therefore infer that the 2:1 integrase:gp3 stoichiometry accounts for the remaining $\sim$1% that separates the re-integration efficiencies of the two models. With a contribution of $\sim$1%, the stoichiometry of synaptic complexes follows the previously observed trend of having minimal effect on RAD module dynamics. In contrast, monomeric integrase binding and intermediate DNA:protein complexes provide the vast majority of improvement in HOLD state efficiency for the detailed model, in almost equal measure. We note here that the formation of intermediate DNA:protein complexes is, in part, due to the monomeric integrase binding, however, these complexes also arise from our pairwise dimeric integrase binding and monomeric gp3 binding. The disadvantage of minimal intermediate complexes and protein binding pathways is clear with regard to the model of [11]. Following an efficient OFF operation, the concentration of SSRs is restored to a basal level. At this point, there is insufficient protein to hold the system in the $D_{\mathrm{BP}}I_4R_4$ complex and the first interaction that can possibly occur is the dissociation of gp3. This dissociation immediately produces the $D_{\mathrm{BP}}I_4$ complex which is then able to re-integrate. By contrast, the transition from the $D_{\mathrm{BP}}I_4R_2$ complex to the $D_{\mathrm{BP}}I_4$ complex is not as straightforward in our mechanistic model. In the absence of induction, gp3 dissociation produces the intermediate $D_{\mathrm{BP}}I_4R$ complex which itself can potentially give rise to three other complexes, only one of which, $D_{\mathrm{BP}}I_4$, would then be able to re-integrate.

We ideally require the RAD module to function at 100% efficiency for all three operations. Our results indicate that the efficiency of a HOLD following an OFF switch is proportional to the maintenance of the $D_{\mathrm{BP}}I_4R_2$ complex. This could be problematic when executing a regime whereby increased SSR expression is both induced and relinquished in a simultaneous manner. Alternatively, we examine an approach whereby the induction of integrase and gp3 ceases at separate time points. Ceasing induction of gp3 prior to that of integrase is illogical given that gp3 is required to maintain the $D_{\mathrm{BP}}I_4R_2$ complex and prolonged induction of integrase will only facilitate greater re-integration efficiency. Therefore we investigate the effect of ceasing integrase induction prior to that of gp3 on HOLD state efficiency. Fig. 8 depicts this effect in the form of a plot of natural re-integration efficiency against increasing time intervals, δt, between ceasing integrase induction and gp3 induction. The performance of both the model of [11] and our mechanistic model is plotted, with each y-intercept representing the $\sim$47% and $\sim$23% natural re-integration efficiencies for simultaneous cessation ($\delta t = 0$), respectively. Regarding the model of [11], as δt increases we observe a minimal reduction in natural re-integration efficiency. This suggests that, although delaying the cessation of gp3 provides a small improvement, any gp3 dissociation that occurs during prolonged induction still provokes an almost immediate re-integration given the inherent transitioning to the $D_{\mathrm{BP}}I_4$ complex. However, in the case of our mechanistic model we observe a significant reduction in natural re-integration efficiency as δt increases. The dissociation of gp3 gives rise to the intermediate $D_{\mathrm{BP}}I_4R$ complex and, with

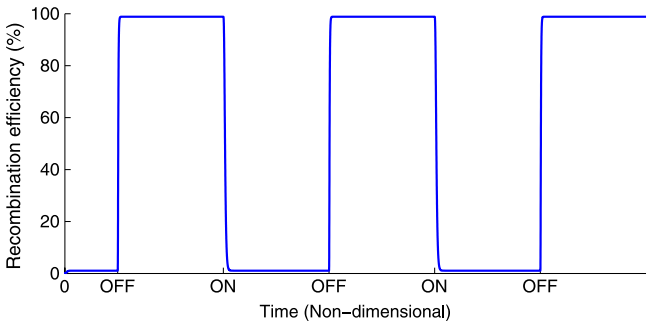

Fig. 9. RAD module *in vivo* switching efficiencies for our mechanistic model. We simulate an OFF-ON-OFF-ON-OFF operative cycle. For the ON operation $a_x = 0.1$ and for the OFF operation $a_x = 10$. Throughout the operative cycle, $a_i = 10$. All remaining non-dimensional parameter values are set equal to 1.

prolonged gp3 induction, the system is therefore weighted in favour of the transition to the $D_{\mathrm{BP}}I_2R$ and $D_{\mathrm{BP}}I_3R$ complexes as well as the reformation of the $D_{\mathrm{BP}}I_4R_2$ complex.

We note that regimes incorporating induction cessation intervals eliminate the HOLD state from the RAD module operative cycle. This may not be desirable for applications regarding biological data storage that are dependent on lasting responses to transient stimuli. However, this may assist in the development of other potential RAD module applications that are not as reliant on HOLD states, such as medical treatments for diseases related to the inheritance of cellular states. We have established that prolonged gp3 induction is capable of reducing re-integration efficiency and thus improving functionality given that we have considered natural re-integration to be synonymous with spontaneous switching and ultimately dysfunctionality of the module. However, if we neglect dependency on HOLD states, then harnessing natural re-integration can provide very simple and highly efficient RAD module functionality. Fig. 9 depicts the dynamical response of the RAD module for 2.5 repeated OFF-ON operative cycles. Here the ON, OFF operations are defined as the cessation of gp3 induction and the induction of gp3 respectively; integrase induction remains constant throughout. Since both operations are mediated, either solely or in part, by integrase, the state of the system is dependent only on gp3 concentration. That is, when there is no induction of gp3 the constant induction of integrase causes a fully efficient ON switch which will remain until gp3 induction causes a fully efficient OFF switch. Induction of gp3 must last for the duration of the desired OFF switch, at which point cessation of gp3 induction is sufficient to cause another fully efficient ON switch through natural re-integration.

V. Conclusion

We have described the development of a detailed mechanistic model of a rewritable recombinase addressable data module, based on a wide-ranging synthesis of available experimental data on the biomolecular interactions underlying DNA recombination. We demonstrated the capability of this model to match and predict *in vitro* experimental data on recombination efficiencies across a range of different concentrations of integrase and gp3, thus validating its usefulness as a design tool for building future synthetic circuitry. Investigation of *in vivo* recombination dynamics revealed the importance of fully accounting for all mechanistic details in models of DNA recombination, in order to accurately predict the effect of different switching strategies on RAD module performance.

VI. Materials and Methods

A. Experimental Procedure

Proteins (ϕC31 integrase and gp3) were purified as described in [20], [28], and [31]. Integrase and gp3 were diluted in integrase dilution buffer [25 mM Tris · HCl (pH 7.5), 1 mM DTT, 1 M NaCl, and 50% (vol/vol) glycerol] and kept at -20 °C. Substrate plasmids containing inverted repeat recombination sites (pZJ56off with attB and attP; pZJ56on with attR and attL) used for *in vitro* assay were prepared from *E. coli* DH5, using a plasmid mini-prep kit (Qiagen). DNA concentrations were determined by measuring the absorbance at 260 nm.

In a typical reaction, premixed integrase and gp3 with 10× their final concentrations were added to a solution of substrate plasmid (∼10 nM) in a reaction buffer [50 mM Tris · HCl (pH 7.5), 0.1 mM EDTA, 5 mM spermidine, and 0.1 mg/ml BSA]. Reactions were carried out at 30 °C, terminated at various time points, by heating the samples to 80 °C for 10 min. Samples were digested with restriction enzymes, then, treated with 5 μl of loading buffer [25 mM Tris · HCl (pH 8.2), 20% (wt/vol) Ficoll, 0.5% sodium dodecyl sulphate, 5 mg/ml protease K, 0.25 mg/ml bromophenol blue] at 37 °C for 30 min prior to loading onto 1.2% (wt/vol) agarose gels. Gels were stained with ethidium bromide, destained in electroporation buffer, and photographed using Bio-Rad UV Transilluminator. Recombinant and non-recombinant DNA bands were quantitated using the volume analysis tool of Quantity One software.

B. Non-Dimensionalzation

While the exact values of the kinetic parameters in the models considered are unknown, we can still understand the general features of the RAD module behaviour by non-dimensionalizing all concentration and time units. The non-dimensionalization process involves identifying model parameters that have the same dimension as the dependent and independent variables, then rewriting the system in terms of these parameters. This is achieved through equating the dimensionality of each individual term in the system of ODEs and solving for each rate constant. The overall number of model parameters is decreased by one as a consequence. See Supplementary Material for a demonstration of non-dimensionalization and a full listing of our models in non-dimensional form.

C. Global Optimization

We employed the Genetic Algorithm function in MATLAB in order to optimize our models against the experimental data. For each model the reaction rate constants k_i were chosen as optimization variables. The GA mimics natural selection; converging to the global minimum within the allocated parameter space by evolving an initial population of randomly generated solutions over a large number of generations. The probability

of obtaining the global optimum solution is maximised by selecting the largest population size and number of generations possible. However, increasing the computational workload in this manner also greatly increases the time frame required for the algorithm to converge. Establishing an effective compromize is key for successful deployment. After a number of trials, the following parameter values for the GA were chosen:

1) Population: 100
2) Generations: 1000
3) Bounds imposed on parameter values: $[10^{-3}, 1000]$.

A population size of 100 was selected for the vast majority of optimizations however, in cases where the number of model parameters was significantly increased, we increased this figure to ensure that the population:parameters ratio never exceeded 3 : 1. We selected a particularly large parameter space due to our focus on establishing optimal model performance in light of the lack of documentation regarding reaction rates. The GA runs the given model under the same conditions used experimentally, with the resulting *in silico* recombination efficiencies subtracted from the *in vitro* data values to give a matrix of error values. We then take the absolute value of each matrix entry and then calculate the total sum to give an overall error, E, specifically

$$E = \sum_{j=1}^{4}\sum_{i=1}^{3}|p_{ij}| + \sum_{j=1}^{4}\sum_{i=1}^{3}|q_{ij}|, \quad p_{ij} \in P, \quad q_{ij} \in Q \quad (26)$$

where $P = S_{\mathrm{BP}} - D_{\mathrm{BP}}$ and $Q = S_{\mathrm{LR}} - D_{\mathrm{LR}}$, i.e., P and Q are the 3×4 matrices calculated by subtracting the matrices of data values $(D_{\mathrm{BP}}, D_{\mathrm{LR}})$ from the matrices of model simulations $(S_{\mathrm{BP}}, S_{\mathrm{LR}})$ for the BP and LR reactions respectively. The matrices are comprised of twelve elements since our data subset contains twelve values for each reaction, and hence the model performs twelve corresponding simulations. The end result is a set of model parameters that gave rise to the minimum overall error.

REFERENCES

[1] R. Danial, S. Woo, L. Turicchia, and R. Sarpeshkar, "Analog transistor models of bacterial genetic circuits," in *Proc. IEEE Biological Circuits and Systems Conf.*, San Diego, CA, USA, Nov. 2011, pp. 333–336.

[2] T. Prescott and A. Papachristodoulou, "Designing conservation relations in layered synthetic biomolecular networks," *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 4, pp. 572–580, Aug. 2015.

[3] P. Wang, X. Yu, and J. Lu, "Identification and evolution of structurally dominant nodes in protein-protein interaction networks," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 1, pp. 87–97, Feb. 2014.

[4] M. Hajimorad and J. Keasling, "Independence among synthetic genetic devices in the bacterium Escherichia coli extends to the time-domain," in *Proc. IEEE Biomedical Circuits and Systems Conf.*, Oct. 22–24, 2014, pp. 89–92.

[5] H. Chiang, J. Jiang, and F. Fagesy, "Building reconfigurable circuitry in a biochemical world," in *Proc. IEEE Biomedical Circuits and Systems Conf.*, Oct. 22–24, 2014, pp. 560–563.

[6] T. Gardner, C. Cantor, and J. Collins, "Construction of a genetic toggle switch in *Escherichia coli*," *Nature*, vol. 403, no. 6767, pp. 339–342, Jan. 20, 2000.

[7] N. Grindley, K. Whiteson, and P. Rice, "Mechanisms of site-specific recombination," *Annu. Rev. Biochem.*, vol. 75, pp. 567–605, 2006.

[8] A. Landy, "The λ integrase site-specific recombination pathway," *Microbiol. Spectr.*, vol. 3, no. 2, Apr. 2015, doi: 10.1128/microbiolspec.MDNA3-0051-2014.

[9] G. van Duyne, "A structural view of cre-loxp site-specific recombination," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 30, pp. 87–104, 2001.

[10] S. Colloms, C. Merrick, F. Olorunniji, M. Stark, M. Smith, A. Osbourn, J. Keasling, and S. Rosser, "Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination," *Nucl. Acids Res.*, vol. 42, no. 4, p. e23, Feb. 2014, doi: 10.1093/nar/gkt1101.

[11] J. Bonnet, P. Subsoontorn, and D. Endy, "Rewritable digital data storage in live cells via engineered control of recombination directionality," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 23, pp. 8884–8889, Jun. 5, 2012, doi: 10.1073/pnas.1202344109.

[12] J. Bowyer, J. Zhao, S. Rosser, S. Colloms, and D. Bates, "Development and experimental validation of a mechanistic model of *in vitro* DNA recombination," in *Proc. 37th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Soc.*, Milano, Italy, 2015, pp. 945–948.

[13] W. Brown, N. Lee, Z. Xu, and M. Smith, "Serine recombinases as tools for genome engineering," *Methods*, vol. 53, no. 4, pp. 372–379, Apr. 2011, doi: 10.1016/j.ymeth.2010.12.031.

[14] P. Fogg, S. Colloms, S. Rosser, M. Stark, and M. Smith, "New applications for phage integrases," *J. Mol. Biol.*, vol. 426, no. 15, pp. 2703–2716, Jul. 29, 2014, doi: 10.1016/j.jmb.2014.05.014.

[15] P. Ghosh, N. Pannunzio, and G Hatfull, "Synapsis in phage Bxb1 integration: Selection mechanism for the correct pair of recombination sites," *J. Mol. Biol.*, vol. 349, no. 2, pp. 331–348, Jun. 3, 2005.

[16] P. Ghosh, L. Bibb, and G. Hatfull, "Two-step site selection for serine-integrase-mediated excision: DNA-directed integrase conformation and central dinucleotide proofreading," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 9, pp. 3238–3243, Mar. 4, 2008, doi: 10.1073/pnas.0711649105.

[17] A. Groth and M. Calos, "Phage integrases: Biology and applications," *J. Mol. Biol.*, vol. 335, no. 3, pp. 667–678, Jan. 16, 2004.

[18] M. Gupta, R. Till, and M. Smith, "Sequences in attB that affect the ability of phiC31 integrase to synapse and to activate DNA cleavage," *Nucl. Acids Res.*, vol. 35, no. 10, pp. 3407–3419, 2007.

[19] R. Keenholtz, S. Rowland, M. Boocock, M. Stark, and P. Rice, "Structural basis for catalytic activation of a serine recombinase," *Structure*, vol. 19, no. 6, pp. 799–809, Jun. 8, 2011, doi: 10.1016/j.str.2011.03.017.

[20] T. Khaleel, E. Younger, A. McEwan, A. Varghese, and M. Smith, "A phage protein that binds $\phi C31$ integrase to switch its directionality," *Mol. Microbiol.*, vol. 80, no. 6, pp. 1450–1463, 2011.

[21] S. Liu, J. Ma, W. Wang, M. Zhang, Q. Xin, S. Peng, R. Li, and H. Zhu, "Mutational analysis of highly conserved residues in the phage phiC31 integrase reveals key amino acids necessary for the DNA recombination," *PLOS ONE*, vol. 5, no. 1, Jan. 25, 2010, Art. no. e8863, doi: 10.1371/journal.pone.0008863.

[22] I. Lucet, F. Tynan, V. Adams, J. Rossjohn, D. Lyras, and J. Rood, "Identification of the structural and functional domains of the large serine recombinase TnpX from Clostridium perfringens," *J. Biol. Chem.*, vol. 280, no. 4, pp. 2503–2511, Jan. 28, 2005.

[23] S. Mandali, G. Dhar, N. Avliyakulov, M. Haykinson, and R. Johnson, "The site-specific integration reaction of Listeria phage A118 integrase, a serine recombinase," *Mob DNA*, vol. 4, no. 1, pp. 2, Jan. 3, 2013, doi: 10.1186/1759-8753-4-2.

[24] A. McEwan, P. Rowley, and M. Smith, "DNA binding and synapsis by the large C-terminal domain of phiC31 integrase," *Nucl. Acids Res.*, vol. 37, no. 14, pp. 4764–4773, Aug. 2009, doi: 10.1093/nar/gkp485.

[25] A. McEwan, A. Raab, S. Kelly, J. Feldmann, and M. Smith, "Zinc is essential for high-affinity DNA binding and recombinase activity of C31 integrase," *Nucleic Acids Res*, vol. 39, no. 14, pp. 6137–6147, Aug. 2011, doi: 10.1093/nar/gkr220.

[26] T. Miura, Y. Hosaka, Y. Yan-Zhuo, T. Nishizawa, M. Asayama, H. Takahashi, and M. Shirai, "*In vivo* and *in vitro* characterization of site-specific recombination of actinophage R4 integrase," *J. Gen. Appl. Microbiol.*, vol. 57, no. 1, pp. 45–57, 2011.

[27] F. Olorunniji and M. Stark, "Catalysis of site-specific recombination by Tn3 resolvase," *Biochem. Soc. Trans.*, vol. 38, no. 2, pp. 417–21, Apr. 2010, doi: 10.1042/BST0380417.

[28] F. Olorunniji, D. Buck, S. Colloms, A. McEwan, M. Smith, M. Stark, and S. Rosser, "Gated rotation mechanism of site-specific recombination by C31 integrase," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 48, pp. 19 661–19 666, Nov. 27, 2012, doi: 10.1073/pnas.1210964109.

[29] P. Rowley and M. Smith, "Role of the N-terminal domain of phiC31 integrase in attB-attP synapsis," *J. Bacteriol.*, vol. 190, no. 20, pp. 6918–6921, Oct. 2008, doi: 10.1128/JB.00612-08.

[30] S. Singh, P. Ghosh, and G. Hatfull, "Attachment site selection and identity in Bxb1 serine integrase-mediated site-specific recombination," *PLoS Genet*, vol. 9, no. 5, May 2013, Art. no. e1003490, doi: 10.1371/journal.pgen.1003490.

[31] M. Smith, R. Till, K. Brady, P. Soultanas, H. Thorpe, and M. Smith, "Synapsis and DNA cleavage in phiC31 integrase-mediated site-specific recombination," *Nucl. Acids Res.*, vol. 32, no. 8, pp. 2607–2617, May 11, 2004.

[32] M. Smith, W. Brown, A. McEwan, and P. Rowley, "Site-specific recombination by phiC31 integrase and other large serine recombinases," *Biochem. Soc. Trans.*, vol. 38, no. 2, pp. 388–394, Apr. 2010, doi: 10.1042/BST0380388.

[33] M. Stark, M. Boocock, F. Olorunniji, and S. Rowland, "Intermediates in serine recombinase-mediated site-specific recombination," *Biochem. Soc. Trans.*, vol. 39, no. 2, pp. 617–622, Apr. 2011, doi: 10.1042/BST0390617.

[34] H. Thorpe, S. Wilson, and M. Smith, "Control of directionality in the site-specific recombination system of the Streptomyces phage phiC31," *Mol. Microbiol.*, vol. 38, no. 2, pp. 232–241, Oct. 2000.

[35] P. Yuan, K. Gupta, and G. van Duyne, "Tetrameric structure of a serine integrase catalytic domain," *Structure*, vol. 16, no. 8, pp. 1275–1286, Aug. 6 2008, doi: 10.1016/j.str.2008.04.018.

[36] L. Zhang, L. Wang, J. Wang, X. Ou, G. Zhao, and X. Ding, "DNA cleavage is independent of synapsis during streptomyces phage phiBT1 integrase-mediated site-specific recombination," *J. Mol. Cell Biol.*, vol. 2, no. 5, pp. 264–275, Oct. 2010, doi: 10.1093/jmcb/mjq025.

[37] H. Bai, M. Sun, P. Ghosh, G. Hatfull, N. Grindley, and J. Marko, "Single-molecule analysis reveals the molecular bearing mechanism of DNA strand exchange by a serine recombinase," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 18, pp. 7419–7424, May 3, 2011, doi: 10.1073/pnas.1018436108.

[38] L. Bibb, M. Hancox, and G. Hatfull, "Integration and excision by the large serine recombinase phiRv1 integrase," *Mol. Microbiol.*, vol. 55, no. 6, pp. 1896–1910, Mar. 2005.

[39] A. Breuner, L. Brondsted, and K. Hammer, "Novel organization of genes involved in prophage excision identified in the temperate lactococcal bacteriophage TP901-1," *Bacteriol*, vol. 181, no. 23, pp. 7291–7297, Dec. 1999.

[40] A. Breuner, L. Brondsted, and K. Hammer, "Resolvase-like recombination performed by the TP901-1 integrase," *Microbiology*, vol. 147, pp. 2051–2063, Aug. 2001, Pt 8.

[41] E. Cho, R. Gumport, and J. Gardner, "Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex," *J. Bacteriol.*, vol. 184, no. 18, pp. 5200–5203, Sep. 2002.

[42] P. Combes, R. Till, S. Bee, and M. Smith, "The streptomyces genome contains multiple pseudo-attB sites for the (phi)C31-encoded site-specific recombination system," *J. Bacteriol.*, vol. 184, no. 20, pp. 5746–5752, Oct. 2002.

[43] A. Keravala, J. Portlock, J. Nash, D. Vitrant, P. Robbins, and M. Calos, "PhiC31 integrase mediates integration in cultured synovial cells and enhances gene expression in rabbit joints," *J. Gene Med.*, vol. 8, no. 8, pp. 1008–1017, Aug. 2006.

[44] A. Kim, P. Ghosh, M. Aaron, L. Bibb, S. Jain, and G. Hatfull, "Mycobacteriophage Bxb1 integrates into the mycobacterium smegmatis groEL1 gene," *Mol. Microbiol.*, vol. 50, no. 2, pp. 463–473, Oct. 2003.

[45] J. Lewis and G. Hatfull, "Identification and characterization of mycobacteriophage L5 excisionase," *Mol. Microbiol.*, vol. 35, no. 2, pp. 350–360, Jan. 2000.

[46] M. Matsuura, T. Noguchi, D. Yamaguchi, T. Aida, M. Asayama, H. Takahashi, and M. Shirai, "The sre gene (ORF469) encodes a site-specific recombinase responsible for integration of the R4 phage genome," *J. Bacteriol.*, vol. 178, no. 11, pp. 3374–6337, Jun. 1996.

[47] L. Nkrumah, R. Muhle, P. Moura, P. Ghosh, G. Hatfull, W. Jacobs, and D. Fidock, "Efficient site-specific integration in Plasmodium falciparum chromosomes mediated by mycobacteriophage Bxb1 integrase," *Nat. Methods*, vol. 3, no. 8, pp. 615–621, Aug. 2006.

[48] E. Olivares, R. Hollis, and M. Calos, "Phage R4 integrase mediates site-specific integration in human cells," *Gene*, vol. 278, no. 1/2, pp. 167–176, Oct. 31, 2001.

[49] C. Pena, J. Kahlenberg, and G. Hatfull, "Protein-DNA complexes in mycobacteriophage L5 integrative recombination," *J. Bacteriol.*, vol. 181, no. 2, pp. 454–461, Jan. 1999.

[50] K. Rutherford, P. Yuan, K. Perry, R. Sharp, and G. van Duyne, "Attachment site recognition and regulation of directionality by the serine integrases," *Nucl. Acids Res.*, vol. 41, no. 17, pp. 8341–8356, Sep. 2013, doi: 10.1093/nar/gkt580.

[51] J. Thomson and D. Ow, "Site-specific recombination systems for the genetic manipulation of eukaryotic genomes," *Genesis*, vol. 44, no. 10, pp. 465–476, Oct. 2006.

[52] B. Thyagarajan, E. Olivares, R. Hollis, D. Ginsburg, and M. Calos, "Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase," *Mol. Cell Biol.*, vol. 21, no. 12, pp. 3926–3934, Jun. 2001.

[53] G. van Duyne and K. Rutherford, "Large serine recombinase domain structure and attachment site binding," *Crit. Rev. Biochem. Mol. Biol.*, vol. 48, no. 5, pp. 476–491, Sep./Oct. 2013, doi: 10.3109/10409238.2013.831807.

[54] L. Zhang, X. Ou, G. Zhao, and X. Ding, "Highly efficient *in vitro* site-specific recombination system based on streptomyces phage phiBT1 integrase," *J. Bacteriol.*, vol. 190, no. 19, pp. 6392–6397, Oct. 2008, doi: 10.1128/JB.00777-08.

[55] L. Bibb and G. Hatfull, "Integration and excision of the Mycobacterium tuberculosis prophage-like element, phiRv1," *Mol. Microbiol.*, vol. 45, no. 6, pp. 1515–1526, Sep. 2002.

[56] M. Rashel, J. Uchiyama, T. Ujihara, I. Takemura, H. Hoshiba, and S. Matsuzaki, "A novel site-specific recombination system derived from bacteriophage phiMR11," *Biochem. Biophys. Res. Commun.*, vol. 368, no. 2, pp. 192–198, Apr. 4, 2008, doi: 10.1016/j.bbrc.2008.01.045.

[57] B. Swalla, E. Cho, R. Gumport, and J. Gardner, "The molecular basis of co-operative DNA binding between lambda integrase and excisionase," *Mol. Microbiol.*, vol. 50, no. 1, pp. 89–99, Oct. 2003.

[58] P. Ghosh, L. Wasil, and G. Hatfull, "Control of phage Bxb1 excision by a novel recombination directionality factor," *PLoS Biol.*, vol. 4, no. 6, p. e186, Jun. 2006.

[59] M. Wu, R. Su, X. Li, T. Ellis, Y. Li, and X. Wang, "Engineering of regulated stochastic cell fate determination," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 26, pp. 10 610–10 615, Jun. 25, 2013, doi: 10.1073/pnas.1305423110.