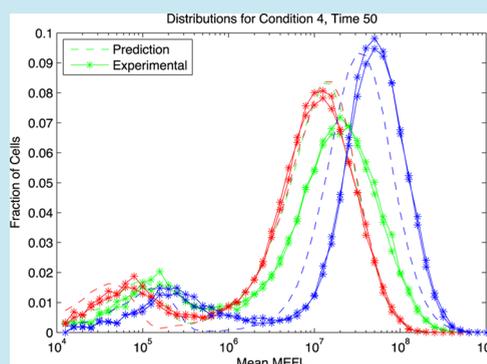


1 **Model-Driven Engineering of Gene Expression from RNA Replicons**2 Jacob Beal,^{*,†,‡} Tyler E. Wagner,^{‡,†} Tasuku Kitada,[§] Odisse Azizgolshani,^{||} Jordan Moberg Parker,[⊥]3 Douglas Densmore,[‡] and Ron Weiss[§]4 [†]Raytheon BBN Technologies, Cambridge, Massachusetts United States5 [‡]Center of Synthetic Biology, Boston University, Boston, Massachusetts 02215, United States6 [§]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States7 ^{||}Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, California 90095-1570, United States8 [⊥]Department of Microbiology, Immunology and Molecular Genetics, University of California Los Angeles, 609 Young Drive, Box
9 148906, Los Angeles, California 90095-1570, United States10 **S Supporting Information**

11 **ABSTRACT:** RNA replicons are an emerging platform for engineering
12 synthetic biological systems. Replicons self-amplify, can provide persistent
13 high-level expression of proteins even from a small initial dose, and, unlike
14 DNA vectors, pose minimal risk of chromosomal integration. However, no
15 quantitative model sufficient for engineering levels of protein expression from
16 such replicon systems currently exists. Here, we aim to enable the engineering
17 of multigene expression from more than one species of replicon by creating a
18 computational model based on our experimental observations of the
19 expression dynamics in single- and multireplicon systems. To this end, we
20 studied fluorescent protein expression in baby hamster kidney (BHK-21) cells
21 using a replicon derived from Sindbis virus (SINV). We characterized
22 expression dynamics for this platform based on the dose–response of a single
23 species of replicon over 50 h and on a titration of two cotransfected replicons
24 expressing different fluorescent proteins. From this data, we derive a quantitative model of multireplicon expression and validate
25 it by designing a variety of three-replicon systems, with profiles that match desired expression levels. We achieved a mean error of
26 1.7-fold on a 1000-fold range, thus demonstrating how our model can be applied to precisely control expression levels of each
27 Sindbis replicon species in a system.

28 **KEYWORDS:** quantitative modeling, circuit prediction, replicon, alphavirus, Sindbis, TASBE characterization, expression control,
29 flow cytometry



30 Recently, interest has been growing in RNA replicons as
31 standalone gene delivery vehicles.¹ Two of the primary
32 advantages of replicon-based systems are self-amplification
33 and safety, making replicons an attractive modality for medical
34 applications such as vaccine delivery, gene therapy, and cellular
35 reprogramming.^{2–4} Because they are self-amplifying, replicons
36 can generate high expression of a gene product, such as an
37 antigen, from a relatively low initial dose, compared to
38 nonreplicating RNA.^{5,6} The self-amplifying feature of replicons
39 is particularly valuable for mammalian systems, in which
40 replicating DNA-based vectors are uncommon and difficult to
41 construct due to their complexity.⁷ Moreover, with regard to
42 safety, replicons do not reverse transcribe and are confined to
43 the cytoplasm of the cell, so the risk of undesired integration
44 into the genome is minimal.^{8,9}

45 Of the various replicons that have been developed, we have
46 chosen to focus on a replicon derived from a Sindbis virus
47 (SINV), the most well-characterized alphavirus.⁵ Sindbis
48 replicons with reduced cytopathicity and increased duration
49 of expression have been developed.^{10–14} SINV is a positive-
50 strand RNA virus with an 11.7 kilobase genome. The SINV

genome has a 5′-7-methylguanosine cap and a 3′-poly(A) tail, 51
both of which are characteristics of cellular mRNA, allowing 52
SINV to utilize existing cellular machinery to initiate translation 53
of its nonstructural proteins.¹⁵ The 5′ two-thirds of the SINV 54
genome encodes the four nonstructural proteins that act 55
together to form a replicase complex. Replication occurs 56
through a minus-strand intermediate template, which is used 57
for the synthesis of both the full length positive strand genome, 58
as well as the subgenomic RNA that comprises the 3′ one-third 59
of the genome.^{15–18} In the wild type virus, the subgenomic 60
RNA is the source of the structural proteins that allow the virus 61
to propagate to other cells. These structural proteins can be 62
replaced with an engineered sequence to produce alternate 63
gene products, creating a *replicon* (Figure 1a): a general RNA 64
delivery vehicle that amplifies by self-replication, expresses 65

Special Issue: Summer Synthetic Biology Conference 2014

Received: March 7, 2014

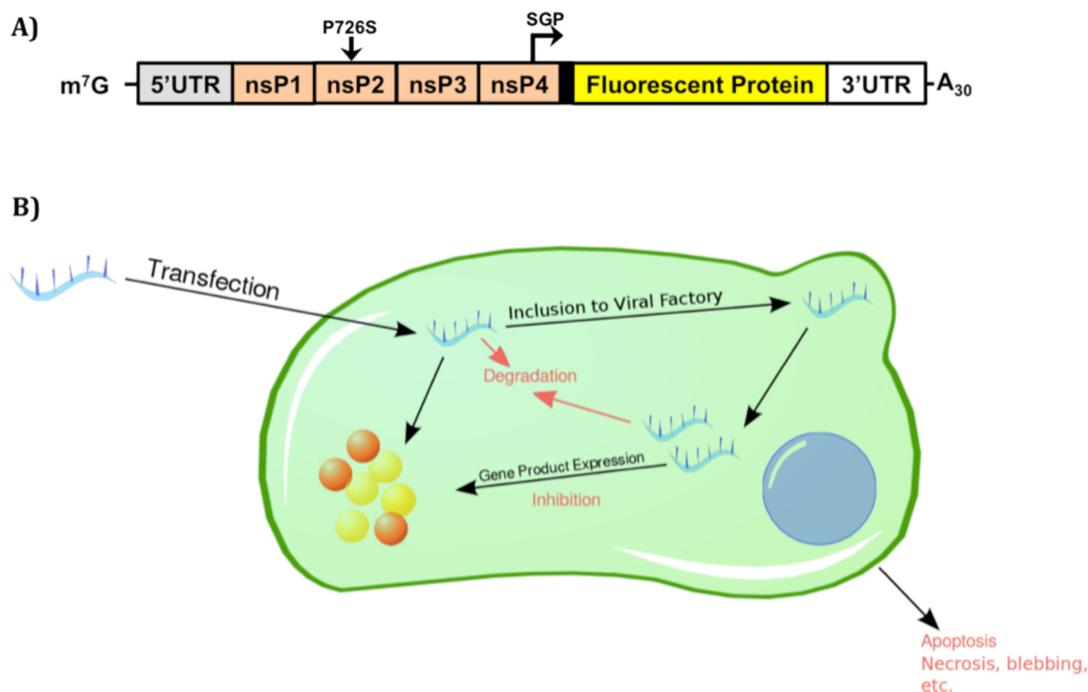


Figure 1. (a) We investigated replicon dynamics using Sindbis replicons with reduced cytopathicity¹⁰ containing the nsP2 P726S mutation (top), in which the portion of the RNA sequence coding for the SINV structural proteins is replaced by an engineered “payload” sequence, such as a fluorescent protein. (b) In replicon-based expression, initially a set of “founder” replicons enter the cell during transfection. The nonstructural proteins (nsPs) expressed from the replicons create “viral factories” in which the number of replicons is amplified exponentially. This process also produces subgenomic transcripts containing the “payload” sequence, which are expressed into proteins. Expression levels may be limited by the ordinary processes of degradation and dilution, as well as resource limits, antiviral innate immune response, and gross cell effects such as apoptosis, necrosis, and blebbing.

66 specified gene products, and lacks the necessary structural
67 proteins that would allow it to infect other cells.^{19–21}

68 Initial work using replicon systems has focused mainly on
69 producing a single protein product.¹⁵ Many applications,
70 however, may require more sophisticated replicon-based
71 circuits with multiple gene products at various different
72 expression levels; these products may also need to be separately
73 regulated or to regulate one another. A number of approaches
74 to express multiple genes from replicons have been
75 investigated, including the use of multiple subgenomic
76 promoters,^{22–25} insertion of internal ribosome entry sites
77 (IRES),^{25–27} and use of 2A sequences.^{28–30} Such approaches,
78 however, have a number of limitations, including decreased or
79 uneven expression, RNA recombination, and inherent scaling
80 limits on the size of an effective replicon.^{31–33} Moreover, many
81 RNA regulatory mechanisms, such as RNAi or those involving
82 cellular nuclease complexes, operate via RNA degradation and
83 thus will act on the replicon as a whole, thereby preventing the
84 use of these mechanisms in single replicon systems to control
85 individual gene products.³⁴

86 In this manuscript, we adopt an alternate approach based on
87 cotransfection of multiple replicons. With this approach, the
88 coding sequences of a system are distributed onto more than
89 one species of replicon, and the system is installed by mixing
90 these species together in an appropriate titration and
91 cotransfecting. This approach overcomes issues of replicon
92 size and allows each replicon species to be titrated and
93 regulated independently. Similar to multiplasmid systems,
94 however, it introduces variability into the stoichiometry of the
95 system, including the possibility that a cell may receive only
96 part of the system. Precise regulation of a gene product from a

single replicon may also benefit from this approach: as we show 97
in this manuscript, single replicon systems tend to saturate to 98
the same final expression level, independent of initial dose, 99
which poses a problem when engineering expression of a single 100
protein product. Using the cotransfection model that we 101
develop in this manuscript, however, the expression of a desired 102
protein can be tuned using a second “ballast replicon” that 103
competes for expression resources. 104

Thus, we develop cotransfection as a method to enable the 105
precision engineering of protein expression from replicons by 106
constructing a predictive quantitative model of replicon 107
expression dynamics. To this end, we measure the dose– 108
response behavior of a single species of Sindbis replicon in 109
BHK-21 cells over time, and of varying titrations of two 110
cotransfected replicon species. From this data, we derive a 111
quantitative model of multireplicon expression, which we then 112
validate by applying it to design a collection of six three- 113
replicon systems with a range of expression profiles. The 114
experimental behavior of these systems matches our desired 115
expression levels to a mean accuracy of 1.7-fold on a 1000-fold 116
range, thus demonstrating the capability of our model to 117
precisely control gene expression of each Sindbis replicon in a 118
system. 119

■ RESULTS 120

The design of the Sindbis replicons we used and the main 121
processes involved in expression from such replicons are shown 122
in Figure 1. First, the transfection process introduces some 123
number of “founder” replicons into each cell. Using the native 124
translational machinery, these replicons begin to express 125
proteins, specifically a set of nonstructural proteins used for 126

127 replication. These proteins, through processes not yet fully
 128 understood, modify endosomes and lysosomes to create “viral
 129 factories” in which they replicate the RNA.^{35,36} The replication
 130 process also produces transcripts of the 3′ subgenomic
 131 sequence or sequences encoding an engineered “payload,”
 132 which are translated to express the encoded proteins. These
 133 proteins are then degraded by typical cellular mechanisms.
 134 Protein expression may be inhibited by type-I interferon (IFN)
 135 mediated cellular antiviral innate immune responses,³⁷ or by
 136 gross cellular events such as apoptosis, necrosis, and blebbing.
 137 Here, we use BHK-21 cells (an IFN deficient cell line) as the
 138 host cell for Sindbis replicon expression to reduce possible
 139 complications that the innate immune response may have on
 140 gene expression. Expression is also self-regulated by the
 141 nonstructural protein complex, which self-cleaves and alters
 142 the polarity of the RNA-dependent RNA replication complex.¹⁵
 143 This mechanistic model implies a set of hypotheses regarding
 144 the expression patterns that may be expected in a replicon
 145 system with a constitutively expressed stable reporter protein,
 146 such as mVenus, mKate, or EBFP2. First, as transfection is a
 147 stochastic chemical process, the fraction of cells transfected
 148 should be dose-dependent. Moreover, if the replicons are well-
 149 mixed and transfections proceed independently (a reasonable
 150 null hypothesis), the number of transfections per cell would be
 151 expected to follow a Poisson distribution. Possible behaviors
 152 range between two asymptotic phases (Figure 2a): a “sparse”
 153 phase for lower doses, in which a small fraction of cells are

successfully transfected with a single founder replicon and
 fluoresce, and “dense” phase for higher doses, in which nearly
 all cells receive many founder replicons and fluoresce.
 Multireplicon systems will generally need to operate in the
 “dense” phase such that the probability distributions can be
 engineered to provide a high likelihood of all replicon species
 being delivered to each cell with good control of their
 stoichiometry.

For those cells that are transfected and fluoresce, the
 progress of fluorescent protein expression over time will be
 determined by the relative amount of available resources for
 replication and translation, or by alternative mechanisms for the
 reduction of expression, including antiviral innate immune
 responses or the degradation of RNA (Figure 2b). Replication
 is expected to initially be exponential, but the number of
 replicon copies must eventually saturate as either available
 resources become depleted or the replication process becomes
 inhibited by itself or the cellular host. At low replicon densities,
 translation is expected to be linear in the number of copies of
 the replicon, meaning that during the initial exponential
 replication phase, fluorescence will rise exponentially, tracking
 the rise in replicon copies with some delay. If translational
 resources become depleted or the translation process becomes
 inhibited, then fluorescence will rapidly converge to a stable
 expression level. If such a limit is not reached, on the other
 hand, then fluorescence will continue to increase until protein
 translation is balanced by dilution and/or decay—a long period
 of at least several cell divisions or protein half-lives.

Single Replicon Transfection. To determine which of
 these hypotheses best describes the expression dynamics of
 Sindbis replicons in BHK-21 cells, we measured the dose–
 response behavior of cells over the course of 50 h post-
 transfection (Figure 3a). We transfected BHK-21 cells with
 Sindbis replicon containing a gene for a fluorescent reporter,
 mVenus, at logarithmic doses (21, 41, 103, 206, 411, 1027,
 2055 ng per 1×10^5 cells). Fluorescence was measured using
 flow cytometry at 1, 3, 5, 7, 9, 11, 16, 19, 21, 26, 35, and 50 h
 post-transfection. We found that the transfected population of
 cells exhibits behavior consistent with being translation-limited,
 with an extremely sharp rise in mean fluorescence that rapidly
 saturates. Alternatively, if fluorescent expression were primarily
 replication-limited, the rate of convergence to saturation would
 be dominated by the process of achieving equilibrium between
 production and loss of protein, which should have a time
 constant on the order of the division time of the cells. Instead,
 we observe a much faster convergence consistent with
 translational resource depletion having at least a similar level
 of importance to replication rates: all dosages converged to a
 consistent expression level by 16 h and remained relatively
 stable thereafter (Figure 3a). Before that point, however, there
 is a monotonic relationship between dose and expression
 (Figure 3b), suggesting that, at lower dosages, limits on
 replication are also affecting the course of expression. With this
 data, we constructed an approximate model of mean protein
 expression using the following formula derived from the
 hypothesis that protein expression is strongly regulated by
 translational resources:

$$E(t) = \max(0, S(1 - 2^{\delta_E - t/\lambda_E}))$$

where $E(t)$ is the expression at time t , S is the saturated
 expression level, δ_E is an initial delay representing the early
 phase of the exponential replication process, and λ_E is the time
 constant on the exponential convergence toward saturation.

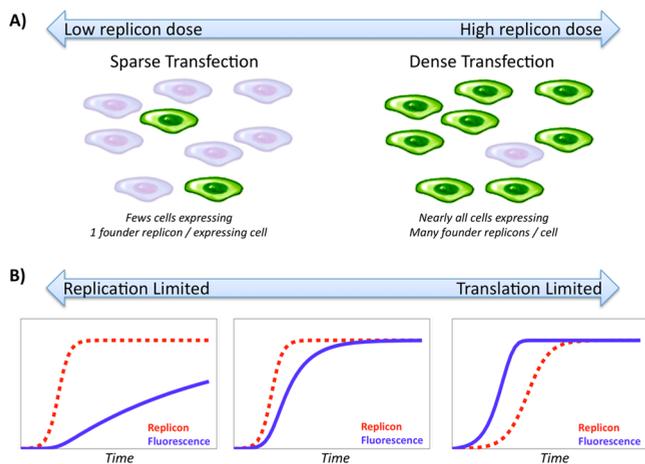


Figure 2. Depending on the relative dose and resources, a replicon system is expected to exhibit qualitatively different phases of behavior. (a) With a very low RNA dose (left), we expect to see only a few cells successfully transfected, and each transfected cell to receive only a single “founder” replicon. At the opposite extreme (right), high doses are expected to put many founder replicons into every cell, in proportion to their fraction of the initial dose. At intermediate levels, most cells will be transfected, but the relative stoichiometry of founder replicons will vary widely. (b) The translation and replication processes are both limited in the amount that they can drive resources. If the replication limit is reached much earlier than the translation limit (left), then expression will rise almost linearly for a long time before slowly reaching an equilibrium. In the opposite extreme, where the translational limit is reached while replication is still proceeding rapidly, then expression will rise exponentially to a sharp limit (right). When both limits are approached at similar rates, expression will rise quickly and then decelerate to an equilibrium. Graphs are ODE simulations varying relative rate constants by approximately 300X, plotted using linear axes.

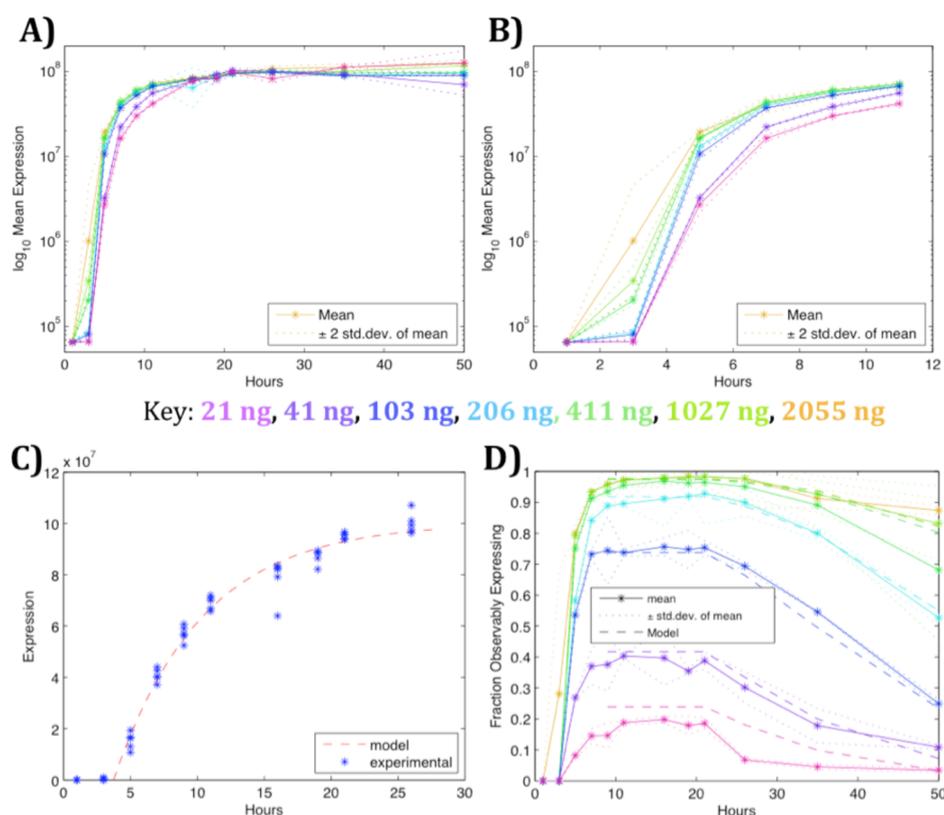


Figure 3. (a) Evolution of fluorescent expression over time for logarithmically varied doses of replicon, showing for all doses a rapid rise toward a consistent dose-independent level. Data shown is mean for upper (transfected) component of bimodal log-normal fit to fluorescence distribution. Dose is indicated by hue, ranging geometrically from 21 ng/ μ L (magenta) to 2055 ng/ μ L (yellow). (b) Detail of the first 11 h. (c) The rise toward saturation over the first 26 h shows that observations are well fit by a resource limitation model. (d) The initial fraction of transfected cells is dose-dependent following a Poisson distribution, then decreases as transfected cells stop or drastically slow their rate of division. Expression units are MEFL: molecules of equivalent fluorescein.

215 The delay before significant expression can be observed is due
 216 to the delivery and replication process; in principle, we should
 217 be able to model the dynamics of the development of viral
 218 factories and the dose-dependent rise of expression with respect
 219 to time. In this work, however, we restrict our model to the 1-
 220 DOF approximation of a fixed delay δ_E (rather than dose-
 221 dependent) because the signal from this process in our data is
 222 not sufficient to fit a more complex model within reasonable
 223 bounds (Supporting Information section 2), and (as will be
 224 seen) abstracting away dose-dependent delay is not the limiting
 225 factor on the precision of our predictions.

226 Figure 3c compares this model (fit to parameters $S = 1.05 \times$
 227 10^8 MEFL, $\delta_E = 4.02$ h, and $\lambda_E = 5.86$ h) against all of the time
 228 points gathered through 26 h. The model fits the data well, with
 229 the fit having a mean geometric error of 1.07-fold between
 230 model and prediction after 5 h, once detectable expression is
 231 expected to begin. Note, however, that although this model
 232 indicates that expression saturates due to a limit on the
 233 translation process, it does not disambiguate between other
 234 reasons that translation might decrease or cease, such as
 235 resource depletion, innate immune response, or cell death.

236 The size of the fluorescently expressing population over time
 237 can also be predicted from time-course data. Figure 3d shows
 238 that the fraction of observably expressing cells versus time is
 239 strongly dose dependent. After the first three time points
 240 (covering the first 5 h, in which expression is still too low to
 241 observe well at most dosages), the fraction of fluorescent cells is
 242 stable at first but then begins to decrease. The pattern of

decrease is consistent with a model in which replicons are 243
 initially distributed into cells following a Poisson distribution 244
 (per our mechanistic hypothesis above) and where cells that are 245
 not transfected continue to divide, while those that are 246
 transfected stop or drastically slow division once replicon and 247
 expression levels are high. We thus model the fraction of 248
 expressing cells: 249

$$F(d, 0) = \tau P(\text{Pois}(\alpha d) > 0)$$

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

where $F(d, t)$ is the fraction of cells expressing at time t given 250
 initial dose d , τ is an ideal transfection efficiency (modeling the 251
 fact that some cells may fail to be transfected regardless of 252
 dose), $P(\text{Pois}(\alpha d) > 0)$ is the probability of a sample being 253
 greater than zero when drawn from a Poisson distribution 254
 parametrized by dose times a constant α , δ_F is the delay before 255
 cell division is affected in transfected cells, and λ_F is the rate at 256
 which untransfected cells out-divide transfected ones. Figure 3d 257
 compares this model (fit to parameters $\tau = 0.977$, $\alpha = 0.0127$, 258
 $\delta_F = 21.5$ h, and $\lambda_F = 8.89$ h) against time points beginning at 9 259
 h, showing that it is a good fit, with a mean error of 2.5%. Note 260
 that although we have quantified the effect on division, it might 261
 actually be caused by any number of mechanisms: candidates 262
 include resource sequestration, toxicity from the action of the 263
 replicon, innate immune response, or some mixture thereof. 264
 This data also does not distinguish between failure to divide 265

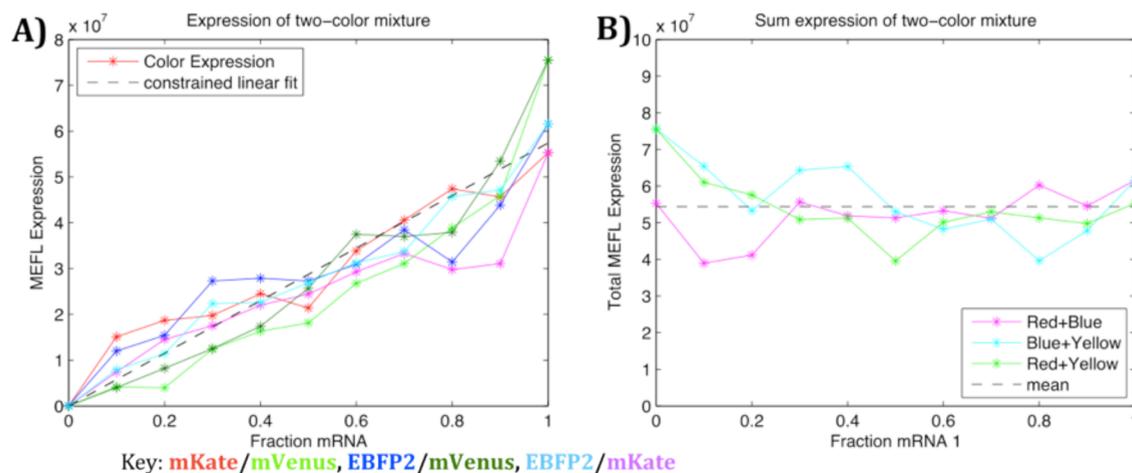


Figure 4. Cotransfection of two replicons at varying titrations and a constant high total dose produces a linear relation between relative dose and fluorescence (a) and a constant total fluorescence (b), as predicted from single-replicon models. Expression units are MEFL: molecules of equivalent fluorescein.

266 and slow division: our model only states how much *faster* the
 267 untransfected population is dividing than the transfected
 268 population.

269 Another interesting implication of this model is that it
 270 appears that no replicon dosage can be low enough to cause
 271 this particular replicon/cell system to be primarily limited by
 272 replication. Given both the value of α that has been identified
 273 and the large number of nonexpressing cells at low dosage, we
 274 may reasonably hypothesize that many cells in the fluorescently
 275 expressing population are initially transfected with only a single
 276 replicon. In the 21 and 41 ng dose conditions, these cells are
 277 expected to make up the vast majority of the expressing
 278 population, and in the 103 ng dose condition are expected to be
 279 48% of the expressing population. Thus, further reduction of
 280 the initial dose may be expected to further decrease the fraction
 281 of expressing cells but will not result in any significant further
 282 shift toward expression being primarily limited by replication.

283 **Multi-replicon Transfection.** In the discussion above, we
 284 established a model that estimates the fraction of cells with
 285 active replicons and the mean expression from a single species
 286 of replicon within these cells. For engineering systems based on
 287 multi-replicon cotransfection, however, it will also be important
 288 to have a model of the degree of variation anticipated from cell
 289 to cell, since small variations in the founder population of
 290 replicons may lead to large differences in the relative
 291 populations following exponential replication. To quantify
 292 variation, we transfected BHK-21 cells with pairs of replicons
 293 that are identical except for the choice of fluorescent reporter:
 294 mVenus/mKate, mKate/EBFP2, or mVenus/EBFP2. Each
 295 sample was transfected with 1390 ng total RNA, at varying
 296 titrations of the two replicons ranging from 0% to 100% in
 297 steps of 10%. Fluorescence was measured via flow cytometry 20
 298 h post-transfection (Figure 4). Based on our prior models, we
 299 expected transfection to be dense, with many founder replicons
 300 per cell. By the law of large numbers, this means that the mean
 301 fluorescence per cell should be approximately linear with
 302 respect to the relative dose (though they may be somewhat
 303 distorted by variation) and when converted to equivalent
 304 MEFL units the sum of the two fluorescence intensities should
 305 be equal to a constant. These predictions are borne out by our
 306 observations (Figure 4). We also refine our estimate of S to
 307 5.44×10^7 MEFL using the means of dual transfections, as the

geometric means of this collection of high-dose conditions 308
 provide a more reliable estimate than the bimodal Gaussian fit 309
 used for tracking small expressing populations in the prior 310
 experiment. 311

When considering distribution of fluorescence, however, 312
 variation is much greater than can be accounted for only by a 313
 pure Poisson distribution. We note two additional sources of 314
 variation that are known to have large effects in biological 315
 systems. First, most cells (including the BHK-21 cells we are 316
 using) vary significantly in their size and state; this is expected 317
 to affect both the resources available for expression and the 318
 amount of founder population of replicons introduced during 319
 transfection. Observing that single-replicon experiments show a 320
 log-normal distribution of fluorescence (Supporting Informa- 321
 tion Figure 1), we model a cell variation factor V as a log- 322
 normal distribution: 323

$$V = 10^{N(0,\sigma)}$$

where $N(0,\sigma)$ is a normal distribution with standard deviation 324
 σ . We include this variation factor as a multiplier in the 325
 distribution of founder replicons, reflecting the hypothesis that 326
 larger cells are likely to take up proportionally more replicons: 327

$$f_i = \text{Pois}(\alpha V d_i)$$

where f_i is the number of founder replicons of the i th replicon, 328
 d_i is the partial dose of the i th replicon, and α is the same 329
 parametrization constant that we fit previously in modeling the 330
 fraction of transfected cells. 331

The second major known source of variation is the 332
 stochasticity of biochemical processes when there are only a 333
 few molecules. In many cases, the number of founder copies of 334
 each replicon expected in a given cell may be quite low, in 335
 which case the first few replication events may have a large 336
 impact on the ultimate ratio of the two cotransfected replicons. 337
 This can be modeled mathematically as a Polya Urn process.³⁸ 338
 The Polya Urn process considers an urn that begins with k 339
 balls, each of which has some color; for each of n iterations, a 340
 ball is drawn, then replaced along with an extra ball of the same 341
 color, increasing the number of balls in the urn. This model has 342
 the property that when the number of balls is small, the ratios 343
 can shift significantly, but as the process continues the ratios 344
 drift less, ultimately converging to a value distributed by a β 345

346 distribution parametrized on the initial set of balls. Applied to
 347 model our replicon systems, the colors are the replicon species,
 348 the initial balls are the founder copies, and the draw and
 349 replacement process is replication events, giving a model of

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

350 where p_i is the proportion of the i th replicon following
 351 replication and $\beta(x,y)$ is the beta distribution.

352 Note that this model of replication via a Polya Urn process is
 353 unlikely to be strongly affected by how replicons are distributed
 354 in viral factories, the process by which viral factories are formed,
 355 or the process by which RNA escapes the viral factories.
 356 Independent factories, each replicating its own collection of
 357 replicons, may be viewed as members of a “meta-urn” that
 358 produces the same distribution of converged ratios, as long as it
 359 is the case that many replicons can fit in a factory. Likewise,
 360 independent factories may be translated in time without a
 361 significant effect, as long as the variance in the time to form a
 362 factory is not so long as to allow other factories to significantly
 363 affect the available transcriptional resources. From our results, it
 364 appears that neither of these cases holds strongly enough to
 365 prevent the Polya Urn model from being a reasonable
 366 approximation of observed behavior. This does not, however,
 367 constrain the system enough to shed any significant new light
 368 on the behavior of viral factories.

369 Combining these factors, we enhance our earlier model of
 370 mean protein expression for single replicons to a model giving a
 371 distribution of expression for multiple replicon species:

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

372 with the expression $E(t,i)$ of the i th replicon at time t varying
 373 from the mean according to the cell variation factor and the
 374 distribution of the replicons’ proportions implied by the initial
 375 dose. Fitting this model against the set of 90%/10% titrations
 376 (to maximize observed variation), we obtain $\sigma = 0.365$ (Figure
 377 5).

378 **Engineering Multi-replicon Systems.** We validated our
 379 expression model, as well as demonstrated how it can be used
 380 to engineer precision control of expression levels, by designing
 381 and predicting expression for a set of six three-replicon
 382 mixtures. These six mixture conditions, shown in Table 1,

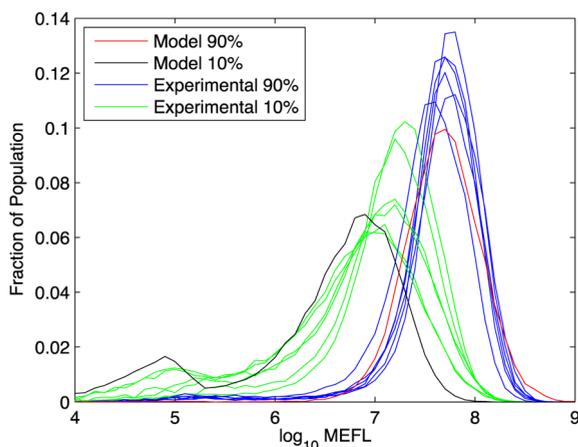


Figure 5. Experimental observations vs model of fluorescence distribution for all 10%:90% ratios of two-replicon transfection, in all six dosage/color combinations.

Table 1. Dosages for Three-Replicon Mixtures Used for Validation of Model-Driven Design

condition	mVenus (ng)	mKate (ng)	EBFP2 (ng)
1	180	180	180
2	540	540	540
3	180	900	720
4	360	360	1080
5	18	180	900
6	720	36	36

were chosen to evaluate the performance of our expression
 model on a variety of systems with dosage levels and dosage
 ratios ranging across approximately 2 orders of magnitude.
 BHK-21 cells were transfected with each replicon mixture at the
 specified dosages and fluorescence was measured by flow
 cytometry at 3, 6, 11, 20, 26, 34, and 50 h post-transfection. In
 all cases, our expression model provides a highly precise
 prediction of the observed value, with a mean prediction error
 of only 1.7-fold (Figure 6a). With respect to individual
 mixtures, there is no significant difference in prediction quality
 (Figure 6b). With respect to time, prediction errors are slightly
 lower at around 24 h, when the greatest number of cells are
 fluorescing most strongly, but show no significant pattern for
 when predictions are better or worse (Supporting Information
 Figure 2). Means versus time for each mixture are shown in
 Figure 6c and Supporting Information Figure 3. Finally, the
 expression model predicts well not only the mean but also the
 distributions of fluorescence within each sample (Figure 6d and
 Supporting Information Figures 4–10).

DISCUSSION

The work presented in this manuscript enables the use of RNA
 replicons as a predictable platform for synthetic biology. We
 have demonstrated that Sindbis replicon in BHK-21 cells has a
 highly systematic pattern of gene product expression over time.
 Expression can be predicted with high precision from the initial
 replicon dosage, using an 8-DOF model. This model was
 derived from the basic mechanisms of transfection, replication,
 and translation, and parametrized using experimental observa-
 tion of single- and dual-replicon transfections. Our model
 allows predictive engineering of a multireplicon system, as we
 have demonstrated by designing and precisely predicting the
 expression distributions of a collection of six mixtures of three
 replicons across a wide range of dosages and ratios.

The same approach can be applied to precision control of the
 expression from a single-replicon system, by titrating with a
 competing “ballast” replicon (note that changing the initial
 dosage of a single replicon does not change the per-cell
 expression level after the first few hours, only the fraction of
 cells successfully transfected). Although our work enables the
 immediate engineering of a wide range of RNA replicon
 expression systems, further investigation is necessary to support
 a full range of replicon-based applications. At present, our
 models only apply to constitutive expression from replicons of
 uniform size, with similar sized products, under control of
 identical subgenomic promoters. Replicons of different sizes
 may replicate at different rates, which would change the
 mapping from founder population to the distribution of
 converged ratios. Similarly, different gene products may require
 different resources, which may affect their share of the
 translational limit, and differences in subgenomic promoters
 are also expected to affect expression. Regulatory interactions

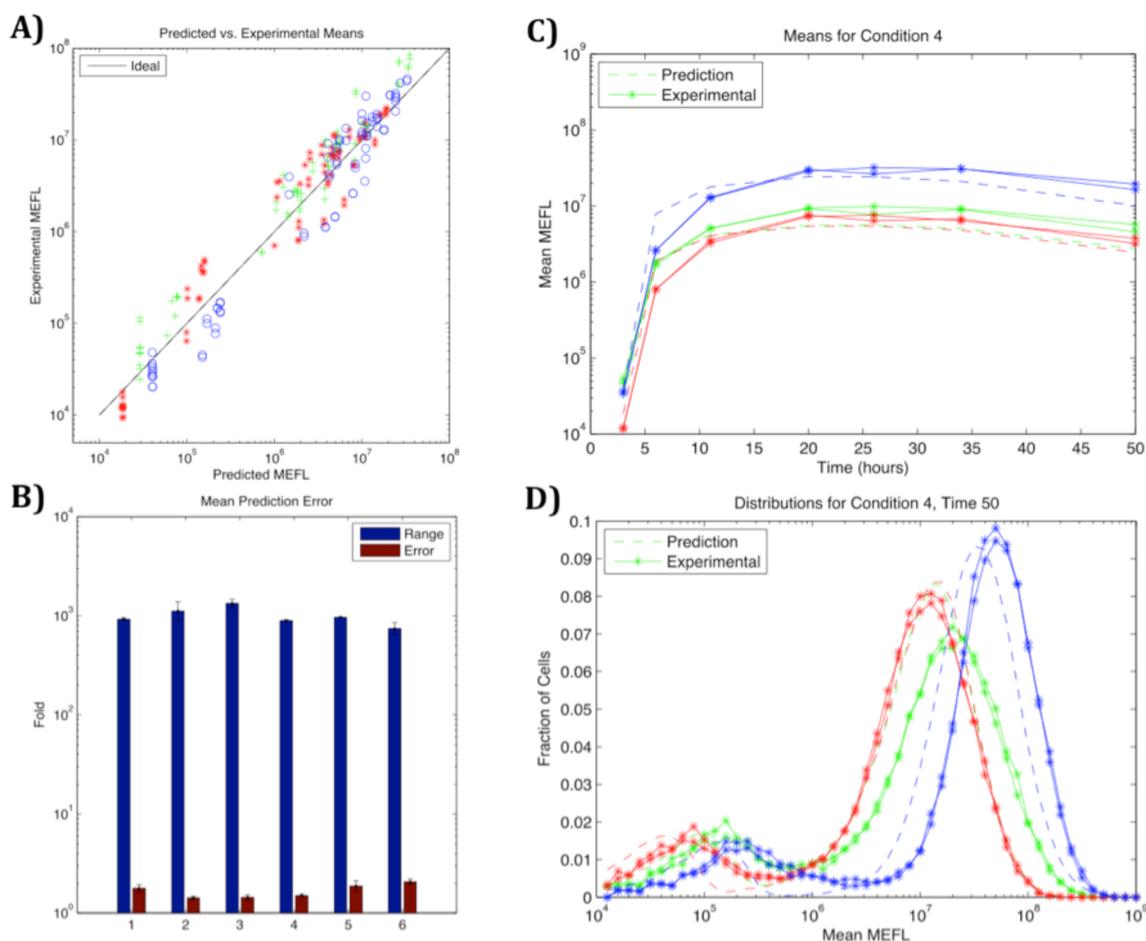


Figure 6. Our Sindbis/BHK expression model successfully predicts the evolution of expression distributions for six three-replicon mixtures across a wide range of initial dosages: (a) mean predicted vs measured expression for each color for all time/mixture combinations measured (mKate is red, EBFP2 is blue, and mVenus is green); (b) geometric mean fold error vs fold expression range for each of the six mixtures. Standard deviation is across replicates for range and across replicates and colors for error. (c, d) Examples of prediction detail: (c) predicted vs experimental evolution of mean expression for Mixture 4 (360 ng mVenus, 360 ng mKate, 1080 ng EBFP2) for 50 h post-transfection; (d) predicted vs experimental distribution of fluorescence values for Mixture 4 at 50 h post-transfection. Full details of predictions are shown in Supporting Information Figure 2 through 10. Expression units are MEFL; molecules of equivalent fluorescein.

434 between replicons will also affect expression dynamics,
 435 particularly when the mechanism of regulation involves RNA
 436 degradation. Finally, the nonstructural proteins of the replicon
 437 are translated as well; these are not included in our model at
 438 present because their effect can be factored out for constitutive
 439 expression from uniform replicons. However, if these non-
 440 structural proteins have a significant effect on expression, they
 441 too may need to be included in a more general computational
 442 model. In many cases, extending models to cover such systems
 443 is also likely to require a more explicit model of replication than
 444 the current coarse abstraction. Although the parameters of the
 445 system that we studied precluded building a comprehensive
 446 quantitative model from fluorescence data, such a model should
 447 be able to be acquired either from other replicons with lower
 448 levels of protein expression or by studying RNA levels directly,
 449 for example, via qRT-PCR. Understanding the contributions of
 450 different mechanisms to the translation limit will also be
 451 important for future engineering work.

452 A concern for certain applications is the fact that the BHK-21
 453 cells stopped or drastically slowed dividing *in vitro*, which
 454 indicates gross cell effects that may preclude the use of Sindbis
 455 replicons in many therapeutic applications. However, other
 456 combinations of replicons and cell lines may show less impact,

and published reports of sustained *in vivo* replicon expression 457
 are quite promising.³⁹ Since other replicon/cell combinations 458
 are generally expected to have the same underlying biochemical 459
 processes (except for the immune mechanisms that BHK-21 460
 lacks), we expect that the same general models will likely apply, 461
 though they will have different parameter values. 462

In summary, the methods and approach that we have 463
 presented here provide a solid foundation for developing a 464
 general capability for precision engineering of biological 465
 systems using RNA replicons. Beyond replicons, this work 466
 also provides an example of systematic establishment of a 467
 quantitative engineering method for a biological mechanism, 468
 some principles of which may be able to be extended more 469
 broadly to other biological systems, such as DNA-based 470
 transcriptional and translational gene circuits. Our model 471
 construction depends critically on the use of per-cell measure- 472
 ments, which allow distinction between expressing and 473
 nonexpressing subpopulations and expose systematic effects 474
 that would otherwise not be directly observable. It also depends 475
 on being able to translate all fluorescence measurements into 476
 equivalent absolute units, which allows data from different 477
 fluorescent proteins to be compared directly in the two- 478
 replicon titration experiments. Finally, we abstract mechanisms 479

480 where necessary in order to ensure that all parameters used by
481 our model are well-supported by experimental data. Applying
482 such methods in other contexts may likewise assist in the more
483 general development of improved biological engineering
484 models and methods.

485 ■ METHODS

486 **Cloning/RNA Generation.** All Sindbis constructs were
487 created from pSINV-EYFP using standard molecular cloning
488 techniques.⁴⁰ The replicon itself originated from the TE12
489 strain of Sindbis and was altered to contain the previously
490 characterized, less cytopathic P726S mutation in nsP2.^{10,11} The
491 plasmids were linearized using SacI prior to *in vitro* tran-
492 scription using the mMESSAGING mMachine SP6 Kit (Life
493 Technologies). The resulting RNA was purified using the
494 RNeasy Mini Kit (Qiagen) and the concentration was
495 measured using the NanoDrop 2000.

496 **Transfection.** All Sindbis transfections were conducted in
497 BHK-21 cells (a kind gift from Dr. James H. Strauss) cultured
498 in EMEM (ATCC) supplemented with 10% FBS (PAA) at 37
499 °C and 5% CO₂. BHK-21 cells at approximately 70%
500 confluence were electroporated using the Neon Transfection
501 System (Life Technologies) following optimization, according
502 to the manufacturers' instructions. In general, for a single well
503 of a 24-well plate (Corning), approximately 100 000 cells were
504 electroporated with RNA ranging from 18 to 2055 ng.

505 For the time-course experiment, samples were taken in
506 duplicate. For the dual transfection experiment, one sample was
507 taken for each titration, with the single replicon samples
508 (100%/0% titration) shared between color pairings. For the
509 three-replicon mixtures, samples were taken in duplicate. Note
510 that the statistical strength of results is not significantly
511 weakened by using smaller numbers of samples per condition,
512 due to the fact that conclusions are drawn from the joint
513 analysis of groups of samples across varying conditions.

514 **Flow Cytometry.** Cells for each time point were washed
515 with 1× PBS, trypsinized, and resuspended in 1× PBS. Flow
516 cytometry was performed using the BD LSRFortessa Flow
517 Cytometer System and FACSDiva software was used for initial
518 data collection.

519 **Statistical Analysis and Modeling.** Flow cytometry data
520 was converted from arbitrary units to compensated MEFL
521 (Molecules of Equivalent FLuorescein) using the TASBE
522 characterization method.⁴¹ An affine compensation matrix is
523 computed from single positive and blank controls. FITC
524 measurements are calibrated to MEFL using SpheroTech RCP-
525 30-5-A beads,⁴² and mappings from other channels to
526 equivalent FITC are computed from cotransfection of high
527 equal dose (>500 ng) of replicons identical except for the
528 choice of fluorescent protein. For dual transfection data, we use
529 the 50:50 condition for each pair; for mixture data we use the
530 26-h time point for condition 2. For Figure 3, expression data is
531 fit against a bimodal Gaussian on the log scale to distinguish the
532 subpopulation of cells where the replicon is successfully
533 transfected from the subpopulation where it is not. Unlike
534 gating thresholds, a bimodal Gaussian model allows good
535 estimation of relative population sizes even when subpopula-
536 tions overlap in their distribution of expression. Mean of
537 expressing population is then calculated as the mean of the
538 upper component of a bimodal log-normal fit to the observed
539 distribution, except for time points 35 and 50 of lowest dosage,
540 where expressing population is too small for a good fit, and we
541 instead select the distribution peak at highest MEFL. For all

other figures, where mixture of replicons skews distributions 542
away from log-normal, means are computed as geometric mean 543
of all samples >10⁴ MEFL. Histograms are generated by 544
segmenting MEFL data into logarithmic bins at 10 bins/decade, 545
with geometric mean and variance computed for those data 546
points in each bin. Model fits are performed using standard 547
least-squares curve fitting on a linear scale. The model of 548
expression rising to saturation is fit against all but the two 549
lowest dosages, which appear to be significantly affected by 550
replication dynamics. The model of fraction transfected is fit 551
against all but the lowest dosage, where expression was too low 552
in some cases for bimodal Gaussian fits to converge. All 553
parameter estimates are expressed to three significant figures for 554
consistency. 555

■ ASSOCIATED CONTENT

📄 Supporting Information

List of exceptions to replicate numbers, a discussion of choices 558
in modeling replication, replicon sequences, and full details of 559
all predictions. This material is available free of charge *via* the 560
Internet at <http://pubs.acs.org>. 561

■ AUTHOR INFORMATION

Corresponding Author

*Email: jakebeal@bbn.com.

Author Contributions

†J.B. and T.E.W. contributed equally to this work. J.B., T.E.W., 566
and R.W. designed experiments. T.E.W. performed the 567
experiments; T.K. assisted with experiments and performed 568
preliminary experiments. J.B. developed and applied computa- 569
tional analysis methods, developed computational models, and 570
developed methods for prediction and design of expression. 571
J.B., T.E.W., R.W., and T.K. interpreted results. O.A. and J.M.P. 572
supplied the replicon and expertise with the Sindbis replicon 573
system. T.K. supplied source DNA constructs. J.B. and T.E.W. 574
wrote the manuscript; all authors edited the manuscript. 575

Notes

The authors declare no competing financial interest. 576
577

■ ACKNOWLEDGMENTS

The authors thank John Scarpa for preparing Figure 1b (using 579
TinkerCell software, found at <http://www.tinkercell.com/> 580
Home/cite) and Andrey Krivoy for input into discussions. 581
The work presented in this manuscript was supported by 582
DARPA under grant W911NF-11-054. 583

■ ABBREVIATIONS

MEFL, molecules of equivalent fluorescein; EBFP2, enhanced 585
blue fluorescent protein 586

■ REFERENCES

- (1) Lundstrom, K. (2009) Alphaviruses in Gene Therapy. *Viruses* 1, 588
13–25. 589
- (2) Lundstrom, K. (2012) Alphavirus Vectors in Vaccine Develop- 590
ment. *J. Vaccines Vaccin.* 3, DOI: 10.4172/2157-7560.1000139. 591
- (3) Wahlfors, J. J., Zullo, S. A., Loimas, S., Nelson, D. M., and 592
Morgan, R. A. (2000) Evaluation of recombinant alphaviruses as 593
vectors in gene therapy. *Gene Ther.* 7, 472 DOI: 101038/sjgt3301122. 594
- (4) Yoshioka, N., Gros, E., Li, H.-R., Kumar, S., Deacon, D. C., 595
Maron, C., Muotri, A. R., Chi, N. C., Fu, X.-D., Yu, B. D., and Dowdy, 596
S. F. (2013) Efficient generation of human iPSCs by a synthetic self- 597
replicative RNA. *Cell Stem Cell* 13, 246–254. 598

- 599 (5) Xiong, C., Levis, R., Shen, P., Schlesinger, S., Rice, C. M., and
600 Huang, H. V. (1989) Sindbis virus: An efficient, broad host range
601 vector for gene expression in animal cells. *Science* 243, 1188–1191.
- 602 (6) Kulasegaran-Shylini, R., Thiviyathan, V., Gorenstein, D. G., and
603 Frolov, I. (2009) The 5'UTR-specific mutation in VEEV TC-83
604 genome has a strong effect on RNA replication and subgenomic RNA
605 synthesis, but not on translation of the encoded proteins. *Virology* 387,
606 211–221.
- 607 (7) Kouprina, N., Earnshaw, W. C., Masumoto, H., and Larionov, V.
608 (2013) A new generation of human artificial chromosomes for
609 functional genomics and gene therapy. *Cell. Mol. Life Sci. CMLS* 70,
610 1135–1148.
- 611 (8) Robertson, J. S. (1994) Safety considerations for nucleic acid
612 vaccines. *Vaccine* 12, 1526–1528.
- 613 (9) Klinman, D. M., Takeno, M., Ichino, M., Gu, M., Yamshchikov,
614 G., Mor, G., and Conover, J. (1997) DNA vaccines: Safety and efficacy
615 issues. *Springer Semin. Immunopathol.* 19, 245–256.
- 616 (10) Frolov, I., Agapov, E., Hoffman, T. A., Prágai, B. M., Lippa, M.,
617 Schlesinger, S., and Rice, C. M. (1999) Selection of RNA replicons
618 capable of persistent noncytopathic replication in mammalian cells. *J.*
619 *Virol.* 73, 3854–3865.
- 620 (11) Lustig, S., Jackson, A. C., Hahn, C. S., Griffin, D. E., Strauss, E.
621 G., and Strauss, J. H. (1988) Molecular basis of Sindbis virus
622 neurovirulence in mice. *J. Virol.* 62, 2329–2336.
- 623 (12) Heise, M. T., White, L. J., Simpson, D. A., Leonard, C., Bernard,
624 K. A., Meeker, R. B., and Johnston, R. E. (2003) An attenuating
625 mutation in nsP1 of the Sindbis-group virus S.A.AR86 accelerates
626 nonstructural protein processing and up-regulates viral 26S RNA
627 synthesis. *J. Virol.* 77, 1149–1156.
- 628 (13) Dryga, S. A., Dryga, O. A., and Schlesinger, S. (1997)
629 Identification of mutations in a Sindbis virus variant able to establish
630 persistent infection in BHK Cells: The importance of a mutation in
631 the nsP2 gene. *Virology* 228, 74–83.
- 632 (14) Perri, S., Driver, D. A., Gardner, J. P., Sherrill, S., Belli, B. A.,
633 Dubensky, T. W., and Polo, J. M. (2000) Replicon vectors derived
634 from Sindbis virus and Semliki Forest virus that establish persistent
635 replication in host cells. *J. Virol.* 74, 9802–9807.
- 636 (15) Strauss, J. H., and Strauss, E. G. (1994) The alphaviruses: Gene
637 expression, replication, and evolution. *Microbiol. Rev.* 58, 491–562.
- 638 (16) Hardy, W. R., and Strauss, J. H. (1989) Processing the
639 nonstructural polyproteins of sindbis virus: Nonstructural proteinase is
640 in the C-terminal half of nsP2 and functions both in cis and in trans. *J.*
641 *Virol.* 63, 4653–4664.
- 642 (17) Shirako, Y., and Strauss, J. H. (1990) Cleavage between nsP1
643 and nsP2 initiates the processing pathway of Sindbis virus non-
644 structural polyprotein P123. *Virology* 177, 54–64.
- 645 (18) Jose, J., Snyder, J. E., and Kuhn, R. J. (2009) A structural and
646 functional perspective of alphavirus replication and assembly. *Future*
647 *Microbiol.* 4, 837–856.
- 648 (19) Schlesinger, S., and Dubensky, T. W., Jr. (1999) Alphavirus
649 vectors for gene expression and vaccines. *Curr. Opin. Biotechnol.* 10,
650 434–439.
- 651 (20) Agapov, E. V., Frolov, I., Lindenbach, B. D., Prágai, B. M.,
652 Schlesinger, S., and Rice, C. M. (1998) Noncytopathic Sindbis virus
653 RNA vectors for heterologous gene expression. *Proc. Natl. Acad. Sci.*
654 *U.S.A.* 95, 12989–12994.
- 655 (21) Geall, A. J., Verma, A., Otten, G. R., Shaw, C. A., Hekele, A.,
656 Banerjee, K., Cu, Y., Beard, C. W., Brito, L. A., Krucker, T., O'Hagan,
657 D. T., Singh, M., Mason, P. W., Valiante, N. M., Dormitzer, P. R.,
658 Barnett, S. W., Rappuoli, R., Ulmer, J. B., and Mandl, C. W. (2012)
659 Nonviral delivery of self-amplifying RNA vaccines. *Proc. Natl. Acad. Sci.*
660 *U.S.A.* 109, 14604–14609.
- 661 (22) Hahn, C. S., Hahn, Y. S., Braciale, T. J., and Rice, C. M. (1992)
662 Infectious Sindbis virus transient expression vectors for studying
663 antigen processing and presentation. *Proc. Natl. Acad. Sci. U.S.A.* 89,
664 2679–2683.
- 665 (23) Frolov, I., Hoffman, T. A., Prágai, B. M., Dryga, S. A., Huang, H.
666 V., Schlesinger, S., and Rice, C. M. (1996) Alphavirus-based expression
vectors: Strategies and applications. *Proc. Natl. Acad. Sci. U.S.A.* 93,
11371–11377.
- (24) Petrakova, O., Volkova, E., Gorchakov, R., Paessler, S., Kinney,
R. M., and Frolov, I. (2005) Noncytopathic replication of Venezuelan
equine encephalitis virus and eastern equine encephalitis virus
replicons in mammalian cells. *J. Virol.* 79, 7597–7608.
- (25) Wiley, M. R., Roberts, L. O., Adelman, Z. N., Myles, K. M.
(2010) Double subgenomic alphaviruses expressing multiple fluo-
rescent proteins using a *Rhopalosiphum padi* virus internal ribosome
entry site element. *PLoS One* 5.
- (26) Sanz, M. Á., Castelló, A., Ventoso, I., Berlanga, J. J., and
Carrasco, L. (2009) Dual Mechanism for the Translation of
Subgenomic mRNA from Sindbis Virus in Infected and Uninfected
Cells. *PLoS One* 4, e4772.
- (27) Firth, A. E., and Brierley, I. (2012) Non-canonical translation in
RNA viruses. *J. Gen. Virol.* 93, 1385–1409.
- (28) Donnelly, M. L. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E.,
Gani, D., and Ryan, M. D. (2001) Analysis of the aphthovirus 2A/2B
polyprotein “cleavage” mechanism indicates not a proteolytic reaction,
but a novel translational effect: A putative ribosomal “skip. *J. Gen.*
Virol. 82, 1013–1025.
- (29) Thomas, J. M., Klimstra, W. B., Ryman, K. D., and Heidner, H.
W. (2003) Sindbis virus vectors designed to express a foreign protein
as a cleavable component of the viral structural polyprotein. *J. Virol.* 77,
5598–5606.
- (30) Kim, J. H., Lee, S.-R., Li, L.-H., Park, H.-J., Park, J.-H., Lee, K. Y.,
Kim, M.-K., Shin, B. A., Choi, S.-Y. (2011) High cleavage efficiency of
a 2A peptide derived from porcine teschovirus-1 in human cell lines,
zebrafish, and mice. *PLoS One* 6.
- (31) Frolov, I., and Schlesinger, S. (1994) Translation of Sindbis
virus mRNA: Effects of sequences downstream of the initiating codon.
J. Virol. 68, 8111–8117.
- (32) Caley, I. J., Betts, M. R., Davis, N. L., Swanstrom, R., Frelinger, J.
A., and Johnston, R. E. (1999) Venezuelan equine encephalitis virus
vectors expressing HIV-1 proteins: Vector design strategies for
improved vaccine efficacy. *Vaccine* 17, 3124–3135.
- (33) Kallio, K., Hellström, K., Balistreri, G., Spuul, P., Jokitalo, E., and
Ahola, T. (2013) Template RNA length determines the size of
replication complex spherules for Semliki Forest virus. *J. Virol.* 87,
9125–9134.
- (34) Decker, C. J., and Parker, R. (2012) P-bodies and stress
granules: Possible roles in the control of translation and mRNA
degradation. *Cold Spring Harb. Perspect. Biol.* 4, a012286.
- (35) Wengler, G., Koschinski, A., Wengler, G., and Dreyer, F. (2003)
Entry of alphaviruses at the plasma membrane converts the viral
surface proteins into an ion-permeable pore that can be detected by
electrophysiological analyses of whole-cell membrane currents. *J. Gen.*
Virol. 84, 173–181.
- (36) Paul, D., and Bartenschlager, R. (2013) Architecture and
biogenesis of plus-strand RNA virus replication factories. *World J.*
Virol. 2, 32–48.
- (37) Perri, S., Greer, C. E., Thudium, K., Doe, B., Legg, H., Liu, H.,
Romero, R. E., Tang, Z., Bin, Q., Dubensky, T. W., Vajdy, M., Otten,
G. R., and Polo, J. M. (2003) An alphavirus replicon particle chimera
derived from Venezuelan equine encephalitis and sindbis viruses is a
potent gene-based vaccine delivery vector. *J. Virol.* 77, 10394–10403.
- (38) Mahmoud, H. (2008) *Polya Urn Models*. CRC Press, Boca
Raton, FL.
- (39) Petrakova, O., Volkova, E., Gorchakov, R., Paessler, S., Kinney,
R. M., and Frolov, I. (2005) Noncytopathic replication of Venezuelan
equine encephalitis virus and eastern equine encephalitis virus
replicons in Mammalian cells. *J. Virol.* 79, 7597–7608.
- (40) Azizogolshani, O., Garmann, R. F., Cadena-Navar, R., Knobler, C.
M., and Gelbart, W. M. (2013) Reconstituted plant viral capsids can
release genes to mammalian cells. *Virology* 441, 12–17.
- (41) Beal, J.; Weiss, R.; Yaman, F.; Davidsohn, N.; Adler, A. A
Method for Fast, High-Precision Characterization of Synthetic Biology
Devices; Computer Science and Artificial Intelligence Laboratory
Technical Report; MIT: Cambridge, MA, 2012

736 (42) Perfetto, S. P., Ambrozak, D., Nguyen, R., Chattopadhyay, P.,
737 and Roederer, M. (2006) Quality assurance for polychromatic flow
738 cytometry. *Nat. Protoc.* 1, 1522–1530.