

Bayesian design of synthetic biological systems

Chris P. Barnes^{a,b,1}, Daniel Silk^{a,b}, Xia Sheng^{a,b}, and Michael P. H. Stumpf^{a,b,c,d,1}

^aCenter for Bioinformatics, Division of Molecular Biosciences; ^bInstitute of Mathematical Sciences; ^cCenter for Integrative Systems Biology; and ^dInstitute of Chemical Biology, Imperial College London, London SW7 2AZ, United Kingdom

Edited by Peter J. Bickel, University of California, Berkeley, CA, and approved July 27, 2011 (received for review December 1, 2010)

Here we introduce a new design framework for synthetic biology that exploits the advantages of Bayesian model selection. We will argue that the difference between inference and design is that in the former we try to reconstruct the system that has given rise to the data that we observe, whereas in the latter, we seek to construct the system that produces the data that we would like to observe, i.e., the desired behavior. Our approach allows us to exploit methods from Bayesian statistics, including efficient exploration of models spaces and high-dimensional parameter spaces, and the ability to rank models with respect to their ability to generate certain types of data. Bayesian model selection furthermore automatically strikes a balance between complexity and (predictive or explanatory) performance of mathematical models. To deal with the complexities of molecular systems we employ an approximate Bayesian computation scheme which only requires us to simulate from different competing models to arrive at rational criteria for choosing between them. We illustrate the advantages resulting from combining the design and modeling (or *in silico* prototyping) stages currently seen as separate in synthetic biology by reference to deterministic and stochastic model systems exhibiting adaptive and switch-like behavior, as well as bacterial two-component signaling systems.

biochemical circuits | dynamical systems | robustness

As we are beginning to understand the mechanisms governing biological systems we are starting to identify potential ways of guiding or controlling the behavior of cellular and molecular systems. Rationally reengineering organisms for biomedical or biotechnological purposes has become the central aim of the fledgling discipline of synthetic biology. By redirecting regulatory and physical interactions or by altering molecular binding affinities we may, for example, control metabolic processes (1, 2) or alter intra- and intercellular communication and decision making processes (3, 4). The range of potential applications of such engineered systems is vast: designing microbes for biofuel production (5, 6) and bioremediation (7); developing control strategies which drive stem cells through the various decisions to become terminally differentiated (or back) (8, 9), with the aim of developing novel therapeutics (10, 11); construction of new drug-delivery systems with homing microbes delivering molecular medicines directly to the site where they are needed (12); use of bacteria or bacterial populations (employing swarming and quorum sensing) as biosensors (13); and gaining better understanding of all manner of biological systems by systematically probing their underlying molecular machinery.

A range of tools and building blocks for such engineered biological systems are now available which allow us to, at least in principle, build such systems from simple and reusable biological components (14). In electronic systems, such modularity has been crucial and has allowed the cost-effective production of reliable components that can be combined to produce desired outputs. Biology, however, poses different and novel challenges that are intimately linked to the biophysical and biochemical properties of biomolecules and the media in which they are suspended. Especially in crowded environments such as found inside living cells the lack of insulation between different components, i.e., the very real possibility of undesired cross talk, can create problems;

with increasing miniaturization similar, albeit quantum effects, are now also surfacing in electronic circuits (15).

As synthetic biology gears up to bring engineering methods and tools to bear on biological problems the way in which we manipulate biological systems and processes is likely to change. Historically, each new branch of engineering has gone through a phase of what can be described as tinkering before rationally planned and executed designs became common place. Arguably, this practice is the current state of synthetic biology and it has indeed been suggested that the complexity of synthetic biological systems over the past decade has reached a plateau (16). From the earliest days, explicit quantitative modeling of systems has been integral to the vision and practice of synthetic biology and it will become increasingly important in the future. The ability to model how a natural or synthetic system will perform under controlled conditions must in fact be seen as one of the hallmarks of success of the integrative approaches taken by systems and synthetic biology.

Here we present a statistical approach to the design of synthetic biological systems that utilizes methods from Bayesian statistics to train models according to specified input-output characteristics. It incorporates modeling and automated design and is general in the sense that it can be applied to any system that can be described by a mathematical model which can be simulated. Because of the statistical nature of this approach, previously challenging problems such as handling stochastic models, accounting for kinetic parameter uncertainty, and incorporating environmental stochasticity can all be handled in a straightforward and consistent manner.

Bayesian Approach to System Design

The question of how to design a system to perform a specified task can be viewed as an analogue to reverse engineering. In design we want to elucidate the most appropriate system to achieve our design objectives; in reverse engineering we aim to infer the most probable system structure and dynamics that can give rise to some observational data. In this respect, the design question can be viewed as statistical inference on data we wish to observe.

In the Bayesian approach to statistical inference the posterior distribution is the quantity of interest and is given by the normalized product of the likelihood and the prior. In most practical applications the posterior distribution cannot be derived analytically, but if the likelihood (and prior) can be expressed mathematically we can use Monte Carlo methods to sample from the posterior. In many cases where the model structure is complex the likelihood cannot be written in closed form and traditional Monte Carlo techniques cannot be applied. These include inference for the types of stochastic processes encountered in systems and synthetic biology. In these cases a family of techniques known

Author contributions: C.P.B. and M.P.H.S. designed research; C.P.B., D.S., and X.S. performed research; and C.P.B. and M.P.H.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence may be addressed. E-mail: christopher.barnes@imperial.ac.uk or m.stumpf@imperial.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1017972108/-DCSupplemental.

collectively as approximate Bayesian computation (ABC) can be applied: ABC uses model simulations to approximate the posterior distribution directly. Here we use a sequential Monte Carlo ABC algorithm known as ABC SMC to move from the prior to the approximate posterior via a series of intermediate distributions (17). This framework can also be used to perform Bayesian model selection (18) and has been implemented in the software package ABC-SysBio (19).

Fig. 1 depicts the approach presented here. The design objectives are first specified through input-output characteristics. Here these have been depicted as a single time series, though the method can be applied in a much broader sense with multiple inputs and outputs. A set of competing designs is then specified through deterministic or stochastic models, each containing a set of kinetic parameters and associated prior distributions. The distance function measures the discrepancy between the model output and the objective. In principle it is possible to specify a distribution over the objective and each model could also contain experimental error. The ABC SMC algorithm then automatically evolves the set of models toward the desired design objectives. The results are a set of posterior probabilities representing the probability for each design to achieve the specified design objectives in addition to the posterior probability distribution of the associated kinetic parameters. This approach is similar in spirit to some existing methods for the automated design of genetic networks such as those adopting evolutionary algorithms (20, 21), Monte Carlo methods (22, 23), or optimization (24–26) but the advantages of our method over traditional ones are that we can utilize powerful concepts from Bayesian statistics in the design of complex biological systems, including

- the rational comparison of models under parameter uncertainty using Bayesian model selection, which automatically accounts for model complexity (number of parameters) and robustness to parameter uncertainty;

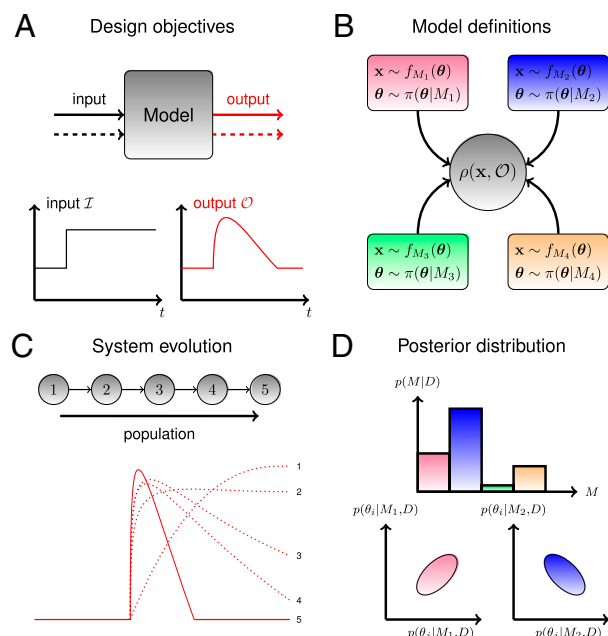


Fig. 1. Bayesian approach to system design. (A) The design objectives are encoded by the specification of input and output characteristics. (B) One or more competing designs for the system are specified together with priors on the parameters. A distance function, $\rho(\mathbf{x}, \mathcal{O})$, relates model output, \mathbf{x} , to the desired output characteristic, \mathcal{O} . (C) The system is evolved using sequential Monte Carlo. Each population more accurately approximates the desired behavior. (D) The model posterior probability encodes the ability of each design to achieve the desired behavior. The parameter posterior shows parameters that are sensitive or insensitive to the input-output specification.

- a posterior distribution over possible design parameter values that can be analyzed for parameter sensitivity and robustness and provide credible limits on design parameters;
- the treatment of stochastic systems at the design stage including the design of systems with required probability distributions on system components; and
- methods for the efficient exploration of high-dimensional parameter space.

In the following we demonstrate the power of this approach by examining, from this unique perspective, systems that have been of interest in the recent literature. First we consider systems that are capable of biochemical adaptation (27), we then look at the ability of two bacterial two-component system (TCS) topologies to achieve particular input-output behaviors, and finally we finish with an analysis of designs for a stochastic toggle switch with no cooperative binding at the promoter.

Biochemical Adaptation

Biochemical adaptation refers to the ability of a system to respond to an input signal and return to the prestimulus steady state (Fig. 14). Ma et al. (27) identified two three-node network topologies that are necessary for biochemical adaptation: a negative feedback loop with a buffering node and an incoherent feedforward loop with a proportioner node (IFFLP). Within these categories they identified eleven simple networks that were capable of adaptation (Fig. 24). We applied the Bayesian design approach to these eleven networks using Michaelis–Menten kinetic models with and without cooperativity [SI Appendix shows ordinary differential equations (ODEs) describing these models]. The desired output characteristics were defined through the adaptation efficiency, E , and sensitivity, S , given by

$$E = \left| \frac{(O_2 - O_1)/O_1}{(I_2 - I_1)/I_1} \right| \quad \text{and} \quad S = \left| \frac{(O_{\text{peak}} - O_1)/O_1}{(I_2 - I_1)/I_1} \right|,$$

where I_1, I_2 are the input values (here fixed at 0.5 and 0.6, respectively), O_1, O_2 are the output steady-state levels before and after the input change and O_{peak} is the maximal transient output level. We defined the two-component distance to be $\rho(\mathbf{x}, \mathcal{O}) = \{E, S^{-1}\}$ such that as $\rho(\mathbf{x}, \mathcal{O})$ decreases the behavior approaches the desired behavior. The final population was defined to obey the tolerances $\epsilon = \{0.1, 1.0\}$, which defines close to perfect adaptation (when $O_1 - O_2 \leq O_1/50$) and a fractional response equal to the fractional change in input.

The results of the model selection are shown in Fig. 2 B and C. When cooperativity is not included the most robust designs for producing the desired input-output characteristics are the incoherent feedforward loops, but when cooperativity is added the posterior shifts significantly toward the negative feedback topology. If a system with these requirements were to be implemented then not only would designs 11 and 4 be clear candidates for further study, but many of the designs can be effectively ruled out and the ranking of the models provides a clear strategy for an experimental program. These results also illustrate how small changes in context or incomplete understanding of a system can produce a large change in the most robust design. The Bayesian framework allows us to incorporate such uncertainty—or safeguard against our ignorance—naturally into the design process.

The posterior distribution provides information on which parameters are correlated and which are the most sensitive to the desired behavior. The posterior for model 11 under no cooperativity is shown in Fig. 2D, where the ODE model is given by

Fig. 2. Biochemical adaptation. (A) Eleven networks capable of biochemical adaptation. A, B, C represent enzymes that catalyze reactions in their active state. For example $A \rightarrow B$ indicates that A converts B from its inactive to active state and $A \leftarrow B$ indicates that A converts B from its active to inactive state. The input is applied to species A and the output is taken to be the concentration of the active form of C . The concentrations of active and inactive forms sum to 1. Reactions with no origin node refer to background activating/deactivations enzymes. (B) Posterior probability for achieving biochemical adaptation when there is no cooperativity. (C) Posterior probability for achieving biochemical adaptation when cooperativity is included. The error bars in B and C indicate the variability in the marginal model posteriors over three separate runs. (D) Parameter posterior distribution, represented by univariate and bivariate marginal distributions, for model 11 in the case of no cooperativity.

$$\begin{aligned}\frac{dA}{dt} &= Ik_{IA} \frac{(1-A)^{n_{IA}}}{(1-A)^{n_{IA}} + K_{IA}^{n_{IA}}} - F_A k_{FA} \frac{A^{n_{FA}}}{A^{n_{FA}} + K_{FA}^{n_{FA}}} \\ \frac{dB}{dt} &= Ak_{AB} \frac{(1-B)^{n_{AB}}}{(1-B)^{n_{AB}} + K_{AB}^{n_{AB}}} - F_B k_{FB} \frac{B^{n_{FB}}}{B^{n_{FB}} + K_{FB}^{n_{FB}}} \\ \frac{dC}{dt} &= Ak_{AC} \frac{(1-C)^{n_{AC}}}{(1-C)^{n_{AC}} + K_{AC}^{n_{AC}}} - Bk_{BC} \frac{C^{n_{BC}}}{C^{n_{BC}} + K_{BC}^{n_{BC}}},\end{aligned}$$

where $X = \{A, B, C\}$ corresponds to the concentrations of the active forms of proteins (1 - X corresponds to the concentrations of the inactive form). I represents the input signal and k and K represent the reaction rate parameters (of which there are 12 in total). F_A and F_B represent background deactivating enzymes with concentration fixed at 0.5.

The posterior shows in particular that the parameters for the background deactivating enzyme (k_{FB} and K_{FB}) on node B are correlated, and that K_{FB} should be large; this criterion is exactly the requirement for the linear regime necessary for the IFFLP system to achieve adaptation (27). A principal component (PC) analysis of the posterior (*SI Appendix*, Fig. S1) shows other correlated parameters on the first few principal components. The last principal component describes the direction of least variance and therefore the most sensitive parameters. From this information we can deduce for example that the nature of the reaction between nodes A and C is relatively unimportant. A similar analysis for model 4 in the case of cooperativity (*SI Appendix*, Figs. S2 and S3) shows for example that the behavior is insensitive to the values of the Hill coefficients and the details of the reaction between nodes A and C.

Bacterial Two-Component Systems

TCSs allow bacteria to sense external environmental stimuli and relay information into the cell, e.g., to the gene expression apparatus. They consist of a histidine kinase (HK) that autophosphorylates upon interaction with a specific stimulus. The phosphoryl group is then passed on to a cognate response regulator

protein (RR) which, once phosphorylated, can regulate transcription (28).

Naturally occurring TCSs differ in the number of phosphate binding domains. In the most common form there are two phosphate binding domains (Fig. 3A), but an alternative form exists that consists of a phosphorelay mechanism with four phosphate binding domains (Fig. 3B). These are referred to as the orthodox and unorthodox TCSs, respectively (29). The reason for the existence of two forms of TCS remains largely unknown but it has been demonstrated that the phosphorelay is robust to noise and can provide an ultrasensitive response to stimuli (29, 30), whereas the orthodox system can provide behavior that is independent of the concentrations of its components (31). Here we have applied the Bayesian approach to directly compare the ability of orthodox and unorthodox designs to achieve various input-output behaviors, using ODE models similar to ones described previously (29) (see *SI Appendix*).

Fig. 3 *C–F* show four types of behavior that may be desired in synthetic TCS systems (e.g., for bioremediation or biopharmaceutical applications), and the corresponding posterior probabilities of the orthodox and unorthodox models to achieve them. In Fig. 3*C* the specified behavior is that of a fast response to a square pulse input signal. That is the output should show a maximum within 0.1 s after the pulse starts and a minimum 0.1 s after the pulse ends. As can be seen from the posterior probabilities, both models achieve this behavior easily, as one would expect from a signaling system, with the orthodox system slightly outperforming the unorthodox system. In Fig. 3*D* the ability of the two systems to achieve a steady output state for $t > 2$ s under a constant input signal is examined, and again both systems perform comparably with the orthodox system appearing slightly more favorable.

In Fig. 3E the input signal is a high-frequency sinusoid with a mean of 0.6, and the desired output is a constant signal with the same mean and root mean square < 0.3 ; this behavior would mimic a system that is robust to high-frequency noise. The output trajectories at some intermediate and final populations are shown

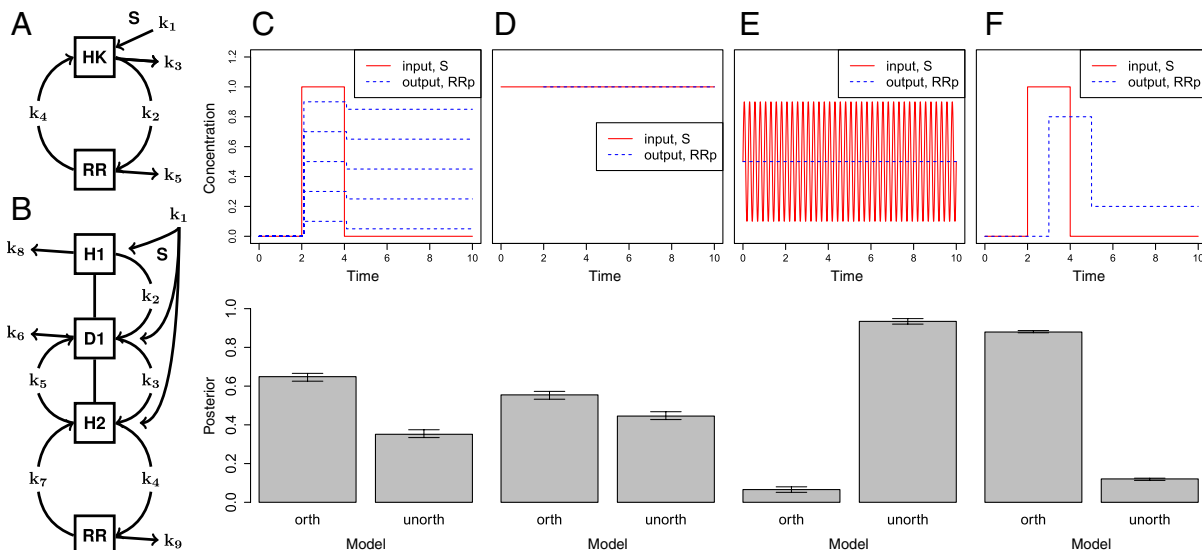


Fig. 3. Bacterial two-component systems. (A) Orthodox system where *HK* denotes the histidine kinase and *RR* the response regulator, both of which have phosphorylated forms, *HKp* and *RRp*. Arrows represent reactions involving phosphate groups and the k_i represent the rate parameters. S is the input stimulus signal that causes autophosphorylation of the histidine kinase. (B) Phosphorelay system with three phosphate-binding domains, where *H*, *D* refer to histidine and aspartate domains, respectively. (C–F) Specified input-output behavior (above) and posterior probabilities for the two designs to achieve it (below). The input signal corresponds to the stimulus, S , and the output signal is represented by the concentration of phosphorylated response regulator, *RRp*. The error bars indicate the variability in the marginal model posteriors over three separate runs.

in *SI Appendix*, Fig. S4). In this example the unorthodox system clearly outperforms the orthodox system, which indicates the increased robust to noisy signals that comes with the relay architecture (29). But the direct comparison of the two models' ability to cope with noise, which is becoming possible in this approach, also reveals some unexpected characteristics: Inspection of the posterior distribution (*SI Appendix*, Fig. S5) shows that all the dephosphorylation reaction rates, k_6, k_8, k_9 , are minimized whereas the rate of the signal induced autophosphorylation (k_1) is large. Thus the noise reduction mechanism in the unorthodox system works by saturating the system.

In Fig. 3F the input is again a step function but the output is more specific; it must reach to >0.8 and drop to <0.2 within 0.5 s of the pulse start and end, respectively, thus approximately reproducing the input. [SI Appendix](#), Fig. S6 shows the evolution of the system in this case. Here the orthodox model clearly outperforms the unorthodox model. Inspection of the posterior distribution ([SI Appendix](#), Fig. S7) shows that both the rate of the signal induced autophosphorylation and the rate of phosphorylation of the response regulator by the histidine kinase, k_1, k_2 , are large whereas the rate of dephosphorylation of the response regulator, k_5 , is small. These properties ensure that the shape of the signal is transferred faithfully through the system.

Stochastic Genetic Toggle Switch Without Cooperativity

The genetic toggle switch is a synthetic realization of a bistable switch that forms the basis of cellular memory (32). It is formed by two genes A and B , whose respective proteins repress the production of the other protein; protein A represses the production of protein B and vice versa (Fig. 4). The presence of an interaction with inducer molecules allows the system to switch between steady states with the probability of spontaneous switching low enough such that, in the absence of an interaction, the system will effectively remain in its appropriate steady state indefinitely.

Here we consider four versions of the stochastic genetic toggle switch that are all bistable without the requirement for cooperative binding of the proteins to the gene promoter (33) (note that the deterministic models are not necessarily bistable). The designs are shown in Fig. 4A and consist of the basic toggle switch, an exclusive version containing only one promoter, a version that includes bound repressor degradation (BRD), and a version con-

taining a protein-protein interaction between A and B with the resulting complex nonfunctional. The additional reactions are always to reduce the probability of the deadlock state where both A and B are bound to the promoters of B and A , respectively (33). We modeled the switches using a continuous time Markov jump process which obeys the chemical master equation. Only protein level reactions were modeled which makes the models simpler (and faster to simulate) whereas retaining the important behavior. The stochastic models for all four switches are given in *SI Appendix*. For such complicated stochastic dynamical systems the advantages of a Bayesian perspective over conventional model design strategies (based, e.g., on optimization) come to the fore: Without an appreciation of the whole distribution choice of the best model would be subject to considerable uncertainty.

Fig. 4B shows the desired toggle switch behavior. The inducer S is added between $t = 10$ and $t = 30$, after which the level of protein B should reach a steady state with a mean number of 40. Between $t = 60$ and $t = 80$ the inducer R is added after which the level of protein B should drop to zero. The inducer numbers are both assumed to be 40 molecules which is fixed in the specified range. The desired output behavior was specified via the two-component distance metric, defined to be $\rho(\mathbf{x}, \mathcal{O}) = \{d_1, d_2\}$,

$$d_1 = \sqrt{\frac{\sum_{t \in \alpha} (x_t - y)^2}{n_\alpha}} \quad d_2 = \sqrt{\frac{\sum_{t \in \beta} (x_t - 0)^2}{n_\beta}},$$

where x_t is the number of protein B at time t , y is the target (here fixed at 40), $\alpha = \{t: 30 < t \leq 60\}$, $\beta = \{t: 0 < t \leq 10 \text{ and } 80 < t \leq 100\}$, $n_\alpha = \#\alpha$ and $n_\beta = \#\beta$. The final population was defined to be $\epsilon = \{7.0, 0.05\}$. Here d_1 represents the distance in the on region and d_2 represents the distance in the off region.

SI Appendix, Fig. S8 shows the evolution of the stochastic simulations toward the desired behavior. Fig. 4C shows the posterior probabilities for each system to achieve the specified behavior, and in particular demonstrates that the exclusive toggle switch outperforms all the others, which chimes with intuition, because the exclusive switch removes the possibility of the dead-lock state without the addition of any extra reactions. The fact that the BRD switch performs worse than the original toggle

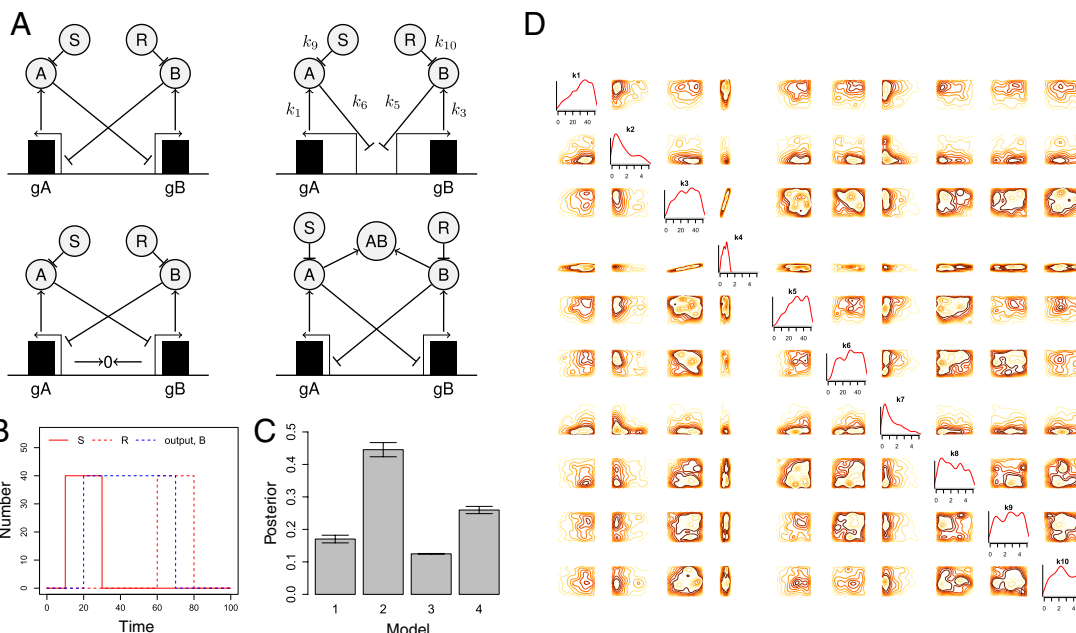
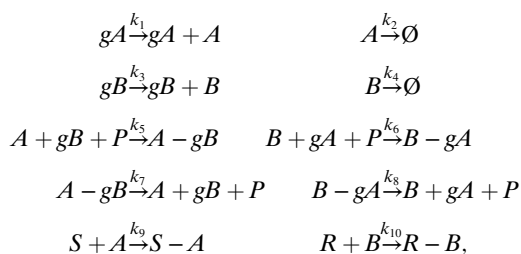


Fig. 4. Stochastic toggle switch. (A) Four different designs for a toggle switch without cooperative binding. Going clockwise from top left we have the basic switch, the exclusive switch where there is only one repressor site, the basic switch with bound repressor degradation, and the basic switch with a protein-protein interaction. Genes *gA*, *gB* express proteins *A*, *B* and *S*, *R* represent inducer molecules. The species *AB* represents complex formation. The rate constants for the reactions are shown for the exclusive switch (protein degradation and transcription factor dissociation from the promoter are not shown). (B) The specified input-output behavior. (C) The posterior probabilities for each model to achieve the toggle switch behavior. (D) Parameter posterior distribution, represented by univariate and bivariate marginal distributions, for model 2 (exclusive switch).

shows that the addition of the two extra degradation reactions does not offer a great enough performance increase for the extra parameters, a manifestation of the parsimony principle or Occam’s razor which is inherent to the Bayesian model selection framework used here.

The reactions that comprise the exclusive switch are given by



where gA, gB represent the gene promoters for protein A, B , respectively, and are fixed at one copy, $A - gB, B - gA$ represent the bound transcription factors and S, R represent the inducer molecules. The P species, fixed to be one copy, ensures only A or B can be bound at any one time. Examination of the posterior distribution for this model, Fig. 4D, clearly shows a large correlation between k_3 and k_4 , which are the production and decay rates of protein B , respectively. This correlation is clearly seen in the principal components (SI Appendix, Fig. S9) through the combination $k_3 + k_4$ dominating the first PC and the combination $k_4 - k_3$ dominating the last PC. Thus the system is sensitive to only the difference in these rates which is typical in birth-death processes.

Discussion

In this paper we have presented a unique method for the design of synthetic biological systems employing ideas from Bayesian statistics. We have demonstrated its utility and generality on three different systems spanning biochemical, signaling, and genetic networks; we also examined oscillatory behavior which is presented in *SI Appendix*. This method has advantages over tra-

ditional design approaches in that the modeling and model evaluation/characterization is incorporated directly into the design stage. The statistical nature of the method has many attractive features including the handling of stochastic systems, the ability to perform model selection and the handling of parameter uncertainty in a well-defined manner. Model selection using summary statistics can be problematic in other settings (34) but is not a concern here because we either use the full dataset with no summary or we define our posterior distributions through the desired system outputs. We used the ABC-SysBio and cuda-sim softwares (19, 35), which takes as input a set of SBML files and as such can be used by bioengineers and experimentalists to rationally compare their competing designs for a system. By using this method we hope that the implementation time of synthetic systems can be reduced by defining a program of experimental work based on the posterior probabilities of each design.

Monte Carlo sampling of parameter spaces has been used to assess the robustness of engineered and biological systems in the past. But like in the statistical case, simple Monte Carlo sampling tends to waste too much effort and time on those regions which are of no real interest for reverse engineering or design purposes. Our statistically based sequential approach homes in onto those regions where the probability of observing the desired behavior is appreciable, which allows us a more nuanced comparative assessment of different design proposals, especially when dynamics are expected (or indeed desired) to exhibit elements of stochasticity. And the Bayesian model selection approach automatically strikes a balance between the systems' abilities to generate the desired behavior effectively but also robustly.

Further developments will include the incorporation of methods for model abstraction to reduce computation time (36) and to handle a database of standard parts as in other existing design software systems (37, 38). Moreover, it is also possible to include the generation of novel structures (by, e.g., using stochastic context-free grammars (39) to propose alterations to a reaction/interaction network) as part of the design process. Just like in the case where ideas from control engineering and statistics can gain-

fully be combined to reverse-engineer the structure of naturally evolved biological systems, we feel that in the design of synthetic systems such a union will also be fruitful.

1. Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat Biotechnol* 21:796–802.
2. Ro DK, et al. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440:940–943.
3. You L, Cox RS, Weiss R, Arnold FH (2004) Programmed population control by cell-cell communication and regulated killing. *Nature* 428:868–871.
4. Kobayashi H, et al. (2004) Programmable cells: Interfacing natural and engineered gene networks. *Proc Natl Acad Sci USA* 101:8414–8419.
5. Fortman JL, et al. (2008) Biofuel alternatives to ethanol: Pumping the microbial well. *Trends Biotechnol* 26:375–381.
6. Savage DF, Way J, Silver PA (2008) Defossilizing fuel: How synthetic biology can transform biofuel production. *ACS Chem Biol* 3:13–16.
7. Cases I, de Lorenzo V (2005) Genetically modified organisms for the environment: Stories of success and failure and what we have learned from them. *Int Microbiol* 8:213–222.
8. Takahashi K, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131:861–872.
9. Hanna JH, Saha K, Jaenisch R (2010) Pluripotency and cellular reprogramming: Facts, hypotheses, unresolved issues. *Cell* 143:508–525.
10. Lu TK, Khalil AS, Collins JJ (2009) Next-generation synthetic gene networks. *Nat Biotechnol* 27:1139–1150.
11. MacArthur BD, Ma'ayan A, Lemischka IR (2009) Systems biology of stem cell fate and cellular reprogramming. *Nat Rev Mol Cell Biol* 10:672–681.
12. Anderson JC, Clarke EJ, Arkin AP, Voigt CA (2006) Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol* 355:619–627.
13. Rajendran M, Ellington AD (2008) Selection of fluorescent aptamer beacons that light up in the presence of zinc. *Anal Bioanal Chem* 390:1067–1075.
14. Canton B, Labno A, Endy D (2008) Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol* 26:787–793.
15. Liang W, Shores MP, Bockrath M, Long JR, Park H (2002) Kondo resonance in a single-molecule transistor. *Nature* 417:725–729.
16. Purnick PEM, Weiss R (2009) The second wave of synthetic biology: From modules to systems. *Nat Rev Mol Cell Biol* 10:410–422.
17. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface* 6:187–202.
18. Toni T, Stumpf MPH (2010) Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26:104–110.
19. Liepe J, et al. (2010) ABC-SysBio-approximate Bayesian computation in Python with GPU support. *Bioinformatics* 26:1797–1799.

ACKNOWLEDGMENTS. C.B., D.S. and M.P.H.S gratefully acknowledge support from the Biotechnology and Biological Research Council. X.S. is supported through a Kwok Foundation scholarship. M.P.H.S. is a Royal Society Wolfson Research Merit Award holder.

20. Bray D, Lay S (1994) Computer simulated evolution of a network of cell-signaling molecules. *Biophys J* 66:972–977.
21. François P, Hakim V (2004) Design of genetic networks with specified functions by evolution *in silico*. *Proc Natl Acad Sci USA* 101:580–585.
22. Battogtokh D, Asch DK, Case ME, Arnold J, Schuttler HB (2002) An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of *Neurospora crassa*. *Proc Natl Acad Sci USA* 99:16904–16909.
23. Feng XJ, et al. (2004) Optimizing genetic circuits by global sensitivity analysis. *Biophys J* 87:2195–2202.
24. Rodrigo G, Carrera J, Jaramillo A (2007) Genetdes: Automatic design of transcriptional networks. *Bioinformatics* 23:1857–1858.
25. Dasika MS, Maranas CD (2008) Optcircuit: An optimization based method for computational design of genetic circuits. *BMC Syst Biol* 2:24.
26. Batt G, Yordanov B, Weiss R, Belta C (2007) Robustness analysis and tuning of synthetic gene networks. *Bioinformatics* 23:2415–2422.
27. Ma W, Trusina A, El-Samad H, Lim WA, Tang C (2009) Defining network topologies that can achieve biochemical adaptation. *Cell* 138:760–773.
28. Stock AM, Robinson VL, Goudreau PN (2000) Two-component signal transduction. *Annu Rev Biochem* 69:183–215.
29. Kim JR, Cho KH (2006) The multi-step phosphorelay mechanism of unorthodox two-component systems in *E. coli* realizes ultrasensitivity to stimuli while maintaining robustness to noises. *Comput Biol Chem* 30:438–444.
30. Csikász-Nagy A, Cardelli L, Soyer OS (2011) Response dynamics of phosphorelays suggest their potential utility in cell signalling. *J R Soc Interface* 8:480–488.
31. Shinar G, Milo R, Martinez MR, Alon U (2007) Input output robustness in simple bacterial signaling systems. *Proc Natl Acad Sci USA* 104:19931–19935.
32. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403:339–342.
33. Lipshtat A, Loinger A, Balaban NQ, Biham O (2006) Genetic toggle switch without cooperative binding. *Phys Rev Lett* 96:188101.
34. Robert CP, Cornuet JM, Marin JM, Pillai N (2011) Lack of confidence in ABC model choice. [arXiv:1102.4432v4](https://arxiv.org/abs/1102.4432v4).
35. Zhou Y, et al. (2011) GPU accelerated biochemical network simulation. *Bioinformatics* 27:874–876.
36. Myers CJ, et al. (2009) Ibiosim: A tool for the analysis and design of genetic circuits. *Bioinformatics* 25:2848–2849.
37. Hill AD, Tomshine JR, Weeding EMB, Sotiropoulos V, Kaznessis YN (2008) Synbioss: The synthetic biology modeling suite. *Bioinformatics* 24:2551–2553.
38. Marchisio MA, Stelling J (2008) Computational design of synthetic gene circuits with composable parts. *Bioinformatics* 24:1903–1910.
39. Baldi P, Brunak Sa (2001) *Bioinformatics: The Machine Learning Approach* (MIT Press, Cambridge, MA), 2nd Ed.