Mechanistic modelling of tyrosine recombination reveals key parameters determining the performance of a CAR T cell switching circuit

Jack E. Bowyer1,2, Deboki Chakravarti3,4, Wilson W. Wong3,4, Declan G. Bates1,2

1School of Engineering, University of Warwick, Coventry, CV4 7AL, UK
2Warwick Integrative Synthetic Biology Centre, Coventry, CV4 7AL, UK
3Department of Biomedical Engineering, Boston University, Boston, MA, 02215, USA
4Biological Design Center, Boston University, Boston, MA, 02215, USA

E-mail: d.bates@warwick.ac.uk

Abstract: Inducible genetic switches based on tyrosine recombinase-based DNA excision are a promising platform for the regulation and control of chimeric antigen receptor (CAR) T cell activity in cancer immunotherapy. These switches exploit the increased stability of DNA excision in tyrosine recombinases through an inversion–excision circuit design. Here, the authors develop the first mechanistic mathematical model of switching dynamics in tyrosine recombinases and validate it against experimental data through both global optimisation and statistical approximation approaches. Analysis of this model provides guidelines regarding which system parameters are best suited to experimental tuning in order to establish optimal switch performance in vivo. In particular, they find that the switching response can be made significantly faster by increasing the concentration of the inducer drug 4-OHT and/or by using promoters generating higher expression levels of the FlpO recombinase.

1 Introduction

Cell-based therapies that employ engineered T cells expressing chimeric antigen receptors (CARs) to target cancer cells have demonstrated promising responses in recent clinical trials, especially with regard to B cell cancers [1–4]. T cell immunotherapy involves extracting T cells from a patient's blood, engineering the cells to improve their ability to detect and target tumour cells, and reintroducing them back into the patient. Synthetic CARs are integrated into the cells which enable them to recognize specific tumour-associated antigens. There are, however, a number of significant safety and efficacy concerns that have arisen regarding T cell therapies. Success against solid tumours has been limited due to difficulties relating to autoreactive ‘on-target, off-tumour’ responses as well as cytokine release syndrome, a response to antigen stimulation that accelerates the immune response to potentially fatal levels. Strategies such as kill switches, designed to terminate the T cell response in cases of high toxicity, terminate the entire therapy which often leads to the continuation of tumour development. Reliable modulation of the therapy is therefore of great importance and has prompted research in the development of switching and control circuitry to enable the regulation of engineered T cell responses [5–17].

A promising approach for the creation of such circuitry is the use of site-specific recombinases (SSRs), which are capable of precise DNA manipulation in mammalian cells [18–20], and can be used to engineer synthetic biological circuits capable of performing user-defined functions that are programmed into the cellular DNA [21, 22]. Moreover, SSR-based gene switches elicit permanent alterations to the DNA and thus possess a built-in memory property. As such, SSR-based gene switches are capable of stable recombination events that produce stable genetic states [23, 24], however, the reversal of recombination events via serine integrases is typically dependent on a recombination directionality factor (RDF) [25–28], which can compromise their efficiency [29, 30]. Consequently, serine recombinase-based circuit designs attempt to exploit unidirectional recombination events that are, therefore, unable to facilitate transitioning back to prior system states [31].

Recent mathematical modelling studies regarding serine integrases and their RDFs have revealed that computational models are capable of quantitative replication of the observed in vitro dynamics [30, 32] and can aid the examination of the practical viability of novel synthetic circuits that combine multiple biological elements [30, 33]. Here we seek to develop the first equivalent model for tyrosine recombinase-based circuitry and explore its use in determining key experimental parameters that govern the performance of a CAR T cell switching circuit.

Tyrosine recombinases elicit a primary inversion event in an identical manner to that of serine integrases. The secondary inversion event, however, can also be mediated by the same tyrosine recombinase without any additional cofactors or RDFs, which causes bidirectionality (reversibility) (Fig. 1a) [34, 35]. Tyrosine recombinase binding interactions exhibit positive cooperativity [19, 36, 37] and conserve a series of single strand exchanges known as a Holliday junction [38, 39]. Deletion and insertion events mediated by tyrosine recombinases also mirror that of their serine cousins, with the exception of bidirectionality (Fig. 1b). Hence, adopting an elementary inversion switch design is not viable, since inversion events are susceptible to natural reversibility, which compromises the efficiency of the desired transcriptional response (Fig. 1c). Reversibility is theoretically more prominent in inversion events compared to deletion and insertion events. This is because the deletion of a genetic sequence forms a loop of DNA that dissociates from its original strand, and therefore presents additional spatial disparity to overcome in a subsequent insertion event, compared to successive inversions of the same genetic sequence. Consequently, DNA excision has emerged as a viable method for developing tyrosine recombinase-based circuitry [40, 41]. The inducible control over gene
expression afforded by SSRs has the potential to significantly improve the efficacy of engineered T cells. Recombinase-based genetic switches represent the most versatile switches in T cells, which have the potential to enable optimal T cell activity that can be tuned via drug dosage and duration, whilst retaining the memory of transient stimuli.

Mathematical models have been used to elucidate a wide variety of synthetic biological circuits. These include gene regulatory networks [42, 43], tuneable oscillators [44], and genetic counters [21]. Problems that arise as a side effect of assembling synthetic biological circuits can also be addressed using mathematical models such as the burden placed on cell growth due to the introduction of synthetic constructs [45]. At the very least, in silico simulation of a given circuit can test hypotheses relating to its dynamic performance in a fraction of the time taken to carry out the equivalent experimental study. The development of new models, therefore, has the potential to create a host of valuable tools for synthetic biologists looking to design and implement novel circuitry in vivo [46]. In this study, we present a mechanistic mathematical model of an inversion–excision switch based on our knowledge of the underlying molecular interactions associated with tyrosine recombinases. The model is formulated through the application of mass action kinetics, thus providing a deterministic output via numerical simulation. The model is validated through two separate data-fitting procedures. The first is a global optimisation technique that minimises error through a ‘survival of the fittest’ procedure. The second is a statistical technique that employs Bayes’ theorem to infer the parameterisation most likely to have produced the data. Our validated model is used to demonstrate an inverse proportionality between 4-OHT concentration and switch response time and identify which model parameters are most sensitive to perturbations and are hence best suited to experimental tuning in vivo.

2 Results and discussion
2.1 Mechanistic model of a tyrosine recombinase inversion-excision switch

The tyrosine recombinase genetic switch developed in [41] exploits the increased stability of DNA excision through a gated inversion-excision circuit design (Fig. 2). The input to this circuit is the flippase O (FlpO) recombinase, an optimised variant of flippase (Flp) with increased efficiency within mammalian cells [47]. A metabolite of the Food and Drug Administration-approved drug tamoxifen, known as 4-OHT, is used to activate the FlpO before it can enter the nucleus and mediate the necessary DNA recombination events. FlpO is expressed constitutively in the cytosol until the advent of 4-OHT causes activation and transport into the nucleus. Leaky nuclear localisation of FlpO can also occur in the absence of 4-OHT, resulting in basal switch activity.

The full network of molecular interactions underlying the switch is shown in Fig. 2, and consists of each DNA:protein and Holliday junction complex arising from FlpO binding interactions. The model is formulated through the application of mass action kinetics, thus providing a deterministic output via numerical simulation. The model is validated through two separate data-fitting procedures. The first is a global optimisation technique that minimises error through a ‘survival of the fittest’ procedure. The second is a statistical technique that employs Bayes’ theorem to infer the parameterisation most likely to have produced the data. Our validated model is used to demonstrate an inverse proportionality between 4-OHT concentration and switch response time and identify which model parameters are most sensitive to perturbations and are hence best suited to experimental tuning in vivo.

2 Results and discussion
2.1 Mechanistic model of a tyrosine recombinase inversion-excision switch

The tyrosine recombinase genetic switch developed in [41] exploits the increased stability of DNA excision through a gated inversion-excision circuit design (Fig. 2). The input to this circuit is the flippase O (FlpO) recombinase, an optimised variant of flippase (Flp) with increased efficiency within mammalian cells [47]. A metabolite of the Food and Drug Administration-approved drug tamoxifen, known as 4-OHT, is used to activate the FlpO before it can enter the nucleus and mediate the necessary DNA recombination events. FlpO is expressed constitutively in the cytosol until the advent of 4-OHT causes activation and transport into the nucleus. Leaky nuclear localisation of FlpO can also occur in the absence of 4-OHT, resulting in basal switch activity.

The full network of molecular interactions underlying the switch is shown in Fig. 2, and consists of each DNA:protein and Holliday junction complex arising from FlpO binding interactions. The initial system state (State 1) consists of a blank sequence flanked by antiparallel FRT sites and an inverted CAR gene flanked by parallel FRT sites. This setup does not provide any transcription. FlpO is expressed constitutively in the cytosol until the advent of 4-OHT causes activation and transport into the nucleus. Leaky nuclear localisation of FlpO can also occur in the absence of 4-OHT, resulting in basal switch activity.

The full network of molecular interactions underlying the switch is shown in Fig. 2, and consists of each DNA:protein and Holliday junction complex arising from FlpO binding interactions. The initial system state (State 1) consists of a blank sequence flanked by antiparallel FRT sites and an inverted CAR gene flanked by parallel FRT sites. This setup does not provide any CAR expression since the non-expressing blank sequence is positioned immediately downstream of the promoter and the CAR gene is not suitably oriented for transcription. FlpO expression elicits inversion of either FRT site pair to produce two transient genetic states. The first sees inversion of both genes, resulting in CAR gene expression due to the now suitably oriented CAR gene
localisation is denoted by $K_a$ and the rate of expression of activated FlpO that leaks into the nucleus is denoted by $\alpha_2$. The rate of FlpO activation and nuclear localisation is denoted by $K_a$. The rate of a FlpO monomer binding reversibly to free FRT sites is denoted by $k_{1,1}$. The rate of degradation of drug inducer is denoted by $\beta_d$. The rate of FlpO activation and nuclear localisation is denoted by $K_a$. The rate of a FlpO monomer binding reversibly to an occupied FRT site by $k_{1,1}$. The rate of degradation of the recombinase protein is denoted by $\beta_p$. The rate of a FlpO monomer binding reversibly to free FRT sites is denoted by $k_{1,1}$. The rate of degradation of drug inducer is denoted by $\beta_d$. The rate of FlpO activation and nuclear localisation is denoted by $K_a$. The rate of a FlpO monomer binding reversibly to free FRT sites is denoted by $k_{1,1}$.

See Fig. 2. Schematic diagram of a tyrosine recombinase-based inversion–excision switch. FlpO recombinase in its activated form targets orthogonal pairs of FRT attachment sites in transitioning the switch from State 1 to State 2 via two transient intermediate genetic states. FRT sites are depicted as purple and green triangles with different shades used to illustrate the result of recombination events. Clear and red pointed boxes depict the blank sequence and CAR gene, respectively. Black arrows depict DNA:protein binding reactions comprising unstable inversion and stable excision events. The rate of unactivated FlpO monomer binding, we denote the rate of a FlpO monomer binding reversibly to each corresponding FRT site is sufficient. Two FlpO monomers bound at each attachment site (D$^{1,1}_{F}$), and two monomers bound at one site (D$^{1,1}_{F}$). Details of the mathematical representation of cooperativity in our model can be found in Section 3. Our model consists of five intermediate Holliday junction complexes describing the inversion/excision of the DNA sequence. FlpO dissociates from the resultant composite attachment sites to give free, genetically differentiated DNA. The excision reaction is not symmetrical, due to the spatial disparity of its products. Again, two FlpO monomers bound at each corresponding FRT site is sufficient to facilitate the excision event, and Holliday junction formation is identical to that of the inversion reaction. FlpO dissociation produces free disparate genetic products: the excised sequence of DNA whose exposed ends are ligated to form a loop, and the shortened DNA strand that remains. Transition to prior genetic states is made possible by the reversibility of the molecular reactions, however, spatial localisation is required for maximum efficiency in the case of transitioning back to State 1 from State 2 (not relevant for the applications presented here).
Although this circuit resembles biological logic gate circuitry in the literature, it has just one input and its only desired function is to implement an efficient ON switch from State 1 to State 2. The ON switch is designed to allow the activation of the CAR T cells to be delayed until an appropriate time point, based on the requirements of individual patients. The OFF switch can be considered as a separate circuit in which the initial positions of the blank sequence and CAR gene are swapped, and thereby only State 1 provides CAR expression. The OFF switch is designed to provide controlled cessation of an initial ON state as an alternative to more traditional kill switches that terminate the therapy completely. Both the ON and OFF switches are induced in the same manner, through 4-OHT activation of the FlpO recombinase which then performs the necessary DNA recombination.

Based on a detailed analysis of the underlying reactions, we developed a set of biochemical equations representing the molecular interaction network shown in Fig. 2 (see supplementary material, Table S1). Our mechanistic model is derived through the application of mass action kinetics to these biochemical equations. The result is a system of ordinary differential equations (ODEs) that describes the rate of change in concentration of each molecular entity with respect to time (see supplementary material, Table S2).

Since we are interested in the dynamical response of the system over time and have time course data with which to validate the model, this modelling approach provides suitable approximations to the time-dependent fluctuations in the concentration of model variables in the ON state (the percentage of cells ON). The deterministic output provides an approximation to the average of an ensemble of stochastic simulations without the increased computational workload necessary for such probabilistic approaches.

To simulate the time course concentration dynamics of the genetic states of interest, we compute the total concentration of the system in these states by summing all ODEs describing molecular entities in the same state. Hence, we can derive the following two equations that capture the time course evolution of State 1 and State 2:

$$\frac{dS_1}{dt} = k_1(D_{5}^{H}F_1 + D_{5}^{H}F_2) - k^2_1(H_1^2 + H_2^2). \quad (1)$$

$$\frac{dS_2}{dt} = k_2(2D_2^{H}F_1 + D_2^{H}F_2) + k_2 \left(2D_2^{H}F_1 + D_2^{H}F_2 \right)$$

$$- k^2_2(H_1^2 + H_2^2). \quad (2)$$

where $S_i, iS_1, iS_2, tS_2$ and $S_0$ denote State 1, transient State 1, State 2, transient State 2 and the deleted DNA state respectively; $D, H$ and $F$ denote DNA, Holliday junction complexes and FlpO, respectively; $i, i.e.,$ and $e_i$ each denote each of the two inversion and excision reactions, respectively. Numerical subscripts for $H$ and $F$ denote the relevant Holliday junction complex (out of five) and the number of monomers, respectively. The numerical solutions to (1) and (2) constitute the outputs of our model simulations. The factor of half in the latter two terms in (2) is important for preserving the conservation of concentration (mass) in the model. A description of how this factor arises mathematically, using a simplified example, is provided in Supplementary Material, Section S1.

### 2.2 Model accurately predicts circuit switching efficiency and dynamics

Our experimental data on the performance of the inversion–excision switch consists of four time-course datasets that record the percentage CAR expression of both the ON switch and OFF switch over a ten-day period subject to 0 and 1 μM 4-OHT dosages (see Section 3).

We employed two separate methods to fit the model to our time course data, due to the lack of experimental data regarding the values of our model parameters. The first method employs global optimisation using a genetic algorithm (GA). The GA mimics natural selection by evolving an initial population of randomly generated solutions over a large number of generations until it converges to a near global minimum within the allocated parameter space (see Section 3). Such algorithms have been widely used in mathematical model inference research relating to, e.g. synthetic oscillators and gene regulatory networks [48–50].

The second data fitting method is a statistical approach known as approximate Bayesian computation (ABC). This is implemented using ABC-SysBio, a Python software package designed specifically for statistical parameter inference in biological systems research [51–53]. The programme employs sequential Monte Carlo (SMC) simulations to construct an accurate approximation to the posterior probability distribution defined by Bayes’ theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad P(B) > 0.$$  

Monte Carlo approaches involve computational simulations to generate random candidate solutions, testing their fitness against the desired output and repeating until a viable solution can be identified (see Section 3). ABC establishes distributions on each parameter value in order to infer the parameterisation most likely to have produced the data. The GA identifies a single model parameterisation that provides minimal error between the associated in silico simulation and experimental data. However, a single parameterisation provides no information on the extent of uncertainty in the optimal values and is thus more susceptible to overfitting the data. Consequently, we seek maximum reliability in our results by initially performing global optimisation, then verifying this result by examining the likelihood of the existence of better solutions using ABC.

Global optimisation data fitting results confirm that the mechanistic model is capable of replicating time course percentage CAR expression for both ON and OFF switches over the ten-day period recorded experimentally (Fig. 3, solid lines). The optimal parameter values identified by the GA are listed in supplementary material, Table S3. The error function used to assess the fitness of each solution calculates the absolute error across all four datasets. The optimal parameterisation provides an error of 28.15, which enables us to inform the selection of the sequence of error thresholds, $\epsilon_i$, for the subsequent ABC parameter inference. In this case, we set the penultimate error threshold, $\epsilon_9$ at 29 and the final threshold, $\epsilon_{10}$ at 28.15 under the assumption that solutions providing an error below 29 are sufficiently similar to our global optimal solution, and that convergence to the final threshold will present solutions superior to that of the GA. ABC parameter inference results reveal that a population of 100 parameterisations met the penultimate error threshold, but that convergence to the final threshold did not occur within a feasible time frame (Table 1). This provides strong evidence that the parameterisation identified by the GA is indeed near the global optimal solution since we are unable to statistically verify the existence of other solutions capable of providing a smaller error. Furthermore, our ABC approach provides distributions of each parameter value for accepted solutions at each error threshold. These distributions are non-uniform and hence it is unlikely that multiple distinct parameterisations are capable of providing the same optimal model output. This also suggests that our model is identifiable; at least within a finite region of the parameter space.

In order to clarify the key mechanistic features of the system, we conducted a sensitivity analysis to identify whether, and how, the model could be simplified without compromising performance (supplementary material, Fig. S1). Parameters are varied by up to two orders of magnitude above and below their optimal values. Sensitivity is examined with respect to response time, which is measured as the time taken for the ON switch to reach 90% of its maximum percentage CAR expression, and the time taken for the OFF switch to reach 90% of its total decrease in percentage CAR expression.

All parameters describing FlpO expression/degradation mechanisms within the system ($h_{a}, a_{d}, h_{b}, b_{d}, k_{a}, K_{a}$) showed high sensitivity, and are therefore retained in the simplified model. The
parameters describing Holliday junction formation ($k_1, k_{-1}, k_2, k_{-2}, k_5, k_{-5}, k_6, k_{-6}, k_7, k_{-7}$) showed relatively low sensitivity and are removed. The parameters describing cooperative FlpO monomer binding showed a high sensitivity for one pair of attachment sites ($i_2, e_1$: $k_1, k_{-1}, k_2, k_{-2}$), but comparatively low sensitivity for the other pair ($i_1, e_2$: $k_1, k_{-1}, k_2, k_{-2}$). Since each of the two pathways from State 1 to State 2 features recombination events involving each pair of attachment sites, we remove cooperative monomer binding and describe each recombination event as a single reversible reaction. The reduced model retains the four state variables relating to FlpO expression, activation, and degradation in the mechanistic model, but consists of just five DNA state variables compared to 63, while the number of parameters is reduced from 27 to 12 (Fig. 4).

Applying mass action kinetics to the biochemical reactions depicted in Fig. 4 gives the following system of nine ODEs:

$$\frac{d[F]}{dt} = a_6 - b_6[F] - K_{6}[\text{4-OHT}][F]$$

$$\frac{d[D_{S6}]}{dt} = k_1[F][D_{S6}] - k_1[F][D_{S5}]+ k_1[F][D_{S5}]$$

$$\frac{d[D_{S5}]}{dt} = k_2[F][D_{S5}] - k_2[F][D_{S4}] + k_1[F][D_{S5}] - k_1[F][D_{S4}]$$

$$\frac{d[D_{S4}]}{dt} = k_3[F][D_{S4}] - k_3[F][D_{S3}] + k_2[F][D_{S4}] - k_2[F][D_{S3}]$$

$$\frac{d[D_{S3}]}{dt} = k_4[F][D_{S3}] - k_4[F][D_{S2}] + k_3[F][D_{S3}] - k_3[F][D_{S2}]$$

$$\frac{d[D_{S2}]}{dt} = k_5[F][D_{S2}] - k_5[F][D_{S1}] + k_4[F][D_{S2}] - k_4[F][D_{S1}]$$

$$\frac{d[D_{S1}]}{dt} = k_6[F][D_{S1}] - k_6[F][D_{S0}] + k_5[F][D_{S1}] - k_5[F][D_{S0}]$$

$$\frac{d[D_{S0}]}{dt} = k_7[F][D_{S0}] - k_7[F][D_{-1}] + k_6[F][D_{S0}] - k_6[F][D_{-1}]$$

Each population of accepted particles is formed by assessing the error provided by each particle. If the error is less than the threshold for that population the particle is accepted if not, the particle is rejected. Our inference required that 100 particles be accepted before progression to the next population and error threshold. The final error threshold, $\epsilon_{10}$, was chosen based on our GA optimisation result and ABC-SysBio was unable to identify a population of particles that provide less error than this result.
model parameterisation (Fig. 3, dashed lines). The optimal parameterisation for the reduced model provides an error of 28.39, a negligible increase on the 28.15 provided by the mechanistic model. The optimal parameter values for the reduced model are listed in supplementary material, Table S4. Further attempts to simplify the model by removing other features resulted in large increases in the fitting error, providing strong evidence that this model represents the simplest possible mechanistic representation of the system. All results described hereafter were derived from this simplified model of the system.

2.3 Switching response times are inversely proportional to the concentration of inducer drug

Response time is the key performance characteristic of the switch since ultimately it will need to function reliably in mammalian cells within time frames that are appropriate to the condition of individual patients. By identifying parameters that are both influential over system output and experimentally tuneable, we can establish guidelines to optimise the functioning of the switch in different application scenarios. Model simulations reveal that there is an inverse proportionality between drug concentration and switching response time (Fig. 5). The response time of both switches is significantly decreased when induced by 4-OHT concentrations in the range $10^{-1}$–$10^{1}$ μM. Therefore the 1 μM 4-
OHT concentration used experimentally may require an increase of up to 10 μM to deliver improved response times. The concentration of 4-OHT used to induce the switch is important as it has implications regarding the cost of the treatment. Identifying key operational factors through model simulations ultimately informs the direction of confirmatory experimental work and presents new avenues of investigation in a time efficient manner. Given that the action of the CAR T cells is as effective as to provide significant therapeutic benefit over small time periods, these dynamics are unlikely to be advantageous. By contrast, the performance of the OFF switch improves uniformly as FlpO expression increases. Recent work on engineering Jurkat T cells expressing CAR within a relatively small time frame (∼55% after 2 days). Therefore, unless the action of the CAR T cells is as effective as to provide significant therapeutic benefit over small time periods, these dynamics are unlikely to be advantageous. By contrast, the performance of the OFF switch improves uniformly as FlpO expression increases.

2.4 Increasing Flp0 expression can provide faster switching

Sensitivity analysis of the reduced model parameters (supplementary material, Fig. S2) identifies the most sensitive to be $\alpha_a$, $\beta_p$, $\beta_d$, and $K_a$. Of these, the rates of protein and drug degradation, $\beta_p$ and $\beta_d$, and the rate of FlpO nuclear localisation through 4-OHT activation, $K_a$, are challenging candidates for experimental tuning. However, the rate of FlpO expression in the system, $\alpha_a$, and $\alpha_b$ is potentially tuneable through appropriate choice of the cognate promoter strength. Model simulations reveal faster response times for both the ON and OFF switches as the values of $\alpha_a$ and $\alpha_b$ are increased two- and five-fold (Fig. 6), with a significant reduction in switch response time for even a two-fold increase in FlpO expression. Recent work on engineering Jurkat T cells with inducible promoters of varying strengths illustrates the potential to provide the required increase in FlpO expression. However, the model predicts that increased FlpO expression also results in a decreased percentage of CAR expression over time with respect to the ON switch (Figs. 6a and 6b). For example, a five-fold increase in FlpO expression is predicted to provide >75% cells expressing CAR within a relatively small time frame (∼1 day), but this reduces to ∼55% after 2 days. Therefore, unless the action of the CAR T cells is as effective as to provide significant therapeutic benefit over small time periods, these dynamics are unlikely to be advantageous. By contrast, the performance of the OFF switch improves uniformly as FlpO expression increases.

The model also predicts that increased FlpO expression causes increased basal switching responses from both switches (Figs. 6c and 6d) that could potentially compromise the efficacy of the therapy altogether, regardless of which switch was considered to be most viable. Hence, attempts to improve the efficiency of the switch through increased FlpO expression would not be suitable for applications in which tight regulation of the initial state in the absence of induction is required.

2.5 Optimal switch performance presents a trade-off between 4-OHT concentration and FlpO expression

Investigating the effects of simultaneously modifying the expression of FlpO and the concentration of 4-OHT reveals that improved switch response times are possible for relatively large doses of 4-OHT if FlpO expression is unaltered, and for relatively low doses of 4-OHT if FlpO expression is increased ∼10-fold (Fig. 7). The model predicts that increasing FlpO expression ∼2-fold and using a 4-OHT dosage of ∼3 μM can provide a notable improvement in response times for both the ON and OFF switches. Although this demonstrates an advantageous reduction in the magnitude of the required fold change in FlpO expression for a relatively small increase in drug dosage, the resulting basal activity increases to potentially problematic levels, suggesting that tuning inducer concentration/dosage may be a more viable method for improving switch efficiency.

3 Materials and methods

3.1 Experimental procedure

Primary CD4+ T cells were harvested from blood obtained from the Boston Children's Hospital using the STEMCELL CD4+ enriched cocktail and RosetteSep system. T cells were maintained in X-VIVO 15 media (Lonza) supplemented with 5% AB Serum (Valley Biomedical), 10 mM N-acetyl L-cystein (Sigma), 55 μM 2-mercaptoethanol (Gibco), and 50–100 units/ml recombinant interleukin-2 (IL-2) (Tecin, NCI BRB Preclinical Repository). Lentivirus containing either the inducible FlpO or stable inversion switch was produced via polyethylenimine transfection of the human embryonic kidney 293 FT cells and collected 3 and 4 days post-transfection. The concentrated virus was produced via ultracentrifugation with 20% sucrose (Sigma) for 2 h at 4°C and 22,000 g. T cells were activated with CD3/CD28 Dynabeads (Gibco), and then spinfected with a virus one day later. Using half
of the concentrated virus, both inducible recombinase and stable inversion switch viruses were spun onto the well a six-well retronectin (Clontech)-coated plate for 90 min at 1200 g. T2a ribosomal skip sequence.

Cells were induced with 1 μM 4-hydroxytamoxifen (4-OHT, Sigma) in methanol solution except for dose response experiments. Cells were induced at a starting concentration of 200,000 cells/ml and maintained between 200,000–1200,000 cells/ml with media containing 4-OHT. Uninduced cells were also plated and maintained at the same cell concentrations in inducer-free media. CAR expression was observed via phycoerythrin (PE)-conjugated antibody staining (R&D Systems IC3696P) for a human myc epitope tag expressed in the extracellular portion of the CAR, which was then quantified via flow cytometry (Attune NxT Flow Cytometer).

The percentage switching efficiency is thought to reflect the efficiency of integrating both the FlpO-producing component and the switch itself into the T cells. Given that successful integration is not guaranteed for every cell, there are four possible outcomes: R−/FS− (no components), R+/FS− (recombinase only), R−/FS+ (switch only) and R+/FS+ (recombinase and switch). Only cells from dark blue (0 days) to dark red (10 days) bound by another monomer is greater than that of a monomer to a free site. This involves setting $k_1 = k_2$ and $k_{-1} = k_{-2}$ and introducing the multipliers $0 < m_1 < 1$ and $0 < m_2 < 1$ such that $m_1$ multiplies $k_1$ and $m_2$ multiplies $k_{-2}$, hence ensuring that $k_1/k_{-1} < k_2/k_{-2}$ in our parameter inference.

### 3.2 Model parameter inference and simulation

#### 3.2.1 Global optimisation:
We employed the built-in GA function in MATLAB for parameter inference purposes. The GA proceeds by initially generating a population of solutions at random within the predefined parameter space. The initial solutions are then scored based on their fitness. This is typically calculated by virtue of an error function that determines how well each solution is able to match the relevant experimental data. The best solutions are selected as parents that will produce the best offspring to populate the next generation of solutions. Parents produce offspring through crossover, whereby a random place in their binary genotype is selected and the information beyond that point is swapped over.

The mutation is also incorporated, whereby single point alterations in the offspring's genotype are imposed to increase diversity within the population. This procedure is repeated indefinitely until a predefined termination criterion is reached; typically, the scores are unchanged over a specified number of generations, the algorithm reaches a specified number of generations, or a specified solution score is established. Parallelisation of the MATLAB code enabled us to run the GA with a large search population for many generations; this significantly increases the likelihood of establishing the global optimal solution. Given the lack of available data regarding the relevant reaction rate constants in the literature, we require a parameter space large enough to locate optimal solutions, but not so big that convergence timescales become impractical. A large parameter space increases the time taken to produce the next generation of solutions however, this can also decrease the overall number of generations required for convergence, since more solutions are inspected in each case. Large parameter spaces also increase the risk of incurring problems with stiff model simulations that may cause the GA to fail, hence multiple trials are often required to determine effective performance criteria. Hence, the search interval imposed on all model parameters is [0, 1]. We ran the GA over 1000 generations with a population size of at least 1000 in order to maximise the likelihood of convergence and the identification of the global performance criteria.
optimum solution. We define the error function that calculates the mean absolute error between the four time-course datasets and their corresponding model simulations such that

$$d = \frac{1}{10} \sum_{i=1}^{10} |x_i - y_i|,$$

where $x_i$ and $y_i$ are corresponding elements of the experimental data and model simulation vectors, respectively. Each of the four time-course datasets consists of ten values and our distance function calculates the mean absolute error across all four datasets.

### 3.2.2 Approximate Bayesian computation (ABC): The ABC–SMC procedure implemented by ABC-SysBio proceeds in the following manner: an initial population of solutions (particles) is randomly generated in accordance with the prior distributions imposed on the model parameters. Each particle provides a simulated dataset, $D^*$, which is compared to the fixed experimental dataset, $D$, by an appropriate distance (error) function and its fitness is scored accordingly. This error score determines the acceptance of a particle, dependent on a decreasing sequence of error thresholds, $\epsilon$, set to correspond with each population, i.e.

$$d(D^*, D) < \epsilon_i,$$

where $\epsilon_1 > \epsilon_2 > \ldots > \epsilon_k$ and $d$ is the distance function. We define our distance function as the mean absolute error between model simulations and experimental data in the same way as the GA approach. Subsequent populations are obtained by perturbing particles from the previous population in accordance with a predetermined perturbation kernel, proceeding until the model is unable to produce particles of sufficient fitness to satisfy the immediate threshold.

An array of model-specific criteria are required to allow the ABC–SysBio package to run efficiently: the sequence of decreasing error thresholds, $\epsilon$, must be provided whereby only the particles capable of providing error less than that of the threshold will be accepted by the algorithm. Each $\epsilon$ must be satisfied in succession until the particles are unable to satisfy the next threshold. The satisfaction of an individual threshold is dependent on the number of particles accepted; the number of acceptable particles required before passage to the next threshold must also be predetermined. The larger the number of particles, the higher probability of reliable inference results and the longer the time taken by the algorithm to converge. Each individual parameter subject to inference requires a prior probability distribution in order to establish the parameter space within which to locate acceptable particles. We allocate the same parameter space used in GA parameter inference to facilitate direct comparisons of the results. Hence, the prior distributions imposed on all model parameters are uniform distributions on the interval $[0, 1]$. The convergence of the algorithm is dependent on all of the aforementioned factors and hence it may require several trials to establish the appropriate performance criteria. To achieve credible results, it is advised that parameter inference and model selection tasks are repeated multiple times due to the random nature of the Monte Carlo simulations that drive the algorithm.

### 4 Conclusion

Mathematical models of synthetic circuits that are based strictly on underlying biological mechanisms can be used as design tools to reduce development times, lower the risk of circuit failure, and identify key parameters determining circuit performance. We have developed the first mechanistic mathematical model of a tyrosine recombinase-based inversion–excision switch and validated it against dynamic data on CAR expression using both global optimisation and statistical approaches. A sensitivity analysis of this model revealed key mechanistic features that allowed the construction of a significantly simplified model that yielded equivalent data fitting results. Analysis of this model revealed that the response time of the switch can be modulated experimentally by tuning the concentration of the inducer drug and/or by choosing promoters that increase the level of expression of the FlpO recombinase. However, both of these solutions also have potential drawbacks in terms of increasing basal expression activity, indicating that the successful clinical deployment of this and other novel recombinase-based circuits might be dependent on the development of new mechanisms for minimising or even nullifying basal recombinase expression.

### 5 Acknowledgments

Research supported by funding from a Bilateral BBSSC/NSF award (#1614642 Bio-ReWritable biocomputers in mammalian cells) and BBSRC grant BB/P011926/1.

### 6 References

[1] Davila, M.L., Riviere, I., Wang, X., et al. : "Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia," *Sci. Transl. Med.*, 2014, 6, p. 224ra225

[2] Brentjens, R.J., Riviere, I., Park, J.H., et al. : "Safety and persistence of adoptively transferred autologous CD19-targeted T cells in patients with relapsed or chemotherapy refractory B-cell leukemias," *Blood*, 2011, 118, pp. 4817–4828

[3] Maudree, S.L., Frey, N., Shaw, P.A., et al. : "Chimeric antigen receptor T cells for sustained remissions in leukemia," *N. Engl. J. Med.*, 2014, 371, pp. 1507–1517

[4] Lee, D.W., Kochenderfer, J.N., Stetler-Stevenson, M., et al. : "T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukemia in children and young adults: a phase 1 dose-escalation trial," *Lancet*, 2015, 385, pp. 517–528

[5] Wu, C.Y., Roybal, K.T., Puchner, E.M., et al. : "Remote control of therapeutic T cells through a small molecule-gated chimeric receptor," *Science*, 2015, 350, p. aab4077

[6] Fedorov, V., Sadelain, M., Kloss, C.C. : "Novel approaches to enhance the specificity and safety of T cell adoptive immunotherapy for lymphoma," *Nat. Biotechnol.*, 2013, 31, pp. 71–75

[7] Sakemura, R., Terakura, S., Watanabe, K., et al. : "A Tet-On inducible system for controlling CD19-chimeric antigen receptor expression upon drug administration," *Cancer Immunol. Res.*, 2016, 4, pp. 658–668

[8] Di Stasi, A., Tey, S.K., Dotti, G., et al. : "Inducible apoptosis as a safety switch for adoptive cell therapy," *N. Engl. J. Med.*, 2011, 365, pp. 1673–1683

[9] Straathof, K.C., Pulè, M.A., Yotnda, P., et al. : "An inducible caspase 9 safety switch for T-cell therapy," *Blood*, 2005, 105, pp. 4247–4254

[10] Budde, L.E., Berger, C., Lin, Y., et al. : "Combining a CD20 chimeric antigen receptor and an inducible caspase 9 suicide switch to improve the efficacy and safety of T cell adoptive immunotherapy for lymphoma," *PLoS One*, 2013, 8, p. e82742

[11] Wei, P., Wong, W., Park, J., et al. : "Bacterial virulence proteins as tools to rewire kinase pathways in yeast and immune cells," *Nature*, 2012, 488, pp. 384–388

[12] Ma, J.S., Kim, J.Y., Kaneko, S.A., et al. : "Versatile strategy for controlling the specificity and activity of engineered T cells," *Proc. Natl. Acad. Sci. U.S.A.*, 2016, 113, pp. E459–E469

[13] Rodgers, D.T., Mazagopoula, M., Hampton, E.N., et al. : "Switch-mediated activation and retargeting of CAR-T cells for B-cell malignancies," *Proc. Natl. Acad. Sci. U.S.A.*, 2016, 113, pp. E459–E468

[14] Kodu, K., Imai, C., Lorenzana, L., et al. : "T lymphocytes expressing a CD16 signaling receptor exhibit antibody-dependent cancer cell killing," *Cancer Res.,* 2014, 74, pp. 93–103

[15] Chen, Y., Jensen, M.C., Smolke, C.D. : "Genetic control of mammalian T-cell activation and retargeting of CAR-T cells for B-cell malignancies," *Science*, 2015, 348, pp. 831–836

[16] Tortikai, H., Reik, A., Lin, P.G., et al. : "A foundation for universal T-cell based immunotherapy: T cells engineered to express a CD19-specific chimeric antigen-receptor and eliminate expression of endogenous TCR," *Blood*, 2012, 119, pp. 5697–5705

[17] Gresh, A.C., Olivares, E.C., Thyagarajan, B., et al. : "A phage integrase directs efficient site-specific integration in human cells," *Proc. Natl. Acad. Sci. U.S.A.*, 2000, 97, pp. 5995–6000

[18] Ghosh, K., Van Duyne, G.D. : "Cre-loxP biochemistry," *Methods*, 2002, 28, pp. 374–383

[19] Ghosh, K., Lau, C.K., Guo, F., et al. : "Peptide trapping of the Holliday junction intermediate in Cre-loxP-specific recombination," *J. Biol. Chem.*, 2005, 280, pp. 8280–8289

[20] Friesland, A.E., Lu, T.K., Wang, X., et al. : "Synthetic gene networks that count," *Science*, 2009, 324, pp. 1199–1202

[21] Fernandez-Rodriguez, C.A., Grimm, R., et al. : "Memory and combinatorial logic based on DNA inversions: dynamics and evolutionary potential," *ACS Synth Biol.*, 2014, 3, pp. 1361–1372

[22] Gresh, A., Colkes, M. : "Phage integrases: biology and applications," *J. Mol. Biol.*, 2004, 335, pp. 667–678

[23] Fogg, P., Colloms, S., Rosser, S., et al. : "New applications for phage integrases," *J. Mol. Biol.*, 2014, 426, pp. 2703–2716
[25] Ghosh, P., Panunzio, N., Hatfull, G.: ‘Synthesis in phage Bxb1 integration: selection mechanism for the correct pair of recombination sites’, J. Mol. Biol., 2005, 349, pp. 331–348

[26] Ghosh, P., Bibb, L., Hatfull, G.: ‘Two-step site selection for serine-integrate-mediated excision: DNA-directed integrase conformation and central dinucleotide proofreading’, Proc. Natl. Acad. Sci. U.S.A., 2008, 105, pp. 3238–3243

[27] Khaleel, T., Younger, E., McIwan, A., et al.: ‘A phage protein that binds ϕC31 integrase to switch its directionality’, Mol. Microbiol., 2011, 80, pp. 1450–1463

[28] Olorunniyi, F., Buck, D., Colloms, S., et al.: ‘Gated rotation mechanism of site-specific recombination by ϕC31 integrase’, Proc. Natl. Acad. Sci. U.S.A., 2012, 109, pp. 19661–19666

[29] Bonnet, J., Subsoontorn, P., Endy, D.: ‘Rewritable digital data storage in live cells via engineered control of recombination directionality’, Proc. Natl. Acad. Sci. U.S.A., 2012, 109, pp. 8884–8889

[30] Bowyer, J., Zhao, J., Subsoontorn, P., et al.: ‘Mechanistic modeling of a rewritable recombinase addressable data module’, IEEE Trans. Biomed. Circuits Syst., 2016, 10, pp. 1161–1170

[31] Hsiao, V., Hori, Y., Rothemund, P., et al.: ‘A population-based temporal logic gate for timing and recording chemical events’, Mol. Syst. Biol., 2016, 12, p. 869

[32] Pokhilko, A., Zhao, J., Stark, W.M., et al.: ‘A simplified mathematical model of directional DNA site-specific recombination by serine integrases’, J. R. Soc. Interface, 2017, 14, p. 20160618, doi: 10.1098/rsif.2016.0618

[33] Pokhilko, A., Ebenhoh, O., Stark, W.M., et al.: ‘Mathematical model of a serine integrase-controlled toggle switch with a single input’, J. R. Soc. Interface, 2018, 15, p. 20180160, doi: 10.1098/rsif.2018.0160

[34] Sorrell, D.A., Kolb, A.F.: ‘Targeted modification of mammalian genomes’, Biotechniques, 2005, 39, pp. 431–469

[35] Grindley, N.D., Whiteson, K.L., Rice, P.A.: ‘Mechanisms of site-specific recombinatation’, Annu. Rev. Biochem., 2006, 75, pp. 567–605

[36] Guo, F., Gopaul, D.N., Van Duyne, G.D.: ‘Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse’, Proc. Natl. Acad. Sci. U.S.A., 1999, 96, pp. 7143–7148

[37] Van Duyne, G.D.: ‘A structural view of Cre-loxP site-specific recombination’, Annu. Rev. Biophys. Biomol. Struct., 2001, 30, pp. 87–104

[38] Chen, Y., Rice, P.A.: ‘New insight into site-specific recombination from Flp recombinase-DNA structures’, Annu. Rev. Biophys. Biomol. Struct., 2003, 47, pp. 122–131

[39] Rashel, M., Uchiyama, J., Ujihara, T., et al.: ‘A novel site-specific recombination system derived from bacteriophage qMR11’, Biochem. Biophys. Res. Commun., 2008, 368, pp. 192–198

[40] Weinberg, B.H., Hang Pham, N.T., Caraballo, L.D., et al.: ‘Large scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells’, Nat. Biotechnol., 2017, 35, pp. 453–462

[41] Chakravarti, D., Caraballo, L.D., Weinberg, B., et al.: ‘Inducible gene switches with memory in human T cells for cellular immunotherapy’, ACS Synth. Biol., 2019, 8, pp. 1744–1754

[42] Hancock, E.J., Stan, G.-B., Arpino, J.A.J., et al.: ‘Mechanistic modeling: bridging the gap between concept and realisation in synthetic biology’, J. R. Soc. Interface, 2015, 12, doi: 10.1098/rsif.2015.0312

[43] Herath, N., Del Vecchio, D.: ‘Deterministic-like model reduction for a class of multiscale stochastic differential equations with application to biomolecular systems’, IEEE Trans. Autom. Control, 2018, 64, pp. 351–358

[44] Nystrom, A., Papachristodoulou, A., Angel, A.: ‘A dynamic model of resource allocation in response to the presence of a synthetic construct’, ACS Synth. Biol., 2018, 7, pp. 1201–1210

[45] Kranz, A., Fu, J., Durrbach, K., et al.: ‘An improved Flp deleter mouse in C57Bl/6 based on FlpO recombinase’, Genesis, 2010, 48, pp. 512–520

[46] Chen, B.-S., Chen, P.-W.: ‘GA-based design algorithms for the robust synthetic genetic oscillators with prescribed amplitude, period and phase’, Gene. Regul. Syst. Bio., 2010, 4, pp. 35–52

[47] Sirbu, A., Ruskin, H.J., Crane, M.: ‘Comparison of evolutionary algorithms in gene regulatory network model inference’, BMC Bioinformatics, 2010, 11, p. 59

[48] Smith, R.W., van Sluijs, B., Fleck, C.: ‘Designing synthetic oscillatory networks in silico: a generalised evolutionary algorithm approach’, BMC Syst. Biol., 2017, 11, p. 118

[49] Liepe, J., Kirk, P., Filippi, S., et al.: ‘A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation’, Proc. Natl. Acad. Sci. U.S.A., 2012, 109, pp. 1201–1210

[50] Stumpf, M.: ‘Approximate Bayesian inference for complex ecosystems’, F1000Prime Rep., 2014, 6, p. 60

[51] Smith, R., Grohn, Y.: ‘Use of approximate Bayesian computation to assess and fit models of Mycobacterium leprae to predict outcomes of the Brazilian control program’, PLOS One, 2015, 10, pp. e0129535