Bowyer *et al. Journal of Biological Engineering* (2017) 11:30
DOI 10.1186/s13036-017-0071-6

Journal of Biological Engineering

RESEARCH Open Access

Modeling the architecture of the regulatory system controlling methylenomycin production in *Streptomyces coelicolor*

Jack E. Bowyer[1*], Emmanuel LC. de los Santos[2], Kathryn M. Styles[3], Alex Fullwood[3], Christophe Corre[4] and Declan G. Bates[1]

Abstract

Background: The antibiotic methylenomycin A is produced naturally by *Streptomyces coelicolor* A3(2), a model organism for streptomycetes. This compound is of particular interest to synthetic biologists because all of the associated biosynthetic, regulatory and resistance genes are located on a single cluster on the SCP1 plasmid, making the entire module easily transferable between different bacterial strains. Understanding further the regulation and biosynthesis of the methylenomycin producing gene cluster could assist in the identification of motifs that can be exploited in synthetic regulatory systems for the rational engineering of novel natural products and antibiotics.

Results: We identify and validate a plausible architecture for the regulatory system controlling methylenomycin production in *S. coelicolor* using mathematical modeling approaches. Model selection via an approximate Bayesian computation (ABC) approach identifies three candidate model architectures that are most likely to produce the available experimental data, from a set of 48 possible candidates. Subsequent global optimization of the parameters of these model architectures identifies a single model that most accurately reproduces the dynamical response of the system, as captured by time series data on methylenomycin production. Further analyses of variants of this model architecture that capture the effects of gene knockouts also reproduce qualitative experimental results observed in mutant *S. coelicolor* strains.

Conclusions: The mechanistic mathematical model developed in this study recapitulates current biological knowledge of the regulation and biosynthesis of the methylenomycin producing gene cluster, and can be used in future studies to make testable predictions and formulate experiments to further improve our understanding of this complex regulatory system.

Keywords: Synthetic biology, Antibiotics, Gene regulation, Methylenomycin, Mathematical modelling, Approximate Bayesian computation, Global optimization

Background

There is currently an increasing demand for research and development of new antibiotics as their overuse, along with many other factors, has led to increased resistance. Streptomycetes produce approximately 70% of all commercial antibiotics currently available [1]. The bacterium *Streptomyces coelicolor* A3(2) has emerged as the model organism for studying streptomycetes, initially thanks to the production of colored metabolites that facilitated genetic studies, and more recently thanks to the sequencing of its entire genome [2]. These bacteria have a 8,667,507 base pair single linear chromosome containing protein coding genes of which over 12% are thought to be regulatory [2]. These predicted transcriptional regulators are thought to mediate antibiotic synthesis through the production of microbial hormones, as well as influence structural and metabolic cellular responses [3]. The linear SCP1 plasmid (~ 356 kb) and the circular SCP2 plasmid (~ 31 kb) are both present within the *S. coelicolor* genome and have also both been sequenced [4]. This genome sequencing has revealed many cryptic and 'silent' gene clusters: sets of genes predicted to

*Correspondence: J.Bowyer@warwick.ac.uk
[1]Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry CV4 7AL, UK
Full list of author information is available at the end of the article

produce a natural product, but whose product has not been observed. Silent gene clusters have been awakened through genetic manipulation of regulatory elements [5, 6]. Thus, characterization of the regulatory system that mediates the production of specialized metabolites is key to discovering new natural products. Developing improved understanding of the regulatory architectures that underlie natural product biosynthesis can also accelerate the design of novel regulatory systems in synthetic biology.

The antibiotic methylenomycin A is a natural product of *S. coelicolor* A3(2) and is of particular interest since all of the 21 biosynthetic, regulatory and resistance genes, located in a cluster on the SCP1 plasmid [4, 7], have been studied in detail [8], and a series of knockout mutant strains has been generated [9]. The regulation of methylenomycin biosynthesis is mediated by the transcriptional repressor MmfR, a TetR-family homodimeric protein consisting of an N-terminal DNA-binding domain and a C-terminal ligand-binding domain (Fig. 1a) [10, 11]. In the initial growth phase of *S. coelicolor*, the MmfR N-terminal domain is thought to be bound to the DNA at the methylenomycin auto-regulatory response element (MARE) causing the transcriptional repression of downstream genes. MmfR holds the system in this repressed state until the advent of the small signaling molecules, methylenomycin furans (MMFs) [12]. MMFs bind specifically to the C-terminal domain of the MmfR, forming an MmfR:MMF complex that results in a conformational change in the MmfR. Consequently, MmfR is released from the MARE, negating the repression and triggering gene transcription. The biosynthesis of MMFs is controlled by the MmfLHP enzymes which are, themselves, repressed by MmfR, thus forming a feedback control loop that governs the dynamical properties of the system. A second repressor, MmyR, is homologous to MmfR yet its role in methylenomycin regulation is currently less understood. There is, however, clear evidence that the impact of MmyR is particularly significant, since *S. coelicolor* strains with the *mmyR* gene knocked out have been found to overproduce methylenomycin [9].

Homologous architectures to that of the methylenomycin regulatory system have been identified across a plethora of microorganisms [13], regulating different classes of natural products and thus indicating the utility of this specific type of regulatory architecture [12]. Responding to environmental changes is of paramount importance to these bacteria. The soil they live in presents a harsh environment with considerable competition for resources and it is therefore vital that they possess sophisticated, tightly regulated mechanisms to turn on the expression of specific genes when required. Hence, obtaining a detailed mechanistic understanding of the regulatory system controlling the biosynthetic pathway to this antibiotic has the potential to elucidate a host of other, less tractable, biosynthetic gene clusters and help standardize one of the most important regulatory networks for the development of new antibiotics.

Recent mathematical modeling investigations have generated new insights into the operation of numerous systems of interest to synthetic biologists [14–17]. Such models not only provide the capability to accurately simulate synthetic systems during the design and development phase, but are also effective tools for the prediction of system responses to variations in environmental conditions [18]. Here, we develop the first detailed

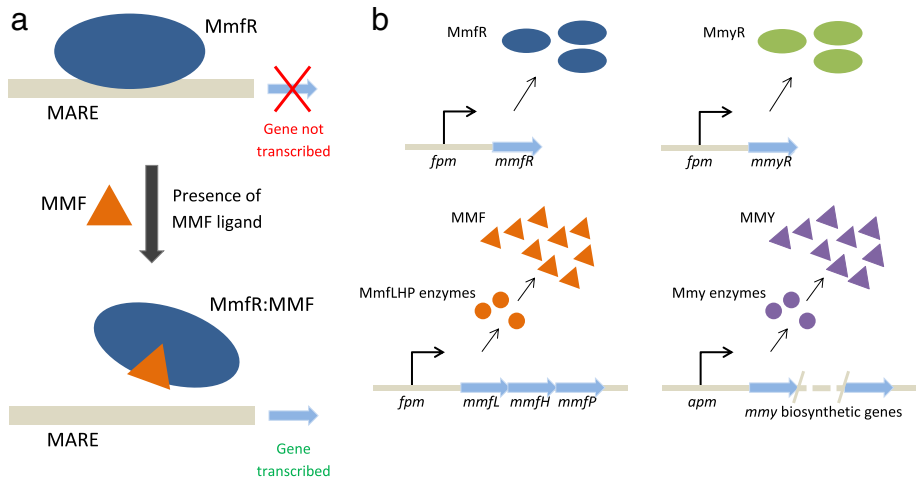

Fig. 1 a Schematic diagram of the MmfR binding mechanism. Binding of MmfR to DNA at the MARE represses gene transcription and therefore negates system output. In the presence of MMF ligand, an MmfR:MMF complex is formed which releases MmfR from the MARE and triggers gene transcription. **b** Schematic diagram of the methylenomycin gene cluster whereby *fpm* and *apm* represent the DNA binding motifs recognized by MmfR and MmyR proteins. The *fpm* controls the expression of *mmfR*, *mmyR* and *mmfLHP* genes while *apm* regulates the expression of the *mmy* biosynthetic genes

mathematical model of the MMF-dependent regulatory system involved in methylenomycin production in *S. coelicolor*, firstly, through a rigorous statistical analysis of the plausibility of multiple candidate model architectures and, secondly, via global optimization of the relevant model parameters against available experimental data. We also validate our candidate model architecture against a range of selection criteria devised in light of experimental observations on methylenomycin production in several mutant *S. coelicolor* strains.

Results and discussion

Formulation of candidate model architectures

The various binding interactions and protein expression summarized in Fig. 1 inform the formulation of our candidate model architectures. MmfR is thought to bind to three distinct intergenic regions on the gene cluster [9]. However, we combine the region associated with MmyR biosynthesis together with the region associated with both MmfR and MMF biosynthesis to form a single DNA module responsible for the biosynthesis of all three molecules (the furan producing module, *fpm*). That is, we use the term *fpm* to refer to five distinct genes that provide control over three distinct molecular products: MmyR, MmfR and MMF. The genes *mmfL*, *mmfH* and *mmfP* are coregulated in an operon and are directly responsible for the production (assembly) of MMF molecules; the *mmfR* and *mmyR* genes control MmfR and MmyR production respectively [9, 12] (Fig. 1b). The third distinct intergenic region is represented by our second DNA module which we consider responsible for methylenomycin (MMY) biosynthesis only (the antibiotic producing module, *apm*). Therefore, our model architectures all consist of two fundamental DNA modules that can both be bound by MmfR, and that have production of their respective proteins repressed as a consequence. Due to its effect on the gene cluster and its homology to MmfR, in this study we hypothesize that MmyR also binds both modules in a similar manner.

Our base architecture accounts for reversible MmfR and MmyR binding to both the *fpm* and *apm* to form four complexes: *fpm*:MmfR, *fpm*:MmyR, *apm*:MmfR and *apm*:MmyR. MMF binds MmfR reversibly at these complexes in order to trigger gene expression; MMF binding MmfR in solution is also accounted for since we have been able to co-crystallize MmfR:MMF complexes and solve the 3D-structure through experimentation void of target DNA modules (data not shown). MmfR:MMF complexes that dissociate from the MAREs return free MmfR and MMF back into the system irreversibly. MmyR, MmfR and MMF production is controlled by the *fpm*. We account for an initial repressed system state by imposing non-zero initial concentrations upon the *fpm*:MmfR and *apm*:MmfR complexes; all remaining model variables have initial concentrations equal to zero. MmfR, MmyR, MMF and MMY all undergo degradation at constant rates (Fig. 2).

This model architecture represents the extent of our current mechanistic understanding, however there are certain details that require further investigation. For example, although we believe that the MMF releases MmfR from the *fpm*:MmfR and *apm*:MmfR complexes and also binds free MmfR in solution, it would be insightful to examine the dynamical influence of each binding mechanism in isolation. Similarly, although we believe there is no interaction between the MMF and MmyR within the system (data not shown), the binding interactions of MMY are not as well documented. It may therefore be possible that MMY is able to inhibit the action of both MmfR and MmyR either through dissociation from their respective *fpm* and *apm* complexes or binding in solution. Consequently, the aim of our modeling investigation is to examine the effect of three key mechanistic properties on model performance:

1. MMF-MmfR interactions occur at existing DNA:MmfR complexes (C), in solution (S) or via both mechanisms (B).
2. MMY-MmfR interactions occur at existing DNA:MmfR complexes (C), in solution (S), via both mechanisms (B) or do not occur at all (N).
3. MMY-MmyR interactions occur at existing DNA:MmyR complexes (C), in solution (S), via both mechanisms (B) or do not occur at all (N).

This set of possible molecular interactions results in 48 distinct candidate model architectures for the methylenomycin regulatory system. Each candidate architecture is given a three letter name corresponding to the interactions accounted for with respect to each of the three properties listed above. The order of the letters in each name corresponds strictly to the numerical order of these properties. For example, our base architecture (described above) is given the name BNN since it accounts for both mechanisms (B) regarding property 1, for no interactions at all (N) regarding property 2 and for no interactions at all (N) regarding property 3.

Each of the 48 candidate architectures presents a distinct reaction network and set of biochemical equations that can be used to derive a dynamical mathematical model. We apply mass action kinetics to the biochemical equations comprising each reaction network to derive a corresponding system of ordinary differential equations (ODEs). Each ODE describes the rate of change in concentration corresponding to each model variable (cellular entity). The solution to each system of ODEs is determined numerically due to the non-linearity of the equations and provides a deterministic output that can be used to simulate and predict in vivo system dynamics

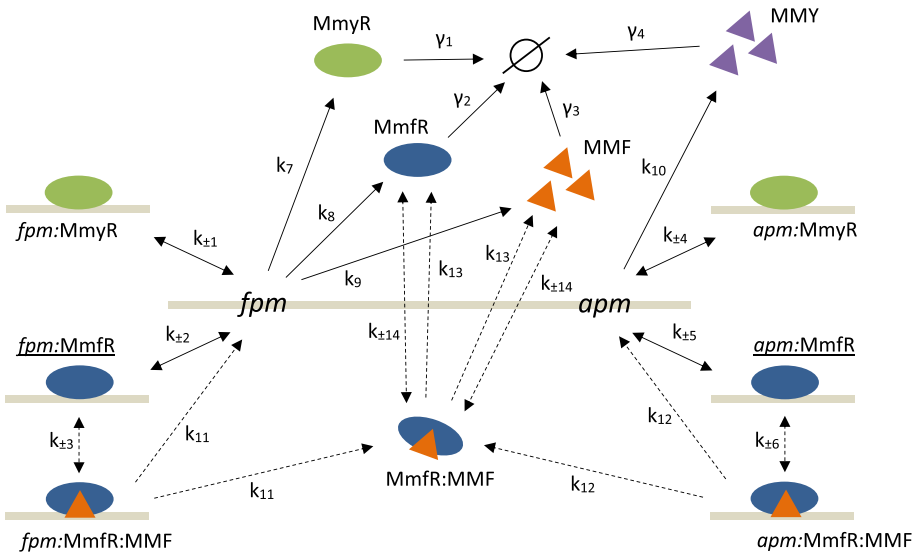

Fig. 2 Schematic diagram of the reaction network comprising the base (BNN) model architecture. Reversible and irreversible reactions are depicted by double and single *arrows* respectively; reaction rate constants are denoted by the corresponding numbered k. The empty set depicts protein degradation, with rate constants denoted by the corresponding numbered γ. *Solid arrows* depict reactions that are common to all 48 model architectures, whereas *dashed arrows* depict those that are subject to adaptation. Cellular entities with non-zero initial concentrations are underlined

in silico. For example, the BNN model architecture is comprised of the following biochemical equations:

$$fpm + \text{MmyR} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} fpm\text{:MmyR} \quad (1)$$

$$fpm + \text{MmfR} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} fpm\text{:MmfR} \quad (2)$$

$$fpm \xrightarrow{k_7} \text{MmyR} + fpm \quad (3)$$

$$fpm \xrightarrow{k_8} \text{MmfR} + fpm \quad (4)$$

$$fpm \xrightarrow{k_9} \text{MMF} + fpm \quad (5)$$

$$fpm\text{:MmfR} + \text{MMF} \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} fpm\text{:MmfR:MMF} \quad (6)$$

$$fpm\text{:MmfR:MMF} \xrightarrow{k_{11}} fpm + \text{MmfR:MMF} \quad (7)$$

$$apm + \text{MmyR} \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} apm\text{:MmyR} \quad (8)$$

$$apm + \text{MmfR} \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} apm\text{:MmfR} \quad (9)$$

$$apm \xrightarrow{k_{10}} \text{MMY} + apm \quad (10)$$

$$apm\text{:MmfR} + \text{MMF} \underset{k_{-6}}{\overset{k_6}{\rightleftharpoons}} apm\text{:MmfR:MMF} \quad (11)$$

$$apm\text{:MmfR:MMF} \xrightarrow{k_{12}} apm + \text{MmfR:MMF} \quad (12)$$

$$\text{MmfR:MMF} \xrightarrow{k_{13}} \text{MmfR} + \text{MMF} \quad (13)$$

$$\text{MmfR} + \text{MMF} \underset{k_{-14}}{\overset{k_{14}}{\rightleftharpoons}} \text{MmfR:MMF} \quad (14)$$

$$\text{MmyR} \xrightarrow{\gamma_1} \varnothing \quad (15)$$

$$\text{MmfR} \xrightarrow{\gamma_2} \varnothing \quad (16)$$

$$\text{MMF} \xrightarrow{\gamma_3} \varnothing \quad (17)$$

$$\text{MMY} \xrightarrow{\gamma_4} \varnothing \quad (18)$$

from which we derive the following system of model ODEs:

$$\begin{aligned}\frac{d[\text{MmyR}]}{dt} &= k_7[fpm] - k_1[\text{MmyR}][fpm] + k_{-1}[fpm\text{:MmyR}] \\ &\quad - k_4[\text{MmyR}][apm] + k_{-4}[apm\text{:MmyR}] \\ &\quad - \gamma_1[\text{MmyR}],\end{aligned} \quad (19)$$

$$\begin{aligned}\frac{d[\text{MmfR}]}{dt} &= k_8[fpm] - k_2[\text{MmfR}][fpm] + k_{-2}[fpm\text{:MmfR}] \\ &\quad - k_5[\text{MmfR}][apm] + k_{-5}[apm\text{:MmfR}] \\ &\quad + k_{11}[fpm\text{:MmfR:MMF}] + k_{12}[apm\text{:MmfR:MMF}] + \\ &\quad k_{13}[\text{MmfR:MMF}] - k_{14}[\text{MmfR}][\text{MMF}] \\ &\quad + k_{-14}[\text{MmfR:MMF}] - \gamma_2[\text{MmfR}],\end{aligned} \quad (20)$$

$$\begin{aligned}\frac{d[fpm]}{dt} &= k_{11}[fpm\text{:MmfR:MMF}] - k_1[\text{MmyR}][fpm] \\ &\quad + k_{-1}[fpm\text{:MmyR}] + k_{-2}[fpm\text{:MmfR}] \\ &\quad - k_2[\text{MmfR}][fpm],\end{aligned} \quad (21)$$

$$\begin{aligned}\frac{d[apm]}{dt} &= k_{12}[apm\text{:MmfR:MMF}] - k_4[\text{MmyR}][apm] \\ &\quad + k_{-4}[apm\text{:MmyR}] + k_{-5}[apm\text{:MmfR}] \\ &\quad - k_5[\text{MmfR}][apm],\end{aligned} \quad (22)$$

$$\frac{d[fpm\text{:MmyR}]}{dt} = k_1[\text{MmyR}][fpm] - k_{-1}[fpm\text{:MmyR}], \quad (23)$$

$$\frac{d[apm\text{:MmyR}]}{dt} = k_4[\text{MmyR}][apm] - k_{-4}[apm\text{:MmyR}], \quad (24)$$

$$\begin{aligned}\frac{d[fpm\text{:MmfR}]}{dt} &= k_2[\text{MmfR}][fpm] - k_{-2}[fpm\text{:MmfR}] \\ &\quad - k_3[fpm\text{:MmfR}][\text{MMF}] + \\ &\quad k_{-3}[fpm\text{:MmfR:MMF}],\end{aligned} \quad (25)$$

$$\frac{d[\,apm\text{:MmfR}]}{dt} = k_5[\,\text{MmfR}]\,[\,apm] - k_{-5}[\,apm\text{:MmfR}] - k_6[\,apm\text{:MmfR}]\,[\,\text{MMF}] + k_{-6}[\,apm\text{:MmfR:MMF}]\,, \tag{26}$$

$$\frac{d[fpm\text{:MmfR:MMF}]}{dt} = k_3[fpm\text{:MmfR}]\,[\,\text{MMF}] - k_{-3}[fpm\text{:MmfR:MMF}] - k_{11}[fpm\text{:MmfR:MMF}]\,, \tag{27}$$

$$\frac{d[\,apm\text{:MmfR:MMF}]}{dt} = k_6[\,apm\text{:MmfR}]\,[\,\text{MMF}] - k_{-6}[\,apm\text{:MmfR:MMF}] - k_{12}[\,apm\text{:MmfR:MMF}]\,, \tag{28}$$

$$\begin{aligned}\frac{d[\,\text{MMF}]}{dt} = {} & k_9[fpm] - k_3[fpm\text{:MmfR}]\,[\,\text{MMF}] \\ & + k_{-3}[fpm\text{:MmfR:MMF}] + k_{-6}[\,apm\text{:MmfR:MMF}] \\ & - k_6[apm\text{:MmfR}][\text{MMF}] + k_{11}[fpm\text{:MmfR:MMF}] + \\ & k_{12}[\,apm\text{:MmfR:MMF}] + k_{13}[\,\text{MmfR:MMF}] \\ & - k_{14}[\,\text{MmfR}]\,[\,\text{MMF}] + k_{-14}[\,\text{MmfR:MMF}] \\ & - \gamma_3[\,\text{MMF}]\,,\end{aligned} \tag{29}$$

$$\frac{d[\,\text{MMY}]}{dt} = k_{10}[\,apm] - \gamma_4[\,\text{MMY}]\,, \tag{30}$$

$$\begin{aligned}\frac{d[\,\text{MmfR:MMF}]}{dt} = {} & k_{11}[fpm\text{:MmfR:MMF}] + k_{12}[\,apm\text{:MmfR:MMF}] \\ & - k_{13}[\,\text{MmfR:MMF}] + k_{14}[\text{MmfR}]\,[\text{MMF}] \\ & - k_{-14}[\text{MmfR:MMF}]\,,\end{aligned} \tag{31}$$

where square brackets denote concentration and the reaction rate constants translate to model parameters. Reactions associated with reversible DNA:protein binding (k_1, k_{-1}, k_2, k_{-2}, k_4, k_{-4}, k_5 and k_{-5}), the production of MmyR, MmfR, MMF and MMY (k_7, k_8, k_9 and k_{10}) and each individual protein degradation reaction ($\gamma_{1,2,3,4}$) are common to all of our candidate model architectures. Other reactions that are associated with the release of MmfR from existing DNA:MmfR complexes or the sequestration of MmfR and MmyR via binding in solution are not common to all models and are thus subject to investigation through our computational simulations.

Model simulations are provided by the numerical solutions to the relevant model ODEs, which are calculated using the ODE solver ode45 in MATLAB. We are interested in examining the dynamics of methylenomycin production in each of the 48 candidate models and therefore analyze the simulations of MMY provided by numerical solutions to the corresponding ODE (30).

Available experimental data

Methylenomycin production by *S. coelicolor* has been shown to adopt a typical dynamical profile [19, 20]. Once expression is initiated, usually by environmental conditions that are thought to establish MMF production, it increases relatively quickly towards a global maximum level. Expression then decreases from this maximum, reaching a relatively low level at steady-state. This profile aligns with the premise that the system is initially held in a repressed state until MmfR is released by MMF to trigger methylenomycin expression, which then increases quickly until free MmfR and MmyR cause secondary repression and eventual equilibrium of the feedback loop.

We consider the binding affinity of MmfR to the *fpm* and *apm* to be strong, based on experimental data regarding binding interactions between a similar protein, SAV2270, and its associated DNA motifs (our unpublished data). We characterized the binding of this protein to Streptavidin Immobilized oligonucleotides using a Biocore T200 SPR instrument. Our data reveal that the association and dissociation rates of this protein:DNA binding are on the order of 10^5 $M^{-1}s^{-1}$ and 10^{-2} s^{-1} respectively. As a result, we fix the model parameters relating to MmfR association and dissociation from both the *fpm* and *apm* at 10^5 and 10^{-2} respectively ($k_2 = k_5 = 10^5$; $k_{-2} = k_{-5} = 10^{-2}$). The dimensionality of our experimental measurements agree with the corresponding parameters in our dimensional model and we are therefore able to apply these values directly. We assume that MmyR binding interactions are identical to that of MmfR and hence the same values are fixed for the parameters describing MmyR association and dissociation from the *fpm* and *apm* ($k_1 = k_4 = 10^5$; $k_{-1} = k_{-4} = 10^{-2}$).

Mutant strains of *S. coelicolor* that account for specific gene knockouts reveal qualitatively different methylenomycin production dynamics (Table 1). The mutant strain accounting for *mmyR* deletion, Δ*mmyR*, has been shown to exhibit increased methylenomycin expression compared to the wildtype; in the absence of MmyR, the overall capacity of the system to repress methylenomycin production is reduced and therefore the production of the antibiotic is increased. The Δ*mmfLHP* strain exhibits a complete cessation of methylenomycin expression; in the absence of the *mmfLHP* genes, the system

Table 1 The effects of knocking out certain genes and combinations of genes observed experimentally, adapted from [9]

S. coelicolor strain	Methylenomycin production
Wildtype	+
Δ*mmyR*	+++
Δ*mmfLHP*	-
Δ*mmfLHP*+Δ*mmyR*+Δ*mmfR*	+++
Δ*mmfLHP*+exogenous MMF	+

The wildtype strain is allocated a single '+' to denote typical methylenomycin expression. Over-expression and the cessation of expression are denoted by '+++' and '-' respectively

is locked in the *apm*:MmfR complex since the expression of MmfR, MmyR and particularly MMF is prevented and thus the bound MmfR cannot be released. The Δ*mmfLHP*+Δ*mmyR*+Δ*mmfR* strain exhibits increased methylenomycin production compared to the wildtype; in the absence of MmfR and MmyR, both initially and as a result of any subsequent production by the *fpm*, the *apm* is able to produce methylenomycin in an unrestricted manner. The Δ*mmfLHP* strain with exogenous MMF exhibits relatively similar methylenomycin expression to that of the wildtype; in the absence of endogenous MMFs, exogenous MMF permits the release of MmfR and, in turn, methylenomycin expression. Experimentation with the Δ*mmfR* strain has thus far yielded inconclusive results and, as such, presents the opportunity for mathematical modeling simulations to inform future experimental studies.

Model selection via approximate Bayesian computation

In order to assess the potential of the 48 candidate architectures to reproduce the known characteristics of the system, we perform model selection based on approximate Bayesian computation (ABC) using the ABC-SysBio software package. ABC-SysBio combines Bayes' rule with sequential Monte Carlo (SMC) approaches to solve parameter inference and model selection problems in systems biology [21–23]. The procedure determines the model, from a set of candidate models, that is most likely to have produced the associated experimental data. Extensive quantitative data regarding methylenomycin expression is lacking in the literature, however a time series expression profile is reported in [20]. We therefore provide ABC-SysBio with a dataset designed to replicate this profile (Fig. 3), with two important exceptions. Firstly, we specifically account for the dynamical series of data points in the 40 hour interval between hours 54 and 94 of the time series. This is because the 54 hour experimental time point is when methylenomycin expression commences and translates to the 0 hour time point in our simulations. The time points that precede 54 hours record the repression of methylenomycin production prior to the environmental trigger and are hence excluded when fitting a model that accounts purely for the dynamical response of the system. Secondly, we incorporate additional uniformly distributed 'synthetic' data points, increasing the size of the dataset from 5 points to 41, in order to provide a more rigorous data fitting task to the ABC-SysBio algorithm.

ABC-SysBio also requires a prior probability distribution on each model parameter subject to inference in order to establish the parameter space within which to locate acceptable parameter sets. The prior distributions chosen for all parameters associated with each of the 48 candidate models are uniform distributions on

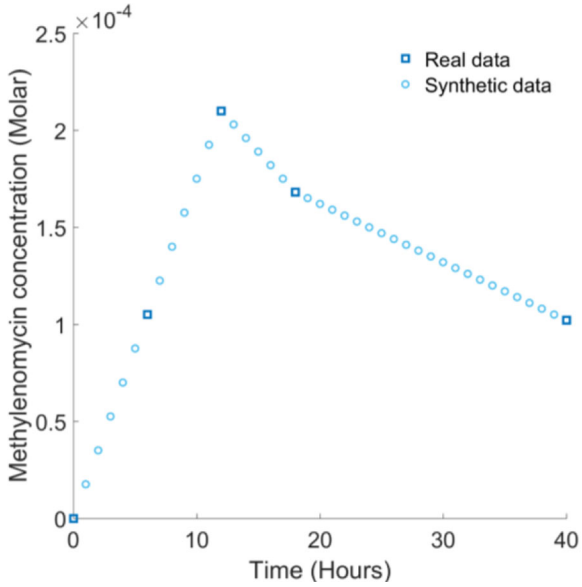

Fig. 3 Experimental data representing current biological knowledge of typical methylenomycin expression in *S. coelicolor*. Real experimental data points taken from [20]. Synthetic data points are added uniformly between real data points to increase the rigor of model selection and parameter inference

the interval $[10^{-4}, 10^{4}]$, that is, all candidate models are given an equal parameter space in attempting to identify parameter values capable of replicating our experimental dataset. We consider uniform priors to be the most suitable for model selection given the complete uncertainty surrounding the parameters subject to inference, and hence all potential parameter values require an equal probability of selection. We also impose prior distributions on the initial conditions of the necessary state variables due to the lack of experimental data regarding the physical quantity of DNA in the system: the prior distributions are uniform distributions on the interval $[0, 1]$ and are assigned only to the MmfR:*fpm* and MmfR:*apm* complexes, all other initial conditions are set equal to zero. ABC-SysBio convergence is dependent on the sequential satisfaction of a predefined series of decreasing error thresholds by a predefined number of solutions (see Methods). Here, the number of solutions required to satisfy each error threshold is 500 [24] in order to reduce the time frame required for convergence; the number of models subject to selection coupled with the inability to parallelize the process presents a particularly time consuming computational workload. The user-defined error function designed to measure the accuracy of simulations takes the mean absolute value of the difference between model outputs and data values:

$$E = \frac{1}{41}\sum_{i=1}^{41}|x_i - d_i|, \tag{32}$$

where E is the error and x_i, d_i are the model outputs and data values at each of the 41 corresponding time points, t_i, respectively.

The results of our model selection are shown in Fig. 4. The final probability distributions reveal that the model most likely to have produced the experimental data is BNN, the model formulated based on our current knowledge (Fig. 4a). The BNN model achieved a 0.916 probability of producing our data which is vastly superior to the remaining models, 36 of which were statistically eliminated through the selection process. This suggests that the most plausible network of molecular interactions underlying this system should account for MMF-MmfR interactions both at existing DNA:MmfR complexes and in solution, no MMY-MmfR interactions at all and no MMY-MmyR interactions at all, as depicted in Fig. 2.

In order to verify that the addition of synthetic data points does not restrict the emergence of other viable candidate models, we repeated the model selection procedure using only the 5 real experimental data points taken from [20]. Mean absolute error generally increases with decreasing numbers of data points which subsequently increases the difficulty for each population of solutions to meet the same error thresholds. Hence, the acceptance rate decreases and the process becomes more time consuming; this run took longer than the original run and met 7 thresholds compared to the previous 11 (Fig. 4b). The probability distribution across all models clearly identified the most likely models as early as P2, which converged further at P4 and P6 to suggest that BNN was a likely model architecture, in agreement with our initial result. However, P3, P5 and P7 identified a different distribution which suggested that models SCS and SNS were also likely candidates. Given that ABC-SysBio appeared to present two alternating probability distributions, it is probable that additional local minima were identified in this case. To further investigate the set of plausible models identified using this Bayesian inference framework, we next employed global optimization methods, as well as analysis of mutant versions of the candidate models, as described in the following sections.

Parameter inference via global optimization

ABC-SysBio performs parameter inference by producing probability distributions on the numerical values that comprise accepted parameter sets during the model

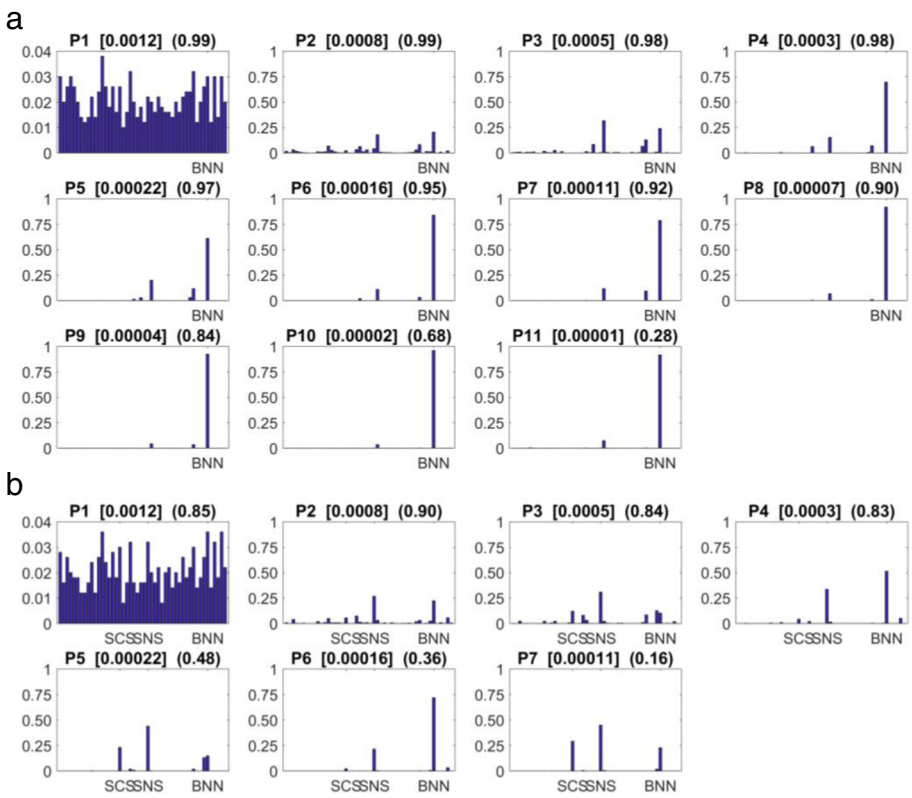

Fig. 4 ABC-SysBio model selection results. **a** Histograms showing the probabilities of producing the full dataset for the 48 candidate models. **b** Histograms showing the probabilities of producing the real experimental dataset for the 48 candidate models. The numbers above each histogram denote the population number, the error threshold ϵ (*square brackets*) and the acceptance rate (*parentheses*) respectively. The number of accepted solutions required to satisfy each error threshold is 500

selection process. For example, the distributions on the initial conditions imposed on the *fpm*:MmfR and *apm*:MmfR complexes reveal that statistically these values can be approximated to be 0.6 and 0.5 respectively (Fig. 5). These distributions are insightful, but cannot provide complete clarity over the numerical values inferred in all cases. Other parameter inference methods, such as global optimization, place greater focus on the identification of specific numerical parameter sets capable of minimizing the user-defined error function. The genetic algorithm (GA), is a particularly powerful global optimization tool and is exploited regularly in biological model parameter inference [25, 26]. The GA converges to the solution providing the global minimum error within the allocated parameter space by evolving an initial population of randomly generated solutions over a large number of generations. This process is based on natural selection, giving the best solutions in the population the best chance of creating the next generation of solutions (see Methods).

In our case, the error function minimized by the GA is the same absolute mean error function used for ABC-SysBio model selection (32). We also allocate the same parameter space to the GA by imposing lower and upper bounds on the inferred parameters of 10^{-4} and 10^{4} respectively. Again, the initial conditions imposed on the model variables are zero with the exception of those regarding the *fpm*:MmfR and *apm*:MmfR complexes which we approximate to be 0.55 in light of our ABC-SysBio probability distributions and given that we require both initial concentrations to be equal. The results of our global optimization are shown in Fig. 6. The BNN model is capable of accurately matching the experimental time-course data when optimized within the same parameter space used for model selection. The optimal parameter set identified by the GA is listed in Table 2 and provides an absolute mean error of 6.12×10^{-6}. The four parameter values describing protein degradation ($\gamma_{1,2,3,4}$) vary by one order of magnitude at most; the remaining parameter values all describe protein:protein association and dissociation and vary by three orders of magnitude at most. Hence, we conclude that the numerical ranges of these optimal parameter values are reasonable within this biological context.

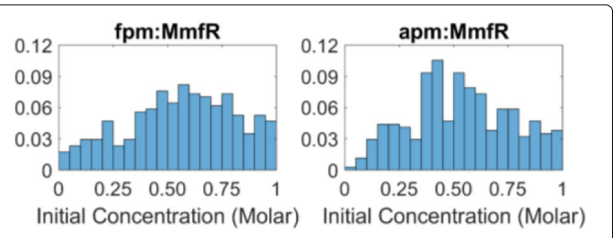

Fig. 5 ABC-SysBio parameter inference results. Histograms show the probability distributions on the two parameters describing the initial concentration of the *fpm*:MmfR and *apm*:MmfR complexes

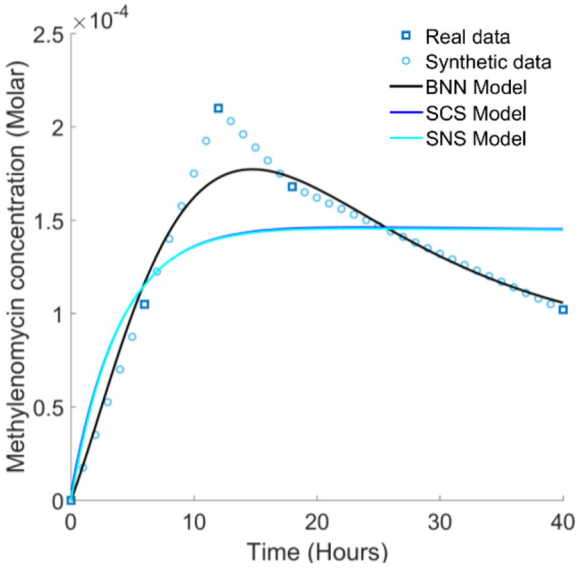

Fig. 6 Genetic algorithm global optimization results. The BNN model is able to fit the experimental data using the optimal parameter set identified by the GA. The absolute mean error provided by this optimal solution is 6.12×10^{-6}. The optimal fits provided by the SCS and SNS models are similar and are not as accurate as the BNN model. The absolute mean error provided by both of these optimal solutions is 2.39×10^{-5}

To investigate further, we also optimized the parameters of the SCS and SNS models against the experimental data using the GA, in an identical manner to that previously

Table 2 Optimal parameter values for our BNN model

Reaction	Value ($M^{-1}s^{-1}$)	Reaction	Value (s^{-1})
k_3	3.6119	k_{-3}	0.1092
k_6	0.9079	k_{-6}	5.7766
k_{14}	0.0065	k_{-14}	0.2208
–	–	k_7	2.6978
–	–	k_8	0.8902
–	–	k_9	5.8903
–	–	k_{10}	0.1101
–	–	k_{11}	0.6296
–	–	k_{12}	0.5307
–	–	k_{13}	0.0880
–	–	γ_1	0.9470
–	–	γ_2	2.7057
–	–	γ_3	0.2248
–	–	γ_4	0.1646

This optimal parameter set is dimensional, with parameters in the first and second reaction columns taking the units $M^{-1}s^{-1}$ and s^{-1} respectively based on standard mass action kinetics

done for the BNN model. This revealed that neither model was able to achieve the same quality of fit to the data as the BNN (minimum error of 2.39×10^{-5} for both SCS and SNS compared to 6.12×10^{-6} for BNN). In addition, neither the SCS or SNS models were able to even qualitatively replicate the non-monotonicity in the response that is clearly exhibited in the experimental data.

Monte Carlo simulations of methylenomycin production in mutant strains

We performed additional model validation by testing the BNN model against our qualitative data regarding methylenomycin production in mutant *S. coelicolor* strains. We employ Monte Carlo simulations to examine methylenomycin production under four distinct conditions corresponding to the mutant strains described in Table 1. By examining the dynamical response to specific gene knockouts against the wildtype strain, represented by the optimal BNN model output in Fig. 6, we are able to investigate the qualitative effect of adapting our BNN model to emulate these mutant strains.

When simulating MMY production in the different mutant strains, we account for Δ*mmyR* by simply setting the parameter describing MmyR production from the *fpm*, k_7, to zero. However, Δ*mmfR* strains are incapable of producing MmfR and therefore cannot be simulated in the initial repressed state comprised of the *fpm*:MmfR and *apm*:MmfR complexes. Hence, the parameter describing MmyR production from the *fpm*, k_8, is set to zero and the allocation of initial concentrations is adapted to exclude the *fpm*:MmfR and *apm*:MmfR complexes. The Δ*mmfLHP* strain is simulated by setting the initial concentration of the *fpm* and its associated complexes to zero, since the entire DNA module has been knocked out. The addition of exogenous MMF involves allocating this variable an initial concentration of 0.55 to align with the initial concentrations allocated to the relevant variables, that is, no new model parameters are introduced to describe production of exogenous MMF. Mutant strains comprising combinations of gene knockouts are simulated by combining the appropriate adaptations.

Specifically, in order to simulate the Δ*mmyR* strain we set $k_7 = 0$. To simulate the Δ*mmfLHP* strain the initial concentration of 0.55 is imposed on the *apm*:MmfR complex only, all other initial concentrations are set equal to zero. To simulate the Δ*mmfLHP*+Δ*mmyR*+Δ*mmfR* strain we set $k_7 = k_8 = 0$ and all initial concentration are set equal to zero with the exception of the *apm* which is set equal to 0.55. To simulate the Δ*mmfLHP*+exogenous MMF strain initial concentrations of the *apm* and MMF are set equal to 0.55, and all other initial concentrations are set equal to zero.

Monte Carlo simulations assign random values in the interval $[10^{-4}, 10^{4}]$ to all model parameters, excluding those that retain their fixed values assigned for previous model selection and parameter inference purposes, as we continue to examine dimensional dynamic responses. We run a total of 10^4 Monte Carlo simulations to allow for substantial sampling of the parameter space within a feasible time frame. Each simulation outputs MMY production for each of the four mutant strains and calculates the ratio of the mean value to that of the optimal wildtype simulation. We utilize these ratios to investigate the ability of our model to satisfy the following four criteria, which capture the experimentally observed responses of the mutant strains:

1. $\frac{\Delta mmyR}{\text{wildtype}} > 1.1$,
2. $\frac{\Delta mmfLHP}{\text{wildtype}} < 0.9$,
3. $\frac{\Delta mmfLHP+\Delta mmyR+\Delta mmfR}{\text{wildtype}} > 1.1$,
4. $0.9 < \frac{\Delta mmfLHP+\text{MMF}}{\text{wildtype}} < 1.1$,

where overproduction translates to an increase in mean MMY production of $<$ 10%, cessation translates to a decrease in MMY production of $<$ 10% and comparable production translates to a maximum increase or decrease in MMY production of less than 10%. The results of our Monte Carlo simulations are shown in Fig. 7. Parameter sets were identified for BNN that are capable of satisfying each of the four criteria, within the same dimensional solution space as the optimized wildtype model (Fig. 7a). Given the uncertainty regarding the effect of gene knockouts on the reaction kinetics and MMY production of the system, this qualitative agreement offers further validation of the replication and prediction capabilities of the BNN model.

The SCS and SNS models are also able to simulate the responses observed experimentally for the Δ*mmyR* and the Δ*mmfLHP*+Δ*mmyR*+Δ*mmfR* knockout strains, but not for the Δ*mmfLHP* and Δ*mmfLHP*+exogenous MMF knockouts (Fig. 7b and 7c). This is likely due to the most significant mechanistic property separating them from BNN, i.e. the interaction of MMY with one or both of MmyR and MmfR. This interaction results in decreased repression of the *apm* since the MMY is negating the action of either or both regulators and hence the *apm* is less restricted in producing MMY, which causes an overproduction of the antibiotic for the Δ*mmfLHP* and Δ*mmfLHP*+exogenous MMF knockouts that has not been observed experimentally (Fig. 7b and 7c). We therefore conclude that the BNN model remains the most likely candidate model to explain all the available experimental data for this system.

Experimental design for future studies

We are able to inform the design of future experimental studies in light of our results. For example, we are

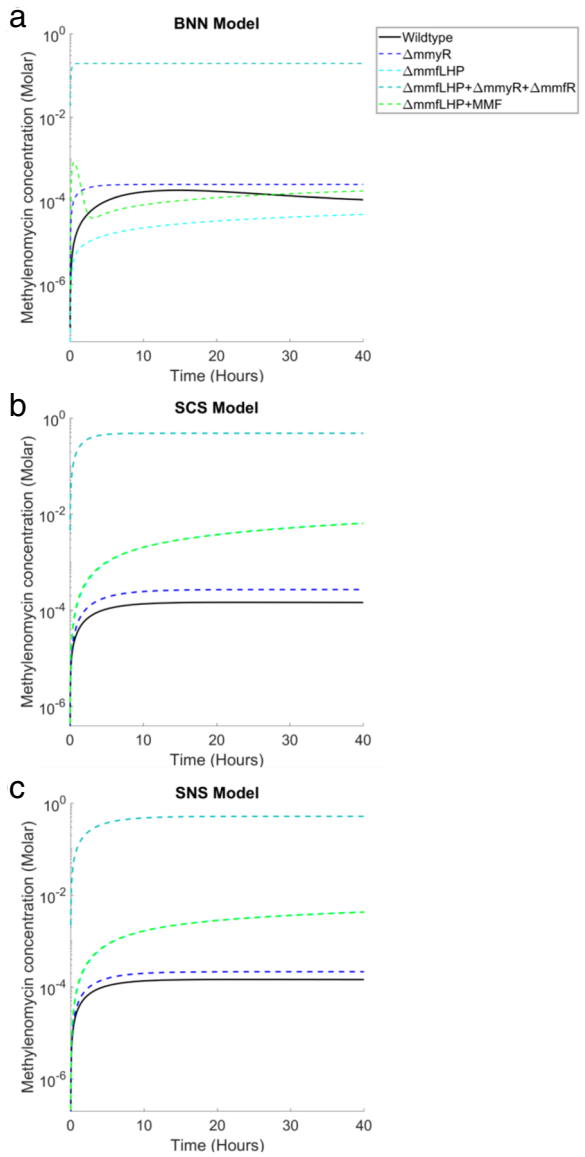

Fig. 7 Monte Carlo simulation results. **a** The BNN model is able to simulate the qualitative data regarding four mutant *S. coelicolor* strains when adapted to replicate the corresponding gene knockouts. **b** The SCS model is unable to simulate qualitative data regarding the ΔmmfLHP and ΔmmfLHP+MMF knockout strains. **c** The SNS model is unable to simulate qualitative data regarding the ΔmmfLHP and ΔmmfLHP+MMF knockout strains

interested in quantifying the response of the $\Delta mmyR$ and $\Delta mmfLHP + \Delta mmyR + \Delta mmfR$ strains in order to verify our model prediction that the five gene mutant elicits a more rapid and significantly greater overproduction of MMY. This has implications both in terms of improving product yields for industrially relevant natural products, and also regarding the potential adverse effects this might cause in the cells, such as toxicity. The result of these experiments would subsequently reveal whether the $\Delta mmfLHP + \Delta mmyR + \Delta mmfR$ is the most effective knockout for improving antibiotic production in novel synthetic regulatory systems.

In the event that directly quantifying MMY production is inconclusive, we would be interested in replacing the gene controlled by the *apm* with a reporter gene coding for fluorescence or luminescence such as green fluorescent protein (GFP) or *lux* genes respectively. This output may enable us to measure the response of the different mutant strains with greater clarity, since experiments of this nature are already well characterized, particularly in the related bacterium *S. venezuelae*.

Finally, we are also interested in examining the $\Delta mmfLHP$+MMF mutant in order to establish the quantity of exogenous MMF and the specific time point of induction that provides optimal MMY production. Our model predicts a narrow production window for this strain which may suggest that direct MMY quantification is not straightforward and that, again, experimental designs incorporating reporter genes would provide improved results.

Conclusions

We have developed a plausible model architecture for the regulatory system controlling methylenomycin production in *Streptomyces coelicolor*. This architecture was found to most closely reproduce the various dynamical responses described by experimental time series data for this system, when tested against 47 other candidate architectures. Global optimization of the model parameters produced close agreement with the experimental data. Appropriate adjustments to the proposed model architecture allow it to replicate observed changes in the dynamics of methylenomycin production in a number of mutant *S. coelicolor* strains.

The mechanistic details captured in the proposed regulatory architecture provide useful insights for the design of future experiments to further investigate the operation of this system, and demonstrate the potential of mathematical models to elucidate the design principles of complex biological control systems. We expect that the emergence of further quantitative experimental data for this system will inform further model development and validation, and allow for the generation of optimized models that are capable of accurately predicting the dynamical responses of one of the most prevalent and important gene regulatory networks in nature.

Methods

ABC-SysBio model selection

ABC-SysBio is a Python software package that is designed specifically for parameter inference and model selection in biological systems research using the approach

of approximate Bayesian computation (ABC) [21]. The program enables ABC inference of mathematical models via sequential Monte Carlo (SMC) approaches [21–23]. Monte Carlo approaches to computational simulations involve generating random candidate solutions, testing their fitness against a desired output and repeating until a viable solution can be identified. In this way, vast numbers of randomly selected parameter sets can be examined in building an accurate approximation to the posterior distribution defined by conditional probabilities known as Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{33}$$

where $P(B) > 0$. The ABC-SMC approach proceeds in the following manner: the first 'population' of accepted solutions or 'particles' is generated randomly based on the prior distributions imposed on the model parameters. Each particle gives rise to a simulated dataset, $D^{\star}$, which is compared to the fixed experimental dataset, D, by an appropriate distance function and its fitness is scored accordingly. Acceptance of a particle is dependent on a decreasing sequence of error thresholds, ϵ, set to correspond with each population. That is,

$$d(D^{\star}, D) < \epsilon_i, \tag{34}$$

where $\epsilon_1 > \epsilon_2 > \ldots > \epsilon_n$ and d is the distance function. Each subsequent population is obtained by perturbing particles from the previous population in accordance with a predetermined perturbation kernel, proceeding until the model is unable to produce particles of sufficient fitness to satisfy the immediate threshold.

An array of model-specific criteria are required to allow the ABC-SysBio package to run efficiently: the sequence of decreasing error thresholds, ϵ, must be provided whereby only the particles capable of providing error less than that of the threshold will be accepted by the algorithm. Each ϵ must be satisfied in succession until the particles are unable to satisfy the next threshold. Satisfaction of an individual threshold is dependent on the number of particles accepted; the number of acceptable particles required before passage to the next threshold must also be predetermined. The larger the number of particles, the higher probability of significant inference and the longer the duration of algorithm to reach convergence. Each individual parameter subject to inference requires a prior probability distribution in order to establish the parameter space within which to locate acceptable particles. Sequences of numerical values representing the relevant experimental data and the corresponding time points must also be provided; the number of data points and time points must be equal. Time series data is currently the only supported data format. One or more distinct datasets can be supplied and can be fitted to any individual model variable or combination of variables. Convergence of the algorithm is dependent on all of the aforementioned factors and hence it may require several trials to establish the appropriate performance criteria. To achieve credible results, it is advised that identical parameter inference and model selection tasks are repeated multiple times due to the random nature of the Monte Carlo simulations that drive the algorithm. Note that all models submitted to the ABC-SysBio package must be written in Systems Biology Markup Language (SBML), a systems biology programming language based on Extensible Markup Language (XML).

Global optimization

We employ the genetic algorithm function in MATLAB in order to optimize our model against our experimental data. The reaction rate constants k_i are chosen as optimization variables with the exception of those fixed in light of our kinetic data. The GA mimics natural selection; converging to the global minimum within the allocated parameter space by evolving an initial population of randomly generated solutions over a large number of generations. The probability of obtaining the global optimum solution is maximized by selecting the largest population size and number of generations possible. However, increasing the computational workload in this manner also greatly increases the time frame required for the algorithm to converge. Establishing an effective compromise is key for successful deployment. We ran the GA under the following conditions:

- Population: 1000
- Generations: 1000
- Bounds imposed on parameter values: $[10^{-4}, 10^{4}]$

The default GA population size in MATLAB is 200 for inference of over 5 parameters however, we selected a population size of 1000 for inference of 17 parameters since a single time series dataset presents a relatively low computational workload and thus allows optimal solutions for large populations to be obtained in feasible time frames. We selected a large parameter space due to our focus on establishing optimal model performance in light of the lack of documentation regarding reaction rates.

Abbreviations

ABC: approximate Bayesian computation; *apm*: antibiotic producing module; DNA: deoxyribonucleic acid; *fpm*: furan producing module; GA: genetic algorithm; GFP: green fluorescent protein; MARE: methylenomycin auto-regulatory response element; MMF: methylenomycin furan; MMY: methylenomycin; ODE: ordinary differential equation; SBML: Systems biology markup language; SMC: Sequential Monte Carlo; *S. coelicolor*: *Streptomyces coelicolor*; XML: Extensible markup language

Acknowledgements
Not applicable.

Funding
This research was supported by: EPSRC via a DTA studentship to JB, EPSRC and BBSRC via research grant BB/M017982/1 to EdIS, BBSRC via MIBTP to KS, the Royal Society via a University Research Fellowship UF090255 to CC and BBSRC via grant BB/M022765/1 to CC.

Availability of data and materials
The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions
JB designed the study, performed all mathematical and statistical analyses, analyzed and interpreted the results and contributed to the writing of the manuscript. EdIS designed the study, analyzed and interpreted the results, provided experimental insight and was a contributor to the writing of the manuscript. KS provided experimental insight and was a contributor to the writing of the manuscript. AF provided the experimental data which informed the numerical values of the fixed dimensional model parameters. CC designed the study, provided experimental insights and was a contributor to the writing of the manuscript. DB designed the study, analyzed and interpreted the results, and was a contributor to the writing of the manuscript. All authors give final approval for publication.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry CV4 7AL, UK. [2]Warwick Integrative Synthetic Biology Centre, Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK. [3]School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK. [4]Warwick Integrative Synthetic Biology Centre, Department of Chemistry and School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK.

Received: 14 February 2017 Accepted: 18 July 2017
Published online: 03 October 2017

References

1. Watve M Tickoo. How many antibiotics are produced by the genus Streptomyces? Arch Microbiol. 2001;176:386–390.
2. Bentley S, et al. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature. 2002;417(6885):141–7.
3. Willey J, Gaskell A. Morphogenetic signaling molecules of the streptomycetes. Chem Rev. 2011;111(1):174–87.
4. Bentley S, et al. SCP1, a 356 023 bp linear plasmid adapted to the ecology and developmental biology of its host, Streptomyces coelicolor A3(2). Mol Microbiol. 2004;51(6):1615–28.
5. Sidda J, Song L, Poon V. Al-Bassam M, Lazos O, Buttner M, Challis G, Corre C, Discovery of a family of γ-aminobutyrate ureas via rational derepression of a silent bacterial gene cluster. Chem Sci. 2014;S(1):86–89.
6. Laureti L, Song L, Huang S, Corre C, Leblond P, Challis G, Aigle B. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens. Proc Natl Acad Sci U S A. 2011;108(15):6258–63.
7. Chater K, Bruton C. Resistance, regulatory and production genes for the antibiotic methylenomycinare are clustered. EMBO J. 1985;4(7):1893–7.
8. Corre C, Challis G. Evidence for the unusual condensation of a diketide with a pentose sugar in the methylenomycin biosynthetic pathway of Streptomyces coelicolor A3(2). ChemBioChem. 2005;6:2166–2170.
9. O'Rourke S, Wietzorrek A, Fowler K, Corre C, Challis G, Chater K. Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in Streptomyces coelicolor. Mol Microbiol. 2009;71(3):763–78.
10. Corre C. In search of the missing ligands for TetR family regulators. Chem Biol. 2013;20:140–142.
11. Ramos J, et al. The TetR family of transcriptional repressors. Microbiol Mol Biol Rev. 2005;69(2):326–56.
12. Corre C, Song L. O'Rourke S, Chater K, Challis G, 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by Streptomyces coelicolor genome mining. Proc Natl Acad Sci U S A. 2008;105(45):17510–5.
13. Liu G, Chater K, Chandra G, Niu G, Tan H. Molecular regulation of antibiotic biosynthesis in streptomyces. Microbiol Mol Biol Rev. 2013;77(1):112–43.
14. Bonnet J, Subsoontorn P. Endy D.Rewritable digital data storage in live cells via engineered control of recombination directionality. Proc Natl Acad Sci USA. 2012;109(23):8884–9.
15. Hsiao V, Hori Y, Rothemund P, Murray R. A population-based temporal logic gate for timing and recording chemical events. Mol Syst Biol. 2016;12(5):869.
16. Bowyer J, Zhao J, Subsoontorn P, Wong W, Rosser S, Bates D. Mechanistic modeling of a rewritable recombinase addressable data module. IEEE Trans Biomed Circuits Syst. 2016;10(6):1161–70. doi:10.1109/TBCAS.2016.2526668. Epub 2016 May 24.
17. Hsiao V, de los Santos E, Whitaker W, Dueber J, Murray R. Design and implementation of a biomolecular concentration tracker. ACS Synth Biol. 2015;4(2):150–61.
18. Karlebach G. Shamir R.Modelling and analysis of gene regulatory networks. Nat Rev Mol Cell Biol. 2008;9(10):770–80.
19. Hayes A, Hobbs G, Smith C, Oliver S, Butler P. Environmental signals triggering methylenomycin production by Streptomyces coelicolor A3(2). J Bacteriol. 1997;179(17):5511–5.
20. Hobbs G, et al. An integrated approach to studying regulation of production of the antibiotic methylenomycin by Streptomyces coelicolor A3(2). J Bacteriol. 1992;174(5):1487–94.
21. Liepe J, Kirk P, Filippi S, Toni T, Barnes C, Stumpf M. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. Nat Protoc. 2014;9(2):439–56.
22. Stumpf M. Approximate Bayesian inference for complex ecosystems. F1000Prime Rep. 2014;6:60.
23. Smith R, Grohn Y. Use of approximate Bayesian computation to assess and fit models of mycobacterium leprae to predict outcomes of the Brazilian control program. PLoS One. 2015;10(6):e0129535.
24. Woods M, Barnes C. Mechanistic modelling and bayesian inference elucidates the variable dynamics of double-strand break repair. PLoS Comput Biol. 2016;12(10):e1005131. doi:10.1371/journal.pcbi.1005131.
25. Chen B-S, Chen P-W. GA-based design algorithms for the robust synthetic genetic oscillators with prescribed amplitude, period and phase. Gene Regul Syst Bio. 2010;4:35–52.
26. Fernandez M, Caballero J, Fernandez L. Sarai A.Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM). Mol Divers. 2011;15(1):269–89.