Host-aware modelling of a synthetic genetic oscillator*

Alexander Darlington and Declan G. Bates

***Abstract*— Synthetic genetic circuits sometimes exhibit unexpected functionality or even fail entirely when implemented *in vivo*, due to the effects of interactions with the host cell that were not accounted for in the circuit's design. In this paper, we consider the effects that limitations in cellular resources have on the dynamics of a synthetic cellular oscillator. We show that incorporating these effects into a host-aware model of the synthetic oscillator results in significant changes in its dynamics, highlighting the need to take account of host-circuit interactions in mathematical models that are to be used as CAD tools for synthetic circuitry.**

I. Introduction

A key goal of synthetic biology is to achieve the capability to efficiently design and implement novel gene regulatory networks, such as oscillators, in living cells. However, at present many gene circuits require several iterative rounds of testing and redesign after the initial design stage, a process that can significantly increase the cost and time-frames associated with the development of new synthetic systems . Initial circuit designs are often found to fail once implemented in the host cell due to hidden or unforeseen host-circuit interactions and other context-dependencies [1], [2], [3], [4]. A key reason for this is that synthetic circuits are often designed based on simplified models which neglect many components of the host cell which interact with the circuit or are required for it to function. Increasingly, modelling frameworks are being developed which enable host-circuit interactions to be explicitly taken into account during the circuit design phase [5], [6], thus leading to models that are more predictive and hence more useful as design tools.

Here we investigate modelling of the effects of resource limitations in the context of the design of a synthetic genetic oscillator. Cells have finite capacity for protein synthesis due to limitations in the number of macromolecular complexes that are required for gene expression. This results in trade-offs whereby the expression of one gene results in the decreased expression of another [5], [7]. This can result in interactions between genes even in the absence of clear regulatory linkages [8]. Here, we investigate these issues in the context of the design of an oscillator circuit implemented *in vivo* by Stricker et al [9].

This gene circuit consists of an activator and repressor, both under mutual control of each other. The activator (A) triggers its own expression and that of the repressor (I), while the repressor inhibits the expression of both A and I, see Fig. 1.

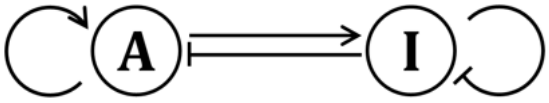

Fig. 1. The activator (A) increases its own expression and that of the repressilator (I). The repressor binds and inhibits its own expression and that of the activator

The paper is organised as follows: Section II describes the model of the isolated oscillator circuit, and the construction of a host-aware version of this model that takes into account resource limitations in the host cell. In Section III we perform a detailed analysis of the dynamics of both models across realistic ranges of their parameter values, in order to investigate the impact of host-circuit interactions on the circuit dynamics and hence the circuit design process. Finally, in Section IV we discuss the implications of our results for the design of synthetic oscillators and offer some conclusions.

II. Model construction

A. Modelling regulation by transcription factors

We model promoter regulation in the oscillator circuit using Hill functions. Given that both transcription factors affect the same promoter, the regulatory effect is the product of the two transcription factors, and thus we have that:

$$R(p_A, p_I) = \frac{\alpha}{1 + \alpha + \beta + \alpha\beta} \tag{1}$$

where α is the positive effect of the activator $(p_A/k_A)^{h_A}$ and β is the inhibitory effect of the inhibitor $(p_I/k_I)^{h_I}$. The variables p_A and p_I are the number of activator and repressor molecules respectively. The parameters k_A and k_I are the number of molecules for half the binding sites to be occupied for the activator and repressor respectively while h_A and h_I are the respective Hill coefficients. The transcription rate of each gene is the product of the maximal rate and the regulation effect.

B. An isolated circuit model ignoring resource limitations

In this model, mRNAs are created spontaneously and then converted into proteins by a process modelled by a first order rate equation, with constant k_{eff} (equivalent to the proxy chemical reaction mRNA $\rightarrow$ protein). This model therefore produces protein directly and so does not take into account utilization of the host ribosomes. All species are degraded and we account for protein dilution at cell division by including an estimated growth rate. Applying the law

*Funding from the University of Warwick and the EPSRC & BBRSRC Centre for Doctoral Training in Synthetic Biology (grant EP/L016494/1). The authors are with the Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry, UK. (e-mail: a.p.s.darlington@warwick.ac.uk)

of mass action to these reactions allows us to develop the following differential equations for the dynamics of mRNA (m_G) and protein (p_G) for each gene, G:

$$\dot{m}_G = w_0 + w_G R(p_A, p_I) - \delta_{m_G} m_G - \lambda_{eff} m_G \quad (2)$$

$$\dot{p}_G = k_{eff} m_G - \delta_{p_G} p_G - \lambda_{eff} p_G \quad (3)$$

where $G = A$ for the activator and $G = R$ for the repressor. In this model, growth rate and translation rate are global parameters linked to the internal status of the host cell - to represent the behaviour of the host we need to estimate fixed values for these parameters. We assume a cell doubling time of approximately 30 minutes, giving an effective growth rate of $log(2)/30$ i.e. 0.0231 per minute. We estimate an effective global protein production rate (k_eff) at 4 molecules per minute, based on the assumption of a protein of length 300 amino acids and an elongation rate of 20 amino acids per second (values retrieved from the Bionumbers database [8]).

C. A host-aware circuit model

Our starting point for the inclusion of more complex host-circuit interactions in the oscillator circuit is the microbial trade-off model developed by Weiße et al [6]. This model consists of 14 differential equations tracking the time evolution of a simple metabolism, including a universal energy carrier (E), and simple proteome, consisting of enzymes, ribosomes and host factors. The proteome mass is restricted while the ribosomes and energy are generated within the model and so are also limited. This creates the simple trade-off that the expression of one protein necessitates the decrease in expression of another as gene expression utilizes free ribosomes and energy and the final protein contributes to the cell's mass.

In this model mRNAs are born spontaneously at a rate proportional to the energy status of the cell ($w_G T_X(E)$, see Table I) and the regulatory effect of the repressors. We assume that the leakiness of the promoter is energy independent and of small value (< 1 molecule per minute). Free ribosomes (p_R) bind mRNAs reversibly to form translation complexes (c_G). Proteins are produced from translation complexes at a rate proportional to both the protein length and the cell's energy status ($T_L(c_G, E)$, see Table I). Protein production also liberates mRNA and free ribosomes. Reactions are modelled according to the law of mass action. All components are diluted at the cell's growth rate and all species, except translation complexes, are degraded. See Table I for the detailed reaction scheme and rate equations.

To model the oscillator circuit in this framework, we introduce the following equations to the host model to describe the activator and repressor:

$$\dot{m}_G = w_0 + w_G R(p_A, p_I) T_X(E) - b_G p_R m_G + u_G c_G + T_L(c_G, E) - (\delta_{m_G} + \lambda) m_G \quad (4)$$

$$\dot{c}_G = b_G p_R m_G - u_G c_G - T_L(c_G, E) - \lambda c_G \quad (5)$$

$$\dot{p}_G = T_L(c_G, E) - (\delta_{p_G} + \lambda) p_G \quad (6)$$

TABLE I

GENE EXPRESSION REACTIONS AND RATE EQUATIONS

Gene expression reactions	Isolated	Host-aware
$\emptyset \xrightarrow{w_0} m_G$ [a]	w_0	w_0
$\emptyset \xrightarrow{w_G R} m_G$ [b]	$w_G R(p_A, p_I)$	$w_G R(p_A, p_I) T_X(E)$ [c]
$m_G + p_R \xrightarrow{b_G} c_G$	n/a	$b_G p_R m_G$
$c_G \xrightarrow{u_G} p_R + m_G$	n/a	$u_G c_G$
$c_G \xrightarrow{k} p_R + m_G + p_G$	n/a	$k c_G = \frac{\gamma(E)}{n_G} c_G$ [c]
$m_G \xrightarrow{k} p_G$	$k_{eff} m_G$	n/a
$m_G \xrightarrow{\delta_{m_G}} \emptyset$	δ_{m_G}	δ_{m_G}
$p_G \xrightarrow{\delta_{p_G}} \emptyset$	δ_{p_G}	δ_{p_G}
$X \xrightarrow{g} \emptyset$ [d]	$\lambda_{eff} X$	λX

[a] Transcription due to promoter leakage. [b] Transcription due to regulation by transcription factors. [c] $T_X(E)$ and $\gamma(E)$ are functions developed by Weiße et al. Both functions scale reactions according to the cell's internal energy status. $T_X(E)$ scales the rate at which mRNAs are spontaneously born and $\gamma(E)$ scales the elongation rate. Rates are scaled by $E/(\nu + E)$ where ν is the threshold value. As $E \to 0$ both expressions tend to 0 so that translation and transcription cease at zero energy and as $E \to \infty$ both tend to 1, so at maximal cellular energy transcription and translation are maximal. See Weiße et al [6] for full derivation. [d] The dilution reaction of the generic species X. For the host-aware model this is λ and is determined by the model. For the isolated and ribosome models the growth rate is estimated as λ_{eff}.

We also modify the host equations as needed to model the effect on host metabolism and sequestration of free ribosomes by the circuit genes. We use the parameters determined by Weiße for the implementation of the host components.

For the purposes of comparison, all reactions and their corresponding rate equations are collected in Table I.

D. Numerical simulations

All models were simulated using ode15s in MATLAB 2015b. Parameter sweeps were conducted by drawing values at random from a discretized parameter space between 0 and a realistic maximum value for each parameter. Parameters which gave rise to oscillations were identified and used for analysis, specific values are presented in the main text.

III. RESULTS

A. Designing the synthetic oscillator

We find a number of parameter sets which lead to oscillations in both the isolated and host-aware model. However, we see large differences in the period and amplitude of the oscillations predicted between the two models (Fig. 2). In only one parameter set do the two models produce approximately the same period (120 minutes) (Fig. 2c) whilst in the others we see the host-aware model producing much longer periods - in many cases nearly twice the time period (Fig. 2a & b). Comparing the phase planes of the isolated model and host-aware models (Fig. 3) we can see two main effects of host-circuit interactions. In some instances the addition of host factors acts to dampen the oscillations observed in the isolated model (Fig. 3a). In others, the host factors stabilize the decaying oscillations observed in the isolated model (Fig. 3c).

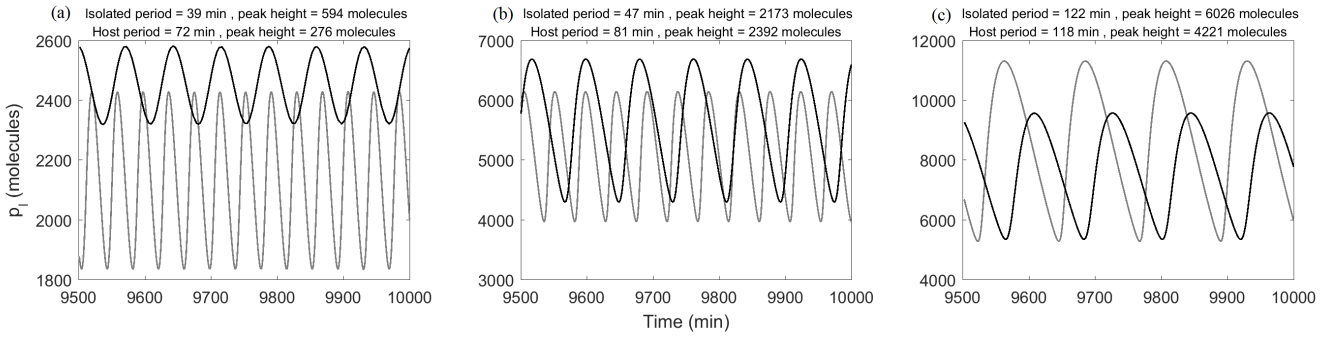

Fig. 2. Comparison of oscillation dynamics between the isolated and host-aware models. (Plots showing the oscillations produced by the isolated model (grey line) and host-aware model (black line) for the same circuit design. We see that the host-aware model predicts much longer periods than the isolated model (a and b) but in some parameter sets there is agreement (c).)

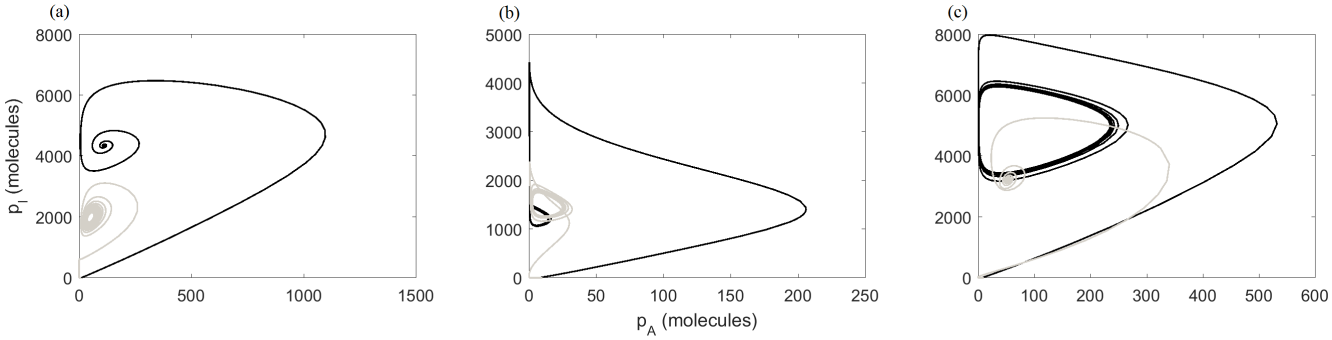

Fig. 3. Phase plane analysis of the isolated and host-aware models. (Phase planes of circuit protein concentrations demonstrating the observed differences between the isolated model (grey line) and host-aware model (black line). The effect of host-circuit interaction can act to dampen oscillations predicted by the isolated model (a), stabilise decaying oscillations predicted by the isolated model (c). In parameter sets where oscillations occur in both models these are of different amplitudes (b).)

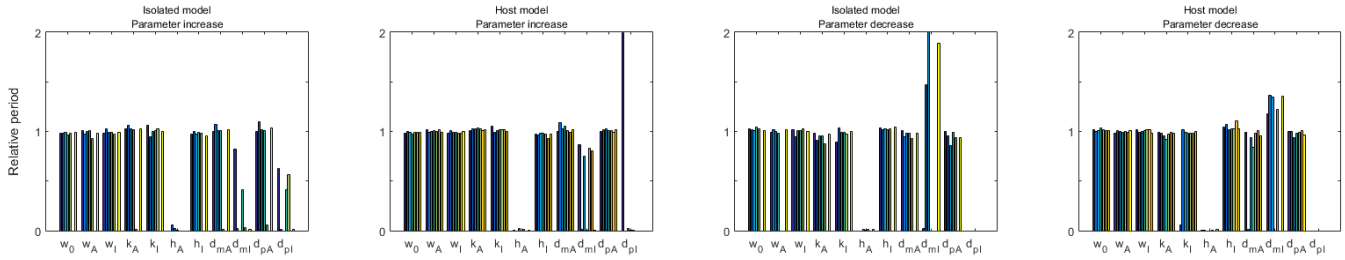

Fig. 4. Comparison of sensitivity analyses (Parameters common to both models were perturbed in turn for all parameter sets which produce oscillations in both models. Results are quoted as the ratio of the new period to the original period so 1 equals no change. Color of bar corresponds to parameter set. Y-axis limited to 2.)

B. Sensitivity analysis

For both the isolated and host-aware models, we perturbed each of the model parameters in turn and assessed the sensitivity of the calculated period on that parameter. Whilst both models produce similar profiles there are large differences in the sensitivities of the decay parameters (Fig. 4). The ratio of new period to the original in the host-aware model deviates from 1 less than in the isolated model. Some parameters when perturbed in the isolated model result in simulation failure but these perturbations are successfully simulated in the host-aware model. We see generally smaller changes in the period of oscillations in the host-aware model for most parameters, implying that host-circuit interactions may increase circuit robustness.

C. Effective host parameters

The isolated model requires the estimation of effective protein synthesis and growth rates. To investigate the effect these estimations, we select a parameter set and vary k_{eff} and λ_{eff} in the isolated model. We find that the emergence or stabilization of oscillations is highly dependent upon choice of the 'lumped' effective host parameter values. We find that at realistic growth rates ($<$0.025 per minute) the number of oscillating parameters is very small (Fig. 5). The isolated model produces oscillations with periods varying between 65 to 434 minutes depending upon the value of k_{eff} chosen.

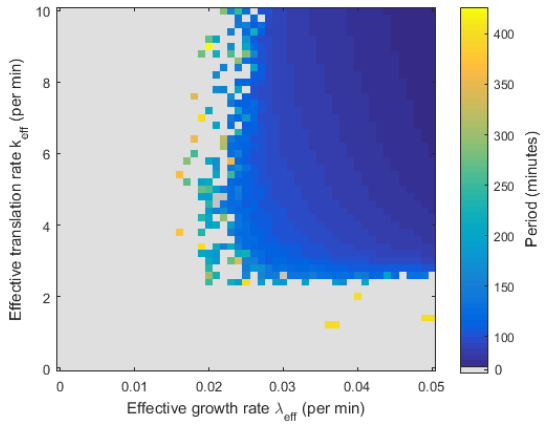

Fig. 5. Effect of varying approximated host parameters (k_{eff} and λ_{eff}) on the period of oscillations in the isolated model. Circuit parameters are kept constant.

We simulate the same parameter set in the host-aware model which does not requires these 'lumped' parameters as the model generates its own translation and growth rates as global properties of the model. In the host-aware model this parameter set yields sustained oscillations with a period of 334 minutes. The agreement between the two models is highly dependent upon the chosen value of k_{eff}.

However, estimating the value of k_{eff} from literature values is difficult as the literature contains a range of average translation rates measured under a variety of growth conditions. Additionally, we base k_{eff} on the ribosome elongation rate only, but this effective rate represents ribosome initiation, elongation and termination all of which have different rates. We ignore the ribosome mRNA binding event.

We find that the choice of values for these 'lumped' parameters has a large impact on the predicted behaviour of the oscillator circuit. The need for these 'effective host' parameters can introduce significant inaccuracies into model predictions and hence extend the circuit design cycle.

IV. Discussion and Conclusion

In this work we have considered the design of a synthetic oscillator using modelling frameworks that neglect and include host-circuit interactions. We find that the inclusion of tradeoffs due to resource and energy limitations in the model greatly changes the predicted dynamic behaviour of the circuit design, when compared to a model lacking these host details.

Our sensitivity analysis shows that for the parameter sets tested oscillation periods appear more stable to parameter perturbations in the host-aware model than in the isolated model.This suggests that taking into account host-circuit interactions may in some cases make the design problem easier.

Synthetic oscillators are often designed to act as timers for downstream circuitry, and it is thus imperative that when such circuits are implemented they exhibit periods close to those specified in their design. We observe that in some parameter regimes host-circuit interactions can destroy oscillations predicted in the isolated model, while in others they can stabilise decaying oscillations. Thus, guidance on appropriate choices of parameter values for *in vivo* implementation of oscillators that is derived from isolated circuit models may be misleading, and lead to non-functional designs. Explicit consideration of models containing resource limitations during the design cycle of synthetic circuits has the potential to significantly increase the efficiency and robustness with which such circuits can be implemented *in vivo*.

TABLE II

Parameter sets for Fig. 2, 3 and 5

Throughout we set both protein lengths to 300 amino acids and the ribosome association and dissociation constants to 1 ($b_A = b_I = u_A = u_I = 1$). The energy-dependent scaling factor θ in the $T_X(E)$ expression is set at 4.38 molecules per cell for all foreign genes. Units: w_G, molecules per minute per cell, k_G molecules per cell, h_G no units, d_G per minute.

	Fig. 2a	Fig. 2b	Fig. 2c	Fig. 3a	Fig. 3b	Fig. 3c	Fig. 5
w_0	0.03	0.10	0.03	0.10	0.08	0.01	0.08
w_A	889	778	444	667	1000	444	444
w_I	444	667	556	778	1000	444	556
k_A	1667	1667	1111	2222	1667	1667	3333
k_I	1667	3889	5000	3889	1111	5000	3333
h_A	1	1	1	1	1	1	1
h_B	6	9	8	7	7	9	3
δ_{m_A}	0.67	0.67	0.33	1.00	0.89	0.67	0.78
δ_{m_I}	0.22	0.11	0.00	0.00	0.11	0.00	0.00
δ_{p_A}	0.56	0.33	0.56	0.33	1.00	0.78	0.33
δ_{p_I}	0.00	0.00	0.00	0.44	0.00	0.22	0.00

References

[1] S. Cardinale and A. P. Arkin, "Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems," *Biotechnology Journal*, vol. 7, no. 7, pp. 856–866, 2012.

[2] F. Moser, N. J. Broers, S. Hartmans, A. Tamsir, R. Kerkman, J. A. Roubos, R. Bovenberg, and C. A. Voigt, "Genetic circuit performance under conditions relevant for industrial bioreactors," *ACS Synthetic Biology*, vol. 1, no. 11, pp. 555–564, 2012.

[3] C. Vilanova, K. Tanner, P. Dorado-Morales, P. Villaescusa, D. Chugani, A. Frías, E. Segredo, X. Molero, M. Fritschi, L. Morales, D. Ramón, C. Peña, J. Peretó, and M. Porcar, "Standards not that standard." *Journal of biological engineering*, vol. 9, p. 17, jan 2015.

[4] S. Cardinale, M. P. Joachimiak, and A. P. Arkin, "Effects of genetic variation on the E. coli host-circuit interface," *Cell Reports*, vol. 4, no. 2, pp. 231–237, 2013.

[5] F. Ceroni, R. Algar, G.-B. Stan, and T. Ellis, "Quantifying cellular capacity identifies gene expression designs with reduced burden," *Nature Methods*, vol. 12, no. 5, pp. 415–423, 2015.

[6] A. Y. Weiße, D. A. Oyarzún, V. Danos, and P. S. Swain, "Mechanistic links between cellular trade-offs, gene expression, and growth," *Proceedings of the National Academy of Sciences*, p. 201416533, 2015.

[7] A. Gyorgy, J. I. Jiménez, J. Yazbek, H.-H. Huang, H. Chung, R. Weiss, and D. Del Vecchio, "Isocost Lines Describe the Cellular Economy of Genetic Circuits," *Biophysical Journal*, vol. 109, no. 3, pp. 639–646, 2015.

[8] Y. Qian and D. del Vecchio, "Effective interaction graphs arising from resource limitations in gene networks," *Proc. of American Control Conference*, 2015.

[9] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, "A fast, robust and tunable synthetic gene oscillator." *Nature*, vol. 456, no. 7221, pp. 516–519, 2008.