

Precision Design of Expression from RNA Replicons

Jacob Beal

Raytheon BBN Technologies
Cambridge, MA, USA
jakebeal@bbn.com

Tyler E. Wagner

Center for Synthetic Biology
Boston University
Boston, MA, USA
wagnert@bu.edu

Tasuku Kitada

Biological Engineering
MIT
Cambridge, MA, USA
kitada@mit.edu

1. MOTIVATION

RNA replicons are an emerging platform of increasing interest, particularly for vaccination and therapeutic applications [3]. A replicon is based on a virus, but replaces the infective capsid proteins with engineered “payload” genes [5]. Here we focus on replicons derived from alphavirus, a positive-strand RNA virus, with architecture and lifecycle shown in Figure 1: the replicon RNA begins with a complex of non-structural proteins (NSPs) that create “viral factories” where it replicates [4]. A subgenomic promoter next induces production of shorter mRNAs containing engineered payload genes, which are translated to produce the proteins encoded by the payload sequences. Finally, both mRNA and proteins are removed by normal processes of dilution and decay.

Replicons provide distinct advantages as a platform for synthetic biology: as they are based on RNA, there is no direct path to affect cellular DNA, which reduces safety issues for medical applications. Unlike ordinary mRNA, they self-amplify, producing much stronger gene expression and over a longer time period, but can trigger a cell’s immune response, thereby curtailing expression. For a full and referenced discussion of the background of this work, see [1].

In order to enable rapid engineering of replicon-based systems, we have characterized the expression dynamics of Sindbis replicons in baby hamster kidney BHK-21 cells with constitutively expressed payloads. From this characterization, we construct a quantitative model that predicts expression in three-replicon systems with better than 2-fold accuracy. This predictive model can then be inverted to produce an algorithm for designing transfection mixtures to produce either desired expression ratios or absolute expression levels.

2. QUANTITATIVE EXPRESSION MODEL

To characterize the expression dynamics for Sindbis replicons in BHK-21 cells, we carried out two experiments: a logarithmic dose-response test of a single species of replicon over a 50-hour timespan, and a linear titration of pairs of co-transfected replicons expressing different fluorescent proteins. Both experiments use the TASBE characterization method [2] to obtain calibrated flow cytometry data in absolute units; in particular, Molecules of Equivalent Fluorescein (MEFL). Full details of this characterization and modeling are given in [1]; for this abstract we merely summarize the key results and the model derived.

IWBDA 2014 Boston, Massachusetts, USA

This work was sponsored by DARPA DSO under grant W911NF-11-054; the views and conclusions contained in this document are those of the authors and not DARPA or the U.S. Government..

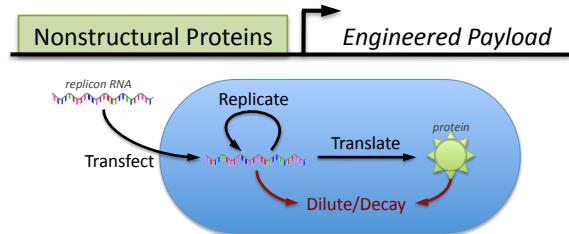


Figure 1: General architecture of RNA replicon design and expression dynamics.

Based on these experiments, we construct the following model: following a short initial delay of $\delta_E = 4.02$ hours as the replicons amplify, fluorescence converges exponentially with half-life $\lambda_E = 5.86$ hours towards a dose-independent translation-limited mean saturation level of $S = 5.44e7$ MEFL. This saturation level is modulated by a log-normal distribution of cell variance $V = 10^{\mathcal{N}(0,\sigma)}$, where the standard deviation $\sigma = 0.365$. The initial transient also appears dose-dependent, but cannot be quantified from these experiments.

Not all cells are transfected, and those that are transfected receive varying initial dosages of plasmids, which affects the final ratio of fluorescence. Cells are transfected with a Poisson distribution dependent on cell variance, receiving an initial “founder population” of each replicon of $f_i = Pois(\alpha \cdot V \cdot d_i)$, where $\alpha = 0.0127$ and d_i is the initial dose of replicon i . This initial dose is then amplified by a Polya Urn process (modeling the transition, as the number of replicons grows, from high sensitivity to stochastic effects to a stable ratio), giving an expected i th replicon proportion

$$p_i = (1 - \sum_{j < i} p_j) \cdot Beta(f_i, \sum_{j > i} f_j)$$

where $Beta(x, y)$ is the beta distribution (the ratio limit of a Polya Urn process). Transfected cells are thus predicted to have a distribution of expressions for the i th replicon at time t of:

$$E(t, i) = V \cdot p_i \cdot max(0, S(1 - 2^{\frac{\delta_E - t}{\lambda_E}}))$$

Finally, as replicon expression rises, the growth of transfected cells drops relative to untransfected. The fraction of expressing cells for total dose d at time t is thus:

$$F(d, 0) = \tau \cdot P(Pois(\alpha \cdot d) > 0)$$

$$F(d, t) = \frac{F(d, 0)}{F(d, 0) + (1 - F(d, 0)) \cdot max(1, 2^{\frac{\delta_F - t}{\lambda_F}})}$$

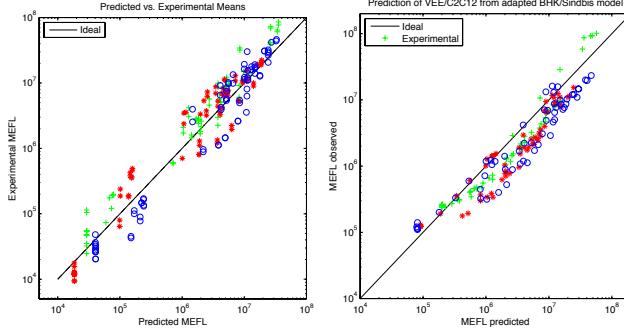


Figure 2: The quantitative expression model gives high precision predictions across a wide range of expression levels both for Sindbis/BHK (left) and VEE/C2C12 (right).

Where $\tau = 0.977$ is the maximum transfection efficiency, $\delta_F = 21.5$ hours is the time before significant impairment of transfected cells, and $\lambda_F = 8.89$ hours is the relative doubling time for untransfected cells after that point.

We have validated this model with a collection of three-replicon mixtures, chosen to test the model with dosage ratios ranging across two orders of magnitude and expression levels ranging across three. Measured at various time points up to 50 hours post-transfection, we find this model provides precise predictions of the mean expression (Figure 2), with a mean prediction error of only 1.7-fold. Model generality was tested by transfer to VEE replicon in C2C12 myoblasts, deriving parameters from a new set of experiments and adding an immune response term, to produce similar accuracy.

3. EXPRESSION DESIGN ALGORITHM

Given this forward model of expression, we can derive an algorithm for designing replicon mixtures with a desired combination of expression levels. In particular, we will design for the peak level of mean expression and peak fraction of cells expressing all elements in the mixture. Precise per-cell control of stoichiometry is, unfortunately, not an option given the degree of cell-to-cell variation, but as we will see the peak mean expression can be accurately designed.

We begin with an assumption that δ_E , λ_E , δ_F , and λ_F are such that cells reach near-peak expression before there is a significant shift in the ratio of expressing to non-expressing cells. This is the time we target, designing for a relative expression of R_i for the i th replicon and for a yield of Y fraction of cells expressing all replicons. The yield for the specified ratio, at a given minimum dosage d_m is:

$$Y = \tau \cdot \prod_i \left(1 - \int e^{-\alpha d_m \frac{R_i}{\min\{R_i\}}} 10^{\sigma x} \phi_{0,\sigma}(x) dx\right)$$

where the integral computes the expected fraction of cells not receiving the i th replicon ($e^{-\lambda}$ for a Poisson of parameter λ) with respect to the log-normal distribution of per-cell variation ($\phi_{0,\sigma}$ being the probability density function for a normal distribution with mean 0 and standard deviation σ). Although complicated, this function is monotonic in d_m , so it is easy to solve numerically for a dosage d_m to produce the specified yield Y . The full set of dosages then proceeds directly from the ratio: $d_i = d_m \cdot \frac{R_i}{\min\{R_i\}}$.

With this algorithm for designing ratios, we can also derive an algorithm for designing absolute expression levels.

Specification Peak Expression	Yield	Designed Dosages (ng)	Simulation	
			Peak Mean MEFL	Yield
1:1 Ratio	0.95	766, 766	2.21e7, 2.21e7	0.950
3:1 Ratio	0.50	212, 71	4.18e7, 1.60e7	0.537
5:2:1 Ratio	0.75	892, 357, 178	3.34e7, 1.21e7, 5.86e6	0.775
3e6 MEFL	0.80	171, 2920	2.50e6	0.799
1e6, 1e7 MEFL	0.50	56, 557, 2420	1.29e6, 1.11e7	0.520
2e7, 1e7 MEFL	0.95	1250, 627, 1530	1.66e7, 7.73e6	0.953

Figure 3: Examples of applying design algorithm, comparing specification with simulated behavior for design (ballast dose last in MEFL designs).

Given a set of desired peak mean expressions E_i , we compute a required “ballast” expression:

$$E_b = S - \sum_i E_i$$

and design for the ratios of the set of E_i and E_b .

Figure 3 shows examples of applying this design algorithm, comparing the specification to a simulation of the designed dosages on a population of 100,000 cells at 20 hours post-transfection using the stochastic model in the prior section. The simulation matches the specification closely, with a maximum expression error of 1.3-fold and a maximum yield error of 4%. Obviously, this could be further improved with additional tuning of the design algorithm, but the need is not urgent since these accuracies are significantly better than the accuracy of simulation for predicting experimental observations.

4. CONTRIBUTIONS AND FUTURE WORK

We have developed a replicon expression model that accurately predicts the behavior of multiple replicon/cell-line combinations, and from this model an algorithm for forward design of expression ratios or levels. Experimental results indicate that this model may be broadly applicable. Future work aims to extend to therapeutically relevant cell lines and *in vivo* systems, as well as regulatory interactions, thus enabling rapid precision engineering of medical applications.

5. ADDITIONAL AUTHORS

Additional authors: Andrey Krivoy (MIT), Odisse Azizgolshani (UCLA), Jordan Moberg Parker (UCLA), Douglas Densmore (BU), Ron Weiss (MIT)

6. REFERENCES

- [1] J. Beal, T. E. Wagner, T. Kitada, A. Krivoy, O. Azizgolshani, J. M. Parker, D. Densmore, and R. Weiss. Model-driven engineering of gene expression from RNA replicons. *submitted*.
- [2] J. Beal, R. Weiss, F. Yaman, N. Davidsohn, and A. Adler. A method for fast, high-precision characterization of synthetic biology devices. Technical Report MIT-CSAIL-TR-2012-008, MIT, April 2012.
- [3] K. Lundstrom. Alphaviruses in gene therapy. *Viruses*, 1:13–25, 2009.
- [4] D. Paul and R. Bartenschlager. Architecture and biogenesis of plus-strand rna virus replication factories. *World J. Virol.*, 2:32–48, 2013.
- [5] C. Xiong, R. Levis, P. Shen, S. Schlesinger, C. M. Rice, and H. V. Huang. Sindbis virus: an efficient, broad host range vector for gene expression in animal cells. *Science*, 243:1188–1191, 1989.