



Predictive biology: modelling, understanding and harnessing microbial complexity

Allison J. Lopatkin^{1,2,3,4} and James J. Collins^{1,2,3}✉

Abstract | Predictive biology is the next great chapter in synthetic and systems biology, particularly for microorganisms. Tasks that once seemed infeasible are increasingly being realized such as designing and implementing intricate synthetic gene circuits that perform complex sensing and actuation functions, and assembling multi-species bacterial communities with specific, predefined compositions. These achievements have been made possible by the integration of diverse expertise across biology, physics and engineering, resulting in an emerging, quantitative understanding of biological design. As ever-expanding multi-omic data sets become available, their potential utility in transforming theory into practice remains firmly rooted in the underlying quantitative principles that govern biological systems. In this Review, we discuss key areas of predictive biology that are of growing interest to microbiology, the challenges associated with the innate complexity of microorganisms and the value of quantitative methods in making microbiology more predictable.

Genetic toggle switch

A synthetic gene circuit consisting of two mutually inhibitory repressor genes, their associated promoters and a reporter gene; bistable feedback results in the circuit stably assuming one of two states (that is, toggling between reporter gene ON and OFF) in response to the transient application of exogenous inducers.

Bistable

A system that can exhibit two distinct stable states.

¹*Institute for Medical Engineering & Science and Department of Biological Engineering, MIT, Cambridge, MA, USA.*

²*Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA.*

³*Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA.*

⁴*Department of Biology, Barnard College, New York, NY, USA.*

✉e-mail: jimjc@mit.edu

<https://doi.org/10.1038/s41579-020-0372-5>

In the current era of big data, predictive analytics has revolutionized our everyday lives. Every second, computer systems around the world are scanning millions of credit card transactions to detect likely fraud in real-time and integrating hundreds of thousands of data points to identify potential mechanical failures in error-prone machines such as airplanes and cars. In this context, predictive biology, or the ability to predict a biological outcome from a set of known inputs (or vice versa), seems well positioned to benefit from the proliferation and use of big data. However, over the past several decades, it has become increasingly apparent that biology does not fall into this framework as readily as we had hoped. This may not be necessarily surprising for complex multicellular organisms, yet it is particularly striking in the world of single-celled microorganisms. *Escherichia coli*, for example, is arguably one of the best-characterized model organisms with over two decades of a fully sequenced and comprehensively annotated genome¹; nevertheless, as of 2019, the function of ~35% of its protein-coding genes remains unknown².

In many ways, the growth of synthetic biology encapsulates both the capabilities and limitations of these analytical advances. On the one hand, increasingly well-characterized genetic components have enabled us to programme biological networks with precisely defined logic. For example, the genetic toggle switch was modelled, designed and built to function as a bistable memory element³, thereby demonstrating one of the most fundamental blocks of cellular decision-making⁴. On the other

hand, even simple gene networks often lead to unforeseen dynamics, forcing us to confront our incomplete grasp of the basic biological rules governing cellular behaviour. Indeed, despite its apparent robustness, small modifications to the original toggle design turned out to be deceptively complex, leading to undesirable behaviour^{5,6}.

The year 2020 marks two decades since the publication of the genetic toggle switch and repressilator papers^{3,7} — of note, the repressilator is a synthetic gene circuit that was modelled, designed and built to function as a ring oscillator. In those 20 years, modern experimental techniques have emerged with the capability of placing a massive amount of biological data at our fingertips. However, translation of these data into mechanistic understanding and meaningful biological insights remains a difficult, labour-intensive task. A recent article estimated that the effort devoted to data analysis and integration can quickly outweigh that needed to obtain the data in the first place⁸, posing challenges to tasks such as identifying microbial community dynamics from metagenomic data⁹, establishing accurate kinetic models of bacterial metabolism¹⁰ or inferring gene regulatory networks¹¹. Although machine learning and other artificial intelligence (AI) techniques excel at analysing and detecting trends and/or clusters within big data, they fall short in the genesis and refinement of underlying biological explanations, particularly as these relationships are often nonlinear. Commensurately detailed quantitative analyses are needed to translate this wealth of data into concrete biological insights.

Box 1 | Types of models in predictive biology

The purpose of a mathematical model is to describe a variable or variables of interest in a precise and quantitative manner. In the simplest case, a variable Y that changes linearly with changes in X can be modelled with the linear equation $Y = b_1 X + b_0$, where b_1 and b_0 are fitted parameters, and b_1 relates the change in Y to a unit change in X . However, because these parameters may not necessarily have biological meaning, cellular-level or molecular-level models are typically derived from biochemical ‘first principles’.

Dynamical models of cellular or molecular systems describe changes in a variable (or group of variables) over time. For example, consider a population N that doubles every generation. If N_g is the population size at generation g , then the model $N_{g+1} = 2 \times N_g$ predicts the population size at the next generation. Discrete dynamical models, such as the above, are advantageous in that they are computationally simple and easily abstracted (for example, this relationship describes population sizes regardless of specific proliferation rate). As such, these models have proven particularly useful when simulating microbial changes over successive time steps such as seasons of growth or mutation acquisition. In contrast to discrete models, continuous dynamical models describe how one or more state variables change over an arbitrary time window, represented by a set of ordinary differential equations (ODEs). For example, consider the reaction rate of an enzyme that converts a substrate (S) into a product (P). From first principles, the ODE describing the formation of P ($\frac{dP}{dt}$) can be derived from the biochemical reactions governing (1) the reversible binding between an enzyme (E) and S, and (2) the catabolic breakdown of ES (enzyme–substrate complex) into P. The resulting ODE $\frac{dP}{dt} = v_{\max} \frac{S}{K_m + S}$, known as the Michaelis–Menten equation, models this process, where v_{\max} is the maximum enzymatic rate, K_m is the substrate concentration at which exactly half of the maximum rate is achieved, and t is time. Moreover, we can readily estimate K_m and v_{\max} by measuring a series of product formation rates across a range of [S] (substrate concentrations).

The Michaelis–Menten model implicitly assumes that the number of enzyme and substrate molecules is sufficiently large such that their binding occurs at a constant rate regardless of [S]. This exemplifies a deterministic process, or one in which the initial conditions and rate parameters entirely determine the model predictions. Deterministic ODE models are primarily used when the system can be assumed to be homogeneously distributed (for example, well-mixed). However, non-homogeneous systems commonly occur and can be divided into two general categories. First, heterogeneity due to sparse and/or variable conditions give rise to stochastic events, meaning that randomness (referred to as biological noise) may affect the model’s predictions. Noise may be extrinsic, which refers to the natural variability inherent in biochemical events, or intrinsic, which refers to variability associated with specific mechanistic interactions. Whether noise is incorporated into a model is context specific and often depends on the relative magnitude of the process of interest. Second, heterogeneity can arise from non-uniform spatial distributions (for example, structural constraints or concentration gradients). In these cases, agent-based models and partial differential equation models are used to describe discrete or continuous spatial scales, respectively. Agent-based models assume a lattice-like structure that divides a given space into individual compartments, or ‘agents’; simulations then describe how these agents depend on and interact with each other (for example, cell growth within a biofilm). Although using infinitely many compartments can approximate a continuous spatial scale, this approach is computationally intensive and inefficient. Instead, partial differential equations are typically used to describe changes in variables of interest as a function of both time and space, for example, chemical or cellular gradients, intercellular signalling, and pattern formation.

Mathematical models (BOX 1) are uniquely suited to address these demands, and thereby serve as a fundamental basis for AI approaches. Of particular relevance to predictive biology are dynamical models due to the ever-changing nature of biological systems, and these are therefore the primary focus of this Review. Indeed, these classical systems biology models are amongst the most well-established techniques to interrogate, explain and parameterize biological networks. Moreover, minimal models derived from biological first principles can serve as important confirmations of big data-inferred network structures. As such, they remain as relevant today as they

were decades ago. Thus, despite the allure of modern ‘black-box’ tools, such as deep learning, the art of mathematically modelling the underlying biological system must not be forgotten nor neglected; rather, it should be treated as a complementary approach to emerging technologies.

We have aimed this Review at readers across all levels of familiarity with predictive biology and biological modelling; as such, we begin with a brief overview of dynamical systems modelling to orient the reader to the field. We then examine the utility of modelling in the context of increasing biological complexity, starting with lower-level cellular processes (for example, transcription) and progressing to community-level ecological and evolutionary dynamics. This bottom-up modelling approach highlights both the opportunities and challenges that currently exist in predictive biology; accordingly, we discuss the key examples in which mathematical models have elucidated fundamental biological design principles and discuss the crucial need for continued emphasis on the underlying dynamics in biological systems. We conclude by considering the open challenges in biological modelling and explore how next-generation techniques such as deep learning can be combined with systems modelling to realize the full potential of predictive biology.

Modelling microbial dynamics

Classical systems biology modelling has proven to be a foundational tool with which we can interrogate and interpret the many complexities of microorganisms. These approaches typically use differential equations to describe the dynamics of gene regulatory networks by simulating biomolecule concentrations over time (BOX 2). Intuitively, this approach directly ties the observed dynamics to specific biochemical interactions. Such network models can be leveraged to generate experimentally testable hypotheses of a proposed mechanism and identify optimal or feasible parameter spaces to achieve the desired functionality. For example, modelling analysis of the genetic toggle switch revealed the conditions that favour bistability (strong promoter binding and increased cooperativity), and thus guided the experimental implementation to optimize circuit performance³.

Although the mathematical representation of a gene network may be sufficient to accurately simulate the observed dynamics (as in the genetic toggle switch), in many cases, the topology of an individual gene network alone is insufficient to fully predict biological behaviour¹². This disconnect often arises due to unforeseen interactions between network components, the increased metabolic burden placed on cells and biological variability. For example, one study¹³ observed that non-cooperative positive feedback of a transcription factor resulted in the bistability of a downstream fluorescent protein. As bistability requires nonlinear positive feedback^{14,15}, this observation could not be explained by the circuit network alone. However, it has become increasingly apparent that coupling between cellular growth and intracellular biochemical networks adds a layer of both complexity and variability to the predicted

Repressilator

A synthetic gene circuit consisting of three repressor genes, arranged such that gene A represses gene B, gene B represses gene C, and gene C represses gene A; this architecture enables the circuit to produce oscillatory outputs (for example, a reporter gene will cycle between ON and OFF states).

Discrete dynamical models
Changes in the state variable(s)
occur at particular time points
and remain fixed in between
time points.

dynamics, thereby having an important role in circuit functionality^{16–19}. Biochemical network models can readily address this issue by integrating feedback through microbial growth dynamics at the whole-cell or population level. To this end, the same study¹³ demonstrated that the effects of gene expression on metabolic burden

due to growth inhibition introduced nonlinearity into the transcription factor dynamics, and incorporating this feedback into their model explained the observed bistability. Specifically, the combination of metabolic burden and a positive feedback loop was sufficient to recreate the observed bistability: at low transcription factor levels, a minimal metabolic burden meant that cells grew rapidly, thereby maintaining a low fluorescent reporter signal. Conversely, high transcription factor levels induced a substantial metabolic burden, which slowed cell growth and proliferation; this prevented the dilution of the reporter and thus maintained a strong fluorescence signal.

Although population dynamics, which are most readily characterized by relative growth rates, are both convenient and comprehensive, they inevitably mask layers of complexity that may each influence emergent function. One such example involves quorum sensing (QS), which refers to a mechanism by which microorganisms sense their local density by secreting and responding to a signalling molecule. This strategy is susceptible to ‘cheater’ cells, which do not produce the signalling molecule, thereby avoiding metabolic burden and attaining a higher growth rate. Based on growth rate alone, these cheaters would be expected to dominate a population. However, using a dynamic model, a study demonstrated that, counter-intuitively, cheaters experienced only a transient benefit: although they initially outcompeted QS-expressing cells, the smaller QS population did not produce sufficient signalling molecules to sustain the cheater population²⁰. In the context of dynamical modelling, it is often necessary to consider ecological (selection, competition, and so on) or evolutionary (adaptation, horizontal gene transfer, and so on) interactions, or both, to precisely view, quantify and ultimately predict the behaviour of natural and engineered systems. Understanding how these underlying processes scale and influence one another is paramount to developing dynamical models of greater accuracy. Indeed, modelling the interplay between population-level interactions has proven useful in a number of contexts, including predicting the effects of mutant emergence²¹, homologous gene exchange²² and ecological differentiation²³ in bacterial systems.

Modelling increasing levels of complexity

Microbial dynamics, in both natural and engineered contexts, are governed by fundamental intrinsic (transcription, translation and metabolic processes) and extrinsic (environment, ecology and evolution) factors (FIG. 1). These factors and their associated processes are inherently interdependent, and often propagate variability and feedback that can dictate emergent behaviours at the circuit, organism, population and community levels²⁴. For example, one study²⁵ found that synthetic networks comprised of multiple gene repression stages could either attenuate or amplify intercellular variability depending on network complexity and input conditions; in scenarios where variability was minimized, these circuits acted as low-pass filters, maintaining population synchrony despite transient environmental perturbations. These interactions are cumulative, such that each layer builds

Box 2 | Basics of gene network model construction

Dynamical models in synthetic and systems biology often describe gene regulatory networks. To illustrate transforming such a network into a set of ordinary differential equations (ODEs), consider the reactions describing the constitutive production of a protein (transcription and translation; see the figure, part a). As shown in the figure, part a, DNA is transcribed into mRNA at a rate k_r , mRNA is translated into protein at a rate k_p , and mRNA and protein degrade at rates d_r and d_p , respectively. Thus, according to classical chemical kinetics, the production of a protein can be modelled according to coupled ODEs (equations 1 and 2):

$$\frac{d[mRNA]}{dt} = k_r - d_r[mRNA]$$

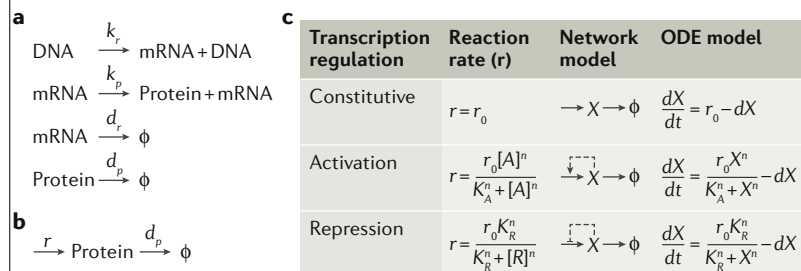
$$\frac{d[protein]}{dt} = k_p[mRNA] - d_p[protein]$$

In this scenario, both [mRNA] and [protein] (mRNA and protein concentrations) will eventually reach a steady state (for example, a constant concentration). Reactions that occur on different timescales can often be simplified, since those that occur more rapidly reach steady state first. Indeed, transcription often occurs substantially faster than translation; thus, to model protein production, we can assume $\frac{d[mRNA]}{dt} \approx 0$, and solve for steady-state mRNA: $[mRNA] = \frac{k_r}{d_r}$. Plugging this into equation 2 and setting $r_0 = \frac{k_r k_p}{d_r}$, we can see that constitutive production of a protein is governed by a single equation (equation 3), where r_0 represents the lumped reaction rate for both transcription and translation. Moreover, this can be represented by a simpler reaction network, as shown in the figure, part b:

$$\frac{d[protein]}{dt} = r_0 - d_p[protein]$$

Molecular networks of all complexity levels can be derived using analogous steps. Since these derivations are well established, a modeller can often transform a network diagram into a set of equations using a few basic transcription regulatory motifs (inhibition and activation) and combinations thereof (feedback). In many cases, transcription rate is proportional to subsequent protein levels. Common models for the concentration of a protein over time for the three general modes of transcription are shown in the figure, part c. Transcriptional regulation occurs either through constitutive production, activation or inhibition. In cases where transcription is proportional to protein levels, each mode of transcription is represented by a particular reaction rate, where r_0 is the lumped production rate, n is the Hill coefficient, [A] and [R] are the activator and repressor concentrations, respectively, and K_A and K_R are the respective half-maximal concentration constants. Assuming these reaction rates, protein X in the network model can be described with the corresponding ODE, where d is the degradation rate of X.

Given the immense complexity in biological systems, modelling is inherently limited by our current knowledge, the accuracy of parameter estimates, and the assumptions we make (for example, the variables we choose to incorporate, separation of timescales, and so on). Establishing and validating a model is therefore an iterative process between parameter estimation, model prediction and testing of relevant hypotheses for further model refinement.



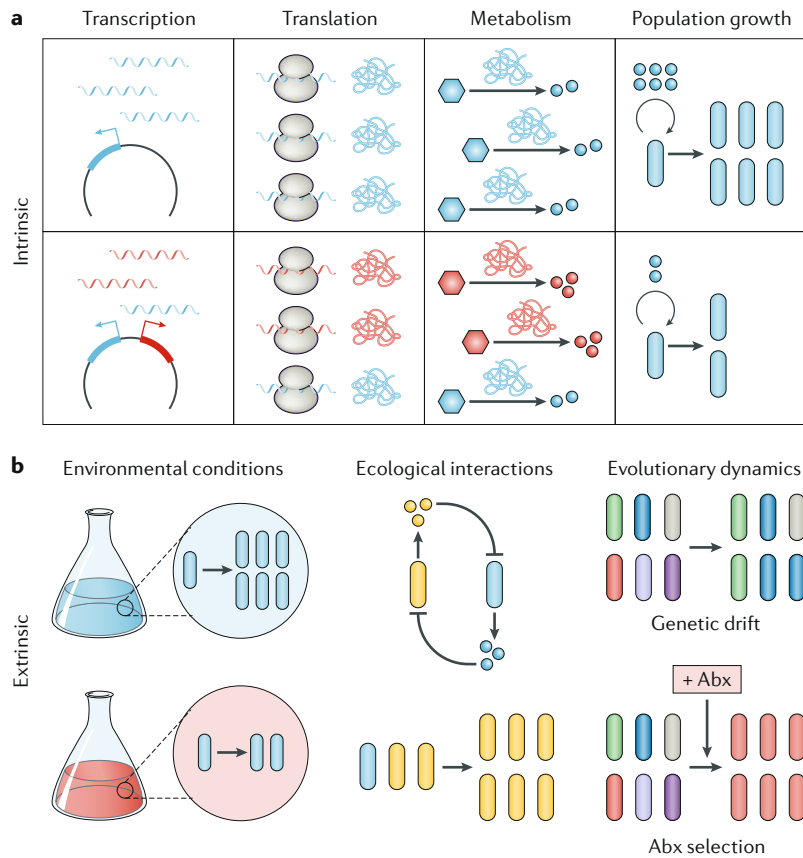


Fig. 1 | Factors that contribute to growth dynamics. **a** | Intrinsic factors that influence growth dynamics. The intrinsic processes that impact overall population dynamics are shown from left to right: transcription, translation, metabolism and population growth. The top row shows the flow of genetic information across all four processes for a single gene (blue). The bottom row shows how changes at the transcription level, represented by the expression of a second gene in red that alters expression of the blue gene, can lead to differences at the population level. **b** | Extrinsic factors that influence growth dynamics. Environmental conditions, represented by a change from a favourable (blue) to less favourable (red) substrate, can differentially impact growth phenotypes (left). The overall population yield and individual species abundance is impacted by local ecological interactions in mixed communities, such as in the suppression of one population (blue) due to the higher prevalence of a second population (yellow) (middle). The same initial population exhibits diverse evolutionary dynamics depending on the absence (top) or presence (bottom) of a selective pressure, represented here by an antibiotic (Abx), that favours a specific genetic variant (right).

of intracellular resources. Consequently, over-expressing just one gene can decrease the expression levels of others²⁶. These hidden and often-overlooked interactions arise due to competition for free ribosomes and RNA polymerases, and result in global changes in growth rates that are attributed to gene expression capacity^{27–29}. Mathematical models have been used both to characterize these interactions and to develop control strategies for gene expression based on resource limitations.

To investigate the effect of gene expression on growth, one study quantified global expression burdens for a variety of diverse synthetic constructs³⁰. This analysis revealed that expression was not strictly inversely proportional to the burden on the host, and this nonlinearity was particularly apparent in constructs with strong ribosome binding sites (RBS), suggesting that RBS strength may underlie gene expression capacity. Complementing these observations, simulations of in vivo gene expression demonstrated that specific transcript sequences (for example, rare codons) only reduced translation for designs with a strong RBS. As such, the authors used their model to optimize the forward design and validate constructs that maximized expression efficiency.

The impact of environmental factors on gene expression can also constrain microbial growth³¹. Another study measured the changes in activity levels of ~1,800 *E. coli* promoters across various environmental conditions, including varying carbon sources, temperature and osmotic stress. This approach revealed that a constant global scaling factor could uniquely describe ~70–90% of changes in gene expression for each condition, regardless of the specific promoter³². In other words, a given promoter's activity levels in any two different conditions were directly proportional to each other; moreover, the magnitude of that proportionality was preserved across the vast majority of promoters tested. Mechanistic models of different potential explanations quantitatively demonstrated that a global resource allocation strategy best fit the experimental data; this model assumed that, in each condition, the total promoter activity was allocated among condition-specific and globally expressed genes, thereby incorporating both the relative fraction of resources dedicated to each and the growth rate differences across conditions. This model was also translatable to *Saccharomyces cerevisiae*³², suggesting kingdom-level generality.

Integrated computational or experimental studies have since leveraged a fundamental understanding of resource allocation to more deeply probe the environment-specific rules governing transcription regulation^{33–35} and to identify the design principles for metabolically efficient gene circuits^{36–40}. For example, one study computationally and experimentally investigated the strategies that reduce indirect coupling between gene circuits within a cell, whereby the expression of one gene circuit alters the expression of a second independent one in the same cell⁴¹. The authors found that negative feedback motifs can control the relative resource utilization (and thus trade-off in expression levels), improving overall efficiency and suggesting a mechanism by which cells prioritize multiple energy-intensive processes. Similarly, a recent study showed that mRNA

upon those below; for example, without understanding complexity in gene expression it is impossible to fully appreciate microbial community dynamics. Below, we highlight how quantitatively teasing apart the various compounding factors at each progressive step of model complexity (from transcription to community-level dynamics) has improved our fundamental ability to predict, understand and design increasingly intricate biological systems. As we will see, in many cases, the goal of modelling is to quantitatively characterize how these factors and their associated processes impact cellular fitness and/or microbial growth dynamics.

Transcription. At the most basic level, biological networks are built on patterns of gene expression, and it is important to recognize that each individual component of a gene network is inherently restricted to the same pool

Continuous dynamical models

Changes in the state variable(s) occur uninterrupted (for example, continuously) over an arbitrary time window.

Ordinary differential equations

(ODEs). A set of equations describing the relationship between the derivative of one or more dependent variables with respect to one independent variable.

degradation targeted to MazF (a sequence-dependent endoribonuclease) improved gene circuit functionality by funneling cellular resources towards circuit expression⁴⁰. Specifically, MazF activity degrades mRNA transcripts based on recognition sites present in most *E. coli* genes. Therefore, exogenously expressing MazF freed resources typically used by native *E. coli* transcription, which led to increased expression of the authors' gene circuit of interest.

These generalized design principles can also be used to inform our understanding of existing regulatory circuits. For example, a follow-up study investigating *mazEF* dynamics demonstrated that quantitatively accounting for MazF-mediated autoregulated mRNA degradation was necessary to describe the observed heterogeneity in cell length, emphasizing the role of this autoregulatory motif in response to stress caused by MazF overexpression³⁵. Similarly, incorporating gene expression capacity was necessary to accurately simulate the pattern formation that was observed in genetically reprogrammed *E. coli* — these results provided fundamental insights into the mechanistic underpinnings of scale invariance in developmental systems⁴².

Translation. The confounding interactions inherent in gene expression discussed above are often further magnified at the protein level, where there is an inherent relationship between synthesis and cellular proliferation. Proteins constitute ~50% of cellular dry mass⁴³, and protein maintenance (for example, repair, turnover, and so on) is the dominant energy-consuming component of maintenance metabolism⁴⁴. Thus, as with gene expression, the metabolic burden associated with translation also influences cellular fitness²⁸. As the cost associated with protein production is readily confounded by the benefit derived from its function, these questions are particularly well-suited for mechanistic or coarse-grained models. Indeed, one study used ribosome profiling⁴⁵ to quantify absolute protein synthesis rates and revealed fundamental cellular control strategies whereby cells optimize protein production to maximize growth efficiency⁴⁶. The authors computationally investigated whether increasing the production of and benefit due to MetE, the limiting step in L-methionine biosynthesis, would improve overall growth despite the cost (due to competition for ribosomes). Remarkably, the model predicted that optimal growth occurred for MetE parameters that closely matched the experimental measurements, demonstrating that cells tune MetE production to maximize biomass⁴⁶. Indeed, the authors experimentally demonstrated that both an increase or a decrease in MetE production from baseline reduced the growth rate.

In addition to the cost associated with production, the benefits of expressing a protein can be directly measured by its catabolic activity, so long as they can be measured separately. To quantify the cost of protein activity, one study established a clever experimental framework that decoupled lactose production from degradation; using a quantitative model for fitness, the authors showed that the cost of lactose permease activity, rather than its production cost, specifically accounted

for the increased burden on the cell⁴⁷. Along these lines, another study similarly avoided confounding growth effects by using a cell-free system to estimate translation efficiencies, and then incorporated these estimates into a growth-burden model to accurately predict efficient construct designs⁴⁸.

Understanding how cells achieve robust protein homeostasis across diverse conditions, and how this translates to overall fitness, has been of particular interest to microbiologists⁴⁹. For example, bacterial cells have extensive protective chaperone networks that maintain correct protein folding while eliminating misfolded or non-functional ones⁵⁰. The ability to model these functional capabilities would be impactful from both a basic biology and a translational perspective: reliably modulating these networks could give us more fine-grain control of protein expression in synthetic systems⁵¹, whereas promoting misfolding may represent an alternative strategy⁵². To this end, quantitative analyses and modelling of resource allocation at the proteome level has revealed the various control strategies that maximize efficient growth^{53–55}. For example, a study showed that, during slow growth, the number of active ribosomes declined, rather than the translation rate as was previously assumed⁵⁶. These principles are particularly useful for predicting or optimizing growth, and other phenotypes, under dynamically changing or harsh environmental conditions.

This approach has recently been extended to other species, including *S. cerevisiae*; multiple studies^{57,58} have demonstrated that models based on proteome constraint and allocation can account for changes in the Crabtree effect, which is reflective of substrate-dependent metabolic dynamics. These and related works use flux-constrained models to predict metabolic phenotypes and intervention strategies^{59,60}, although such models primarily rely on genome-scale frameworks, and are therefore outside the scope of this Review, it is nonetheless crucial to note that these reduced approaches may be particularly useful when kinetic rates and other metabolic parameters (for example, half-maximal constants) are difficult to experimentally estimate⁶¹. Interestingly, one study used a stochastic model to demonstrate that coupling transcription rates, rather than translation rates, to cell division could effectively smooth out protein noise and stabilize protein levels in rapidly dividing cells⁶². These results highlight the potential for multi-level gene circuit optimization in the context of synthetic biology — not only to achieve a certain functionality, but also to simultaneously modulate global host behaviour.

The above studies indicate that building models based on fundamental control strategies, rather than on specific rate parameters, ensures that the models can retain predictive value in a wide range of physiological settings. These strategies have been used in a variety of systems and synthetic biology contexts, including optimizing the biosynthesis of a desired molecule using metabolic pathway analysis⁶³, and designing synthetic gene circuits that rely on host cell metabolism in addition to transcriptional control to achieve the desired functionality⁶⁴. In both of these examples, the authors reported

Deterministic

A type of process whereby the future state of a system involves no randomness and depends entirely on its initial state along with the parameters that govern its change.

Stochastic

A type of process whereby the future state of a system involves a certain degree of randomness.

Dynamical models

Mathematical models that describe the change in one or more variables of interest over time.

Deep learning

A class of artificial intelligence frameworks that are able to learn in an unsupervised manner, typically using structured layers of artificial neural networks.

Quorum sensing

(QS). A control strategy wherein individual cells secrete a signal molecule into their environment; the collective concentration serves as a proxy for local cell density and enables cells to trigger downstream gene expression based on population size.

Crabtree effect

The phenomenon wherein yeast produce ethanol rather than biomass under high-glucose aerobic conditions.

Chemostat

A type of bioreactor that uses inflow and outflow of new or spent media to continuously culture microorganisms at a specified growth rate under chemically defined conditions.

predictable, improved dynamics as a result of multi-level modelling. These efforts represent our increasing ability to leverage native bacterial networks to achieve both outside control and engineered design; increasingly integrated models will likely proliferate as our fundamental understanding progresses.

Metabolic processes. Metabolic activity accounts for both growth (for example, biomass production) and non-growth maintenance functions (for example, stress response, osmotic regulation, and so on), and thus efforts to describe metabolic-specific effects are readily confounded by growth dynamics⁶⁵. Indeed, the classical Monod model for cellular growth on a single substrate⁶⁶ ideally describes dynamics in growth-permissive environments, and is applicable to a wide range of theoretical and experimental studies. However, in the face of nutrient limitation or other inhibitory environments, where metabolism is devoted to survival rather than growth, additional complexity is often required to explain population behaviour; this is particularly relevant when bacteria are subject to various stressors or rapidly changing conditions.

Several decades of carefully designed chemostat experiments has facilitated the quantification of this maintenance metabolism^{67–69}, and incorporating these endogenous metabolic rates into growth models has enabled us to tease apart metabolic-specific dependencies^{70,71}. For example, although antibiotic efficacy increases with bacterial growth rate in a linear manner, we developed and experimentally validated a mathematical model to show that, when growth and metabolism are uncoupled, antibiotic lethality depends on the bacterial metabolic state at the time of treatment rather than on the growth rate⁷². Our work showed that the metabolic response following the initial drug-target interaction drives many aspects of the bacterial response to antibiotic exposure and revealed a metabolic threshold below which antibiotic lethality is negligible. These results suggest that metabolism may be a potential target for future antibacterial strategies, for example, novel adjuvant compounds that modulate cellular activity to make them more susceptible to co-administered antibiotics.

Studies that integrate metabolic flux analyses with substrate kinetics have provided powerful evolutionary insights into laws governing microbial growth and highlight the value of mixed computational models^{73–76}. Indeed, cells grown on switched carbon and nitrogen mixtures⁷⁷ or on switched carbon sources⁷⁸ exhibit specific global transcriptional profiles that can be explained by general control strategies^{79,80}. These findings and analyses provide insights into the temporal dynamics and regulatory plasticity that underlie metabolic adaptation to various environmental stimuli. For example, one study used quantitative estimates of growth rates and promoter activities to establish a hierarchy of non-glucose sugar utilization that can quantitatively predict the growth dynamics on mixed carbon sources⁷⁴. The authors observed that, for some combinations of sugars, simultaneous rather than sequential activation for each system occurred. This could not be explained by a simplified

model that optimized fitness as a function of growth, suggesting that multi-modal fitness may underlie dual sugar utilization. These findings could inform future strategies to control microbial growth in laboratory or natural settings, such as by modulating the exogenous sugar composition to enforce a desired predetermined growth profile.

Insights into the mechanisms that govern metabolic strategies have implications for implementing synthetic biology and metabolic engineering applications with optimized objective functions (for example, efficient use of limited resources) as well as on understanding the interplay between a specific environment and its cognate microbial community. Indeed, metabolism has a major role in defining and shaping the interaction networks of microbial populations in natural environments such as the gut microbiome^{81,82}. For example, although isogenic populations can efficiently perform simple biosynthetic functions, increasingly complex pathways can become burdensome to microorganisms, leading to disadvantages, including a reduced culture density or product yield⁸³ as well as selection for loss-of-function mutations⁸⁴. Interestingly, Morris et al. proposed a related idea, termed the ‘Black Queen Hypothesis’, which suggests that genetic variants with superfluous or burdensome metabolism could be selected against, resulting in reduced genetic diversity but also in more efficient metabolism at a population level⁸⁵.

Alternatively, cell populations can distribute this metabolic burden, wherein distinct subpopulations are responsible for complementary metabolic tasks. To this end, recent studies have used both mechanistic⁸⁶ and genome-scale⁸⁷ models to describe the dynamics of pathway enzymes as well as the host burden; these models elucidate particular population structures capable of maintaining division of labour. For example, through the analysis of 24 common metabolic pathway motifs, a study derived quantitative criteria that define the parameter regimes, wherein a division of labour approach is advantageous to maintaining the entire pathway within a single cell⁸⁶. Such insights can be used to understand how metabolic burden influences cross-species cooperation and joint evolution in natural settings.

Microbial community-level dynamics: ecology and evolution. We have thus far discussed biological complexity largely in the context of individual microorganisms. However, as alluded to at the end of the previous section, it is the collective effect of these individual cells, and their underlying dynamics, that dictate the emergent population-level and community-level behaviours. A typical population often encompasses heterogeneity at both the biochemical and genetic levels^{88,89}. This variability can lead to emergent, diverse ecological and evolutionary dynamics, including competition^{90,91}, protection from stressors^{92,93}, community stability^{94,95}, cooperation in structured environments^{96,97}, altruistic behaviours⁹⁸ and horizontal gene transfer⁹⁹, among others. The full breadth and depth of microbial ecology and evolution are beyond the scope of this Review. Instead, in this section, we discuss heterogeneity in bacterial populations and in communities of increasing complexity,

highlighting select cases that exemplify how modelling such effects has revealed fundamental insights into microbial behaviour.

As noted above, even clonal populations can exhibit heterogeneity due to intrinsic biochemical noise^{100,101}, which may result in complex population-level dynamics. These effects, and their consequences, have been studied in great depth, particularly enabled by stochastic models of evolution and gene expression¹⁰²⁻¹⁰⁴. These models have demonstrated that noise-induced fluctuations can propagate through biological networks, dictating behaviour¹⁰⁵⁻¹⁰⁷. Integrating such approaches with feedback loops and other amplification mechanisms has resulted in microbial models that explain unexpected experimental observations in growth dynamics, including increases in overall fitness¹⁰⁸ or transient oscillations in cell size¹⁰³. Biochemical heterogeneity can also drastically impact the immediate bacterial response to an ecological disturbance. For example, heterogeneous gene expression levels can provide transient population-level tolerance to environmental stressors, including antibiotic treatment¹⁰⁹⁻¹¹² (we note that the topic of bacterial persisters and the biochemical or metabolic components of antibiotic tolerance have been extensively reviewed elsewhere¹¹³⁻¹¹⁵). In some cases, this noise can be amplified by specific gene networks as a form of nonlinear decision-making. For example, several recent studies have demonstrated that sporulation (a dormant state akin to persisters) in *Bacillus subtilis* is driven by an ultrasensitive positive feedback network ensuring that the molecular machinery is synthesized in a 'just-in-time' manner^{116,117}. Other work has shown that integrating sporulation models with stochastic DNA competence results in a transient window of highly variable DNA uptake, generating further heterogeneity that may modulate colony organization¹¹⁸.

Compared to models of biochemical noise, understanding the dynamics of genetically heterogeneous populations or communities is a relatively new and underappreciated challenge; indeed, the advent of metagenomics has only recently revealed the diversity and complexity of microbial communities in situ. Nonetheless, modelling has already proven to be a particularly effective tool in deducing the rules governing microbial community assembly, functionality and long-term dynamics in the face of this variability. Different modelling frameworks have been used to derive multi-species assembly principles based on smaller sets of species interactions¹¹⁹⁻¹²², representing the first step in predicting complex microbial population dynamics. For example, one study used the temporal dynamics of monoculture and paired species to train a computational model of a higher-order, multi-species synthetic gut community — this approach revealed specific ecological interaction networks that could explain coexistence within multi-species communities¹²⁰.

In addition to ecological factors, a community's long-term fate inherently depends on adaptations over time, both in response to other strains and to the local environment. These evolutionary dynamics can lead to unexpected emergent properties; for example, species

evolved within a community were recently shown to exhibit an overall higher productivity compared to those evolved individually and assembled into the same compositions¹²³. Diverse models that integrate both ecological and evolutionary processes have revealed how this feedback constrains and/or promotes stable communities over time. Of note, one study showed that metabolic cooperation can stably emerge in plasmid-encoded auxotrophic yeast populations; in this case, modelling was used to demonstrate that stochastic plasmid loss could not account for the emergent behaviours¹²⁴. A different study used a mechanistic model to demonstrate that feedback loops between species in tri-culture (that is, three individual species) could predictably stabilize emergent social cheating behaviours in a cooperative population¹²⁵. These and other investigations suggest that, although evolution is stochastic, it can be predicted with an appropriate knowledge of community architecture and environment. To this end, recent work derived simple coarse-grained statistical laws that could be used to describe community dynamics in lieu of kinetic modelling¹²⁶; such models could be harnessed to generate long-term predictions without requiring a fully informed mechanistic basis.

Given the myriad of interactions and sources of variability that contribute to microbial consortia assembly and stability, the integration of model-guided approaches and high-throughput experimental platforms¹²⁷, for example, droplet microfluidics¹²⁸, could help capture the full range of potential dynamics, thereby improving our predictive capabilities. Along these lines, a study demonstrated that a stochastic model of the microbial lactose uptake network was not only capable of predicting equilibrium population distributions but also recreated dynamic, pre-steady state, population structures¹²⁹. The authors also showed that relatively few noise parameters are required to effectively convert a deterministic model to its stochastic equivalent. These randomized models are equally applicable to synthetic gene circuits; a recent paper combined dynamic modelling with a high-throughput microfluidic platform to characterize a molecular switch governing cell motility in *B. subtilis*¹³⁰. Specifically, the authors built a gene circuit to implement a known two-protein antagonistic relationship; single-cell quantification confirmed their modelling predictions and suggested that a minimal motif was sufficient to capture population distributions, switch timing and multi-generational inheritance.

As natural selection acts at the level of phenotype, it is unsurprising that biochemical and genetic heterogeneity are inherently linked. Biochemical variability can influence the long-term evolutionary outcomes in microbial populations depending on the particular environment. For example, a study showed that stochastically arising mutants with extended 'lag times' (time to cell division) were selected for in the presence of cyclical antibiotic exposure¹³¹. The authors used a model to predict that lag times would be optimized to match antibiotic exposure durations; these predictions closely matched experimental measurements, providing crucial insights into evolutionary dynamics.

In the study noted above¹³¹, the microbial response to antibiotic treatment and the subsequent evolution of resistance mutations were incorporated into a single mathematical framework. This exemplifies one of the main challenges in predicting complex microbial community dynamics: given the relatively small size of microbial genomes as well as their rapid proliferation and mutation rates¹³², ecological interactions (for example, response to perturbation) often occur concurrently with evolutionary changes — it is therefore particularly challenging to decouple the specific mechanistic contributions from each process individually. Modelling has proven particularly useful in differentiating these aspects, providing fundamental insights into the dominant contributing factors from each process. For example, strong selection pressures — such as exogenous stressors or burdensome gene circuits — enrich for mutants on an expedited timescale. As an example of the former, model-guided experiments have been used to establish quantitative explanations underlying drug-specific evolutionary outcomes¹³³ and to direct the outcome of microbial populations¹³⁴. Similarly, in a synthetic setting, one study implemented a strategy

utilizing inter-strain predation to prevent mutant escapees in a population consisting of three strains expressing a burdensome synthetic lysis circuit. Briefly, the dominant strain was iteratively displaced by a stronger competitor without disrupting the long-term circuit functionality¹³⁵. Modelling was crucial in revealing the ecosystem dynamics among the three strains underlying this cyclic replacement, allowing the overall results to be experimentally validated.

Challenges and opportunities

Clearly, the complexity of both natural and synthetic microbial systems is compounded by multiple levels of biological control, biomolecular variability, and network-level and population-based interactions. A variety of modelling strategies have enabled us to achieve a degree of predictability and control over these systems. However, as experimental capabilities and the breadth of available data continue to increase, next-generation biological models will need to be designed in an increasingly thoughtful and optimized manner. Indeed, as discussed above, increasing model complexity does not necessarily lead to increased benefit or accuracy. This notion — appropriately abstracting a model to the suitable level — is a crucial step in taking full advantage of quantitative approaches. Rather than incorporating as much complexity as possible, models should be based on a clearly defined, biologically meaningful hypothesis, on an understanding of the potential confounding factors and on the surrounding biological context. In some cases, semi-mechanistic or coarse-grained frameworks may be better suited to establish general rules governing observed microbial behaviour^{76,136}.

As with software, architecture or product design, modelling should be considered a case of form following function — minimally viable, fit-for-purpose models are ideal in that they sufficiently describe the dynamics of interest while remaining intuitive and understandable. In practice, the trade-off between complexity and abstraction is often iteratively informed by experimental validation. Although one can imagine certain ‘rules of thumb’, for example, incorporating growth dynamics when studying a burdensome network or integrating noise effects when a topology suggests nonlinearity, it ultimately falls to the modeller to ascertain the necessary scale to reliably capture the biological phenomena of interest.

As enumerated in the above sections, predictive modelling has facilitated numerous advances in our biological understanding, demonstrating its utility. However, as the field moves forward, there remain a number of technical and conceptual challenges that currently restrict the scope and detail of our fundamental knowledge of complex biological systems. In many cases, mathematical modelling approaches can be immediately leveraged to begin addressing these shortcomings. In this section, we identify five primary challenges that we view to be key limitations in the field of predictive biology, with a particular emphasis on synthetic and systems biology (BOX 3). We discuss representative aspects and examples of each challenge and emphasize the ways in which incorporating modelling can be particularly beneficial.

Box 3 | Current challenges and proposed solutions for predictive biology

Complex dynamics in engineered and natural populations

- Systematic characterization of defined ecological modules tested in diverse environments
- Implement high-throughput experimental techniques to increase the parameter space of every isolate-by-environment interaction

Increasingly large data sets

- Minimize ‘fishing expeditions’, where appropriate, by validating insights from large quantitative data sets with specific testable hypotheses through mathematical models and controlled experiments
- Computational tools that facilitate integrating diverse levels of information into predictive models
- Centralized and accessible parameter reporting

Accurate parameter estimates

- Establish an appropriate level of abstractness with which to characterize a given system
- Match parameter definitions with experimental estimation
- Use conditions that more closely mimic the process of interest to improve model accuracy¹⁷¹
- Improved technical tools to standardize parameter estimates
- Efforts to use consistent terminology to facilitate cross-literature data compilation and review

Accounting for stochastic and deterministic evolution in complex populations

- Integration with machine learning and systematic analyses into evolutionary constraints
- Expand sequencing breadth and depth to explore evolutionary responses at lower detection limits

Translating in vitro predictions to in vivo outcomes

- Implement experimental conditions that are biologically relevant
- Inclusion of animal models and other experimental platforms that can better simulate natural environments of interest
- Incorporate physiological conditions into modelling analysis¹⁷²
- Advances in culturing techniques
- Establish in situ quantification methods

As with the preceding discussion of complexity, these challenges often build upon and contribute to one another. Furthermore, we draw particular attention to the growing use of advanced computational techniques, including machine learning and deep learning. As mentioned in the Introduction, the integration of more classical modelling with these AI-based approaches represents exciting new avenues for quantitative biology but must be utilized in a rational, informed manner. Indeed, it is our view that blended strategies that emphasize the respective advantages of dynamical systems modelling and next-generation machine learning represent the future of predictive biology.

Predicting complex gene network dynamics. Even after decades of advances, accurately predicting bacterial cellular dynamics as dictated by underlying gene regulatory networks (whether natural or engineered) remains a significant challenge, and thus we discuss it first here. As gene circuits grow increasingly complex, it becomes commensurately more difficult to predict their dynamics due to intracellular confounding factors as well as the increased sensitivity to environmental conditions. These difficulties are further compounded with increasing heterogeneity, even over short time spans where evolution is not a significant consideration. Despite these challenges, advances in gene circuit engineering, in combination with modelling approaches, provide a blueprint for the future of this field.

State-of-the-art synthetic gene circuits have reached unprecedented levels of precision and complexity, for example, functional integration with biomaterials¹³⁷, in part through systems modelling, which has rapidly streamlined the design and implementation of multifunctional and layered designs^{138,139}. This approach is particularly advantageous in cases where the dynamics are not immediately obvious from the network topology. For example, recent quantitative and experimental integration of the toggle switch and repressilator circuits revealed parameter regions that resulted in emergent novel behaviours, including the control of oscillation coherence¹³⁹. Efforts such as these will likely be aided by improved experimental techniques; to this end, a recent study utilized single-cell microfluidics to examine the variability inherent in a population of cells expressing a mutagenic library of repressilator circuits¹⁴⁰. Although the original circuit did not exhibit oscillatory behaviour in the absence of exogenous inducers, the authors were able to reliably isolate multiple variant lineages capable of doing so.

Moving forward, achieving well-defined behaviour will require optimizing the biophysical constraints of the circuits themselves¹⁴¹, minimizing unwanted secondary effects (for example, resource burdens or noise)^{142,143} and expanding the accompanying models to account for inter-population interactions. This is becoming increasingly possible with the availability of well-characterized genetic parts^{144,145} along with rapid cell-free assembly methods¹⁴⁶, yielding highly complex and layered circuitry^{139,147,148}. Continued mining and characterization of biomolecular components as well as establishing methods to report on the optimal assembly configurations

that minimize efficiency constraints will ensure that models remain intuitive and informative, even as microbial engineering grows in scope.

Beyond the components of gene circuits themselves, libraries of well-characterized microbial communities with defined sensory and actuary functions (for example, population level analogues of gene circuits) could allow modular community assembly much like their gene circuit predecessors¹⁴⁹. Along these lines, a recent study proposed a number of environmental remediation strategies wherein synthetically engineered populations could potentially interact with natural strains in situ to realize population-level behaviours (for example, mutualism or competition), thereby maintaining ecosystems and improving carbon sequestration¹⁵⁰. Despite these and other appealing possibilities, considerable work is needed to systematically define these ecological units (for example, isolate, gene circuit, environment or some combination) and implement their systematic characterization.

Large-scale data inference and parametrization. Ultimately, the ability to predict population dynamics hinges on uncovering biologically meaningful insights from increasingly large data sets. Although an increased accessibility to large computing infrastructure has given us the power to begin interrogating these data sets, with this capability comes significant challenges. One of these challenges, and the second we raise, is that the sheer size of data sets generated by next-generation and high-throughput technologies can lead to unforeseen covariates or factors that obscure the distinction between biological and statistical significance. Machine learning approaches are well suited to addressing large data sets, and previous studies have used these to integrate and search through extensive types and amounts of data to extract specific features of interest such as identifying network architectures^{151,152}. However, machine learning models are prone to overfitting, bias and non-explainability, leading to model structures that are difficult to verify in vitro or that lack predictive value.

Context-appropriate model parameterization and abstraction. Consequently, the third crucial, and related, challenge is how to effectively extract accurate, biologically relevant parameters from large data sets. There are a number of ways to integrate machine learning and AI with dynamical modelling approaches in order to improve the reliability of parameter estimates moving forward (BOX 4). Perhaps most achievable among these is the use of algorithmically generated insights to abstract and simplify modelling approaches. For example, biological relationships uncovered using AI approaches could be used to simplify subsequent (semi-)mechanistic models, ideally reducing the number of parameters needed and/or the difficulty in obtaining reliable values. Indeed, one study used a support vector machine approach to derive a simplified criterion that predicts the outcome (coexistence or collapse) of a generic mutualistic population without explicit characterization of the underlying mutualistic interaction(s)¹⁵³. Although such characterization would be prohibitive due to the diversity of

Support vector machine
A trained machine-learning methodology that uses classification algorithms to separate data into multiple groupings.

Box 4 | Integrating dynamical modelling and ML algorithms

Integration of artificial intelligence (AI) or machine learning (ML) with dynamical systems modelling can be summarized into three general categories, as follows:

Parameterization of known model structure

AI and ML can leverage existing relationships to predict new parameter values at scale. Most often, a parameter of interest will be proxied by a correlator or correlators (often derived from publicly available, pre-defined or easily measurable values) with known relationships; predicted parameters can then be validated using existing models, thereby informing downstream biological hypotheses. Model accuracy and predictive power can be tuned by adding or changing correlators. For instance, enzyme turnover (k_{cat}) correlates with key features such as reaction flux, structural properties and biochemical mechanisms; a recent study showed that ML accurately inferred k_{cat} values that were used to improve whole-cell network model predictions¹⁷³.

Model structure inference

A growing application of AI and ML is the inference of network topologies using traditional Monte Carlo simulation data or, more recently, graph-based methods. Network inference can identify clusters within input data based on a training set of known relations and can combine component relationships to construct an output model framework. For example, gene expression profiles can be used as robust input sets to identify regulatory networks, particularly since many regulatory elements are well defined and thus serve as ideal training data¹⁷⁴. In this application, the output can inform a model for downstream predictions, which should be mechanistically validated to determine relevance.

Generation of novel biological insights

In many cases, AI and ML can be used to detect and/or highlight trends in data sets that are not immediately obvious. In such cases, these trends can and should be used to generate more specific hypotheses for subsequent model development and testing; at the same time, it is important to ensure that outputs of AI or ML models are interpreted through a biological lens, avoiding 'black-box' approaches. As such, the choice of algorithm, cost objective, and so on, can be constrained by incorporating established models; this both ensures that fundamental biological principles are not violated and may reveal mechanistically informative information. For example, one study integrated ML with metabolic network modelling to uncover novel metabolism-related mechanisms of action for bactericidal antibiotics¹⁶⁵. This white-box approach was achieved by training a model to predict changes in drug sensitivity (output) based on simulated metabolic states (input). As both the input and output data consisted of measurements across a panel of the same metabolites, doing so mechanistically linked specific metabolic pathways and heightened antibiotic lethality, which the authors validated experimentally.

mutualistic interactions in nature, the abstracted criterion proved predictive with both simulated and experimental data.

This simplification method could also be applied in the opposite direction — mechanistic models may be used to constrain the interaction space explored by automated algorithms. The importance of precise parameter estimates is particularly apparent for models of mixed microbial communities that must account for both ecological and evolutionary pressures. In these models, teasing apart and incorporating only the relevant attributes is essential for the derivation of useful and translatable insights. For example, although several studies suggest that antibiotics promote horizontal gene transfer by conjugation^{154,155}, a recent study showed, using precise quantitative estimates of plasmid transfer rates in the absence of growth, that antibiotics had a negligible effect on conjugation efficiencies. Instead, a simplified three-population model demonstrated that selection dynamics — an ecological outcome — sufficiently explains the observed antibiotic-associated promotion of conjugation^{156–158}. In conjunction with appropriately abstracted models, employing innovative experimental

methods, for example, metagenomic quantification for growth rate estimates¹⁵⁹, to assess difficult-to-measure parameters will undoubtedly improve biological relevance by eliminating uninformative covariates. By grounding next-generation data sets in concrete, experimentally tractable, semi-mechanistic or fully mechanistic underpinnings, we can help direct studies towards the efficient use of resources and effort.

Predicting evolutionary fates in complex populations.

Even when individual networks or populations are well-characterized in isolation, it is still a major challenge, and goal, to predict (and possibly direct) the long-term evolutionary outcomes, particularly in mixed communities. Recently, model-guided experiments have been leveraged to understand the evolutionary trajectories (for example, drug-specific outcomes)¹³³ and identify the feasible genetic^{134,135,160,161} or biochemical^{162–165} strategies to reduce the evolution of resistant mutants. For example, it has been shown that sequential antibiotic regimens that take advantage of collateral sensitivity networks can maximize treatment efficacy; these approaches took advantage of model-driven exploration of the parameter space to isolate the regions that maximize this favourable outcome^{163,166}. Combining phenotypic measures of interest with advances in machine learning will likely reveal mechanistic explanations that have specific genetic underpinnings. For example, a recent study showed that only 500 diverse *Salmonella* spp. genomes were sufficient to train a machine learning model to accurately predict the minimum inhibitory concentration of 15 antibiotics in a set of 5,278 genomes¹⁶⁷. This type of approach can likely reveal genetic patterns that are otherwise difficult to identify¹⁶⁸ and, in combination with predictive evolutionary models or experimental validation, provide crucial insights into the biological processes that drive the evolution of clinical isolates. Such approaches could be readily extended to other areas of interest, including host adaptations by microorganisms and metabolic engineering.

Turning in vitro observations into in vivo insights. The final challenge we raise is one of translation: how do we best take observations and conclusions from in vitro laboratory settings and apply them to microbial communities arising in in vivo environmental and biomedical contexts? Efforts to increase the relevance of experimental conditions to natural settings, either by moving towards in vivo models or through better simulating the environment of interest by integrating chemical and physical constraints, could be better integrated in a complementary fashion with mathematical model development that aims to incorporate additional biological complexities. A lack of translatability is also often evident in the other direction, that is, recreating natural systems in a laboratory environment. For example, the majority of microorganisms found in nature are unculturable and are often poorly described by laboratory microbial strains; advances in culturing techniques and/or in situ measurements are needed to build accurate predictive models. This represents another arena in which machine learning and other advanced computational

modelling techniques may prove useful. Indeed, a recent paper used a genome-scale model to design growth media for the bacterium *Akkermansia muciniphila* based on a predicted enzyme deficiency; the optimized model was capable of predicting growth dynamics in the tailored environment¹⁶⁹. These approaches may one day be capable of rendering currently unculturable strains amenable to in vitro manipulation.

Conclusions

Dynamic models have led to numerous advances in our collective understanding of biological behaviours, particularly in microbial populations. These models allow us to interrogate biological complexity on multiple levels, ranging from gene expression to evolutionary outcomes. Despite this flexibility, there remain a number of challenges that limit the impact of predictive biology. Undoubtedly, addressing these challenges will require parallel multi-interdisciplinary efforts to push forward both experimental and computational techniques in microbiology beyond the current state-of-the-art. Nonetheless, the utility of dynamical systems modelling in microbiology has proven invaluable over the past two decades.

The growing prevalence of big data and machine learning approaches offer a newer-age alternative to the classic systems approach. Integrating these advances to establish a next-generation paradigm of predictive biology will undoubtedly yield meaningful returns (FIG. 2). Indeed, several recent studies demonstrate that non-canonical uses of machine learning that leverage the utility of computational modelling can provide powerful quantitative insights, including generating coarse-grained predictions for ecological interactions¹⁵³, improving computational efficiency to accelerate model predictions¹⁷⁰, and elucidating causal mechanistic relationships between drug perturbation and cellular response¹⁶⁵. For example, recent work used a ‘white-box’ machine learning approach that integrated genome-scale network modelling with antibiotic half maximal inhibitory concentration data; the authors utilized a

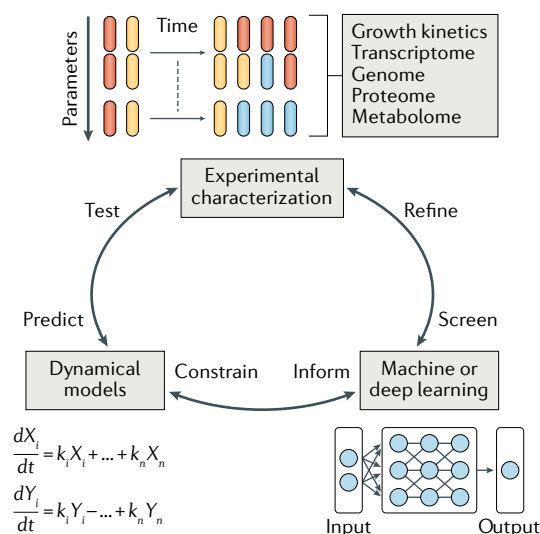


Fig. 2 | A next-generation paradigm of predictive biology. The classical systems approach is defined by the continuous feedback between dynamical modelling and experimental validation. As machine learning and deep learning algorithms become more readily available, a new paradigm in predictive biology will increasingly combine all three complementary approaches, enabled by the breadth of experimental data accessible with next-generation, high-throughput technologies.

perturbation response methodology to tie machine learning predictions to experimental data, which identified nucleotide biosynthesis as a key metabolic pathway contributing to antibiotic lethality¹⁶⁵. These types of integrative approaches encompass the best of both worlds: on the one hand, modelling provides concrete mechanistic insights into microbial systems and parameter spaces that are otherwise difficult (if not impossible) to explore; on the other, the data encompasses sufficient biological information such that, using analytical tools, we can sort through the noise to find the biological meaning.

Published online 29 May 2020

- Blattner, F. R. et al. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462 (1997).
- Ghatak, S., King, Z. A., Sastry, A. & Palsson, B. O. The y-ome defines the 35% of *Escherichia coli* genes that lack experimental evidence of function. *Nucleic Acids Res.* **47**, 2446–2454 (2019).
- Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
Along with the repressilator (reference 7), this study was the first implementation of synthetic biology circuits in living cells.
- Balázsi, G., Van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910–925 (2011).
- Kobayashi, H. et al. Programmable cells: interfacing natural and engineered gene networks. *Proc. Natl Acad. Sci. USA* **101**, 8414–8419 (2004).
- Wu, M. et al. Engineering of regulated stochastic cell fate determination. *Proc. Natl Acad. Sci. USA* **110**, 10610–10615 (2013).
- Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
Along with the toggle switch (reference 3), this study was the first implementation of synthetic biology circuits in living cells.
- Palsson, B. & Zengler, K. The challenges of integrating multi-omic data sets. *Nat. Chem. Biol.* **6**, 787–789 (2010).
- Stein, R. R. et al. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput. Biol.* **9**, e1003388 (2013).
- Mannan, A. A. et al. Integrating kinetic model of *E. coli* with genome scale metabolic fluxes overcomes its open system problem and reveals bistability in central metabolism. *PLoS One* **10**, e0139507 (2015).
- Ocone, A., Haghighverdi, L., Mueller, N. S. & Theis, F. J. Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. *Bioinformatics* **31**, i89–i96 (2015).
- Ellis, T., Wang, X. & Collins, J. J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* **27**, 465–471 (2009).
- Tan, C., Marguet, P. & You, L. Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol.* **5**, 842–848 (2009).
- Smolen, P., Baxter, D. A. & Byrne, J. H. Mathematical modeling of gene networks. *Neuron* **26**, 567–580 (2000).
- Ferrell, J. E. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* **14**, 140–148 (2002).
- Marguet, P., Tanouchi, Y., Spitz, E., Smith, C. & You, L. Oscillations by minimal bacterial suicide circuits reveal hidden facets of host-circuit physiology. *PLoS One* **5**, e11909 (2010).
- Veliz-Cuba, A. et al. Sources of variability in a synthetic gene oscillator. *PLoS Comput. Biol.* **11**, e1004674 (2015).
- Lugagne, J. B. et al. Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nat. Commun.* **8**, 1671 (2017).
- Wu, F., Menn, D. J. & Wang, X. Quorum-sensing crosstalk-driven synthetic circuits: From Unimodality to trimodality. *Chem. Biol.* **21**, 1629–1638 (2014).
- Melke, P., Sahlin, P., Levchenko, A. & Jönsson, H. A cell-based model for quorum sensing in heterogeneous bacterial colonies. *PLoS Comput. Biol.* **6**, 1–13 (2010).
- De Gelder, L. et al. Combining mathematical models and statistical methods to understand and predict the dynamics of antibiotic-sensitive mutants in a population of resistant bacteria during experimental evolution. *Genetics* **168**, 1131–1144 (2004).
- Didelot, X. & Maiden, M. C. J. Impact of recombination on bacterial evolution. *Trends Microbiol.* **18**, 315–322 (2010).
- Shapiro, B. J. et al. Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**, 48–51 (2012).

24. Chalancon, G. et al. Interplay between gene expression noise and regulatory network architecture. *Trends Genet.* **28**, 221–232 (2012).
25. Hooshangi, S., Thiberge, S. & Weiss, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl Acad. Sci. USA* **102**, 3581–3586 (2005).
26. Vind, J., Sørensen, M. A., Rasmussen, M. D. & Pedersen, S. Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes. *J. Mol. Biol.* **231**, 678–688 (1993).
27. Gorochowski, T. E., Avciar-Kucukgoze, I., Bovenberg, R. A. L., Roubos, J. A. & Ignatova, Z. A minimal model of ribosome allocation dynamics captures trade-offs in expression between endogenous and synthetic genes. *ACS Synth. Biol.* **5**, 710–720 (2016).
28. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**, 1099–1103 (2010).
29. Carrera, J., Rodrigo, G., Singh, V., Kirov, B. & Jaramillo, A. Empirical model and in vivo characterization of the bacterial response to synthetic gene expression show that ribosome allocation limits growth rate. *Biotechnol. J.* **6**, 773–783 (2011).
30. Ceroni, F., Algar, R., Stan, G. B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* **12**, 415–418 (2015).
31. Acar, M., Mettetal, J. T. & Van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* **40**, 471–475 (2008).
32. Keren, L. et al. Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.* **9**, 701 (2013).
Measured changes in activity levels of 1,800 *E. coli* promoters and demonstrated that a given promoter's activity levels in any two different conditions were directly proportional to one another.
33. Kochanowski, K. et al. Few regulatory metabolites coordinate expression of central metabolic genes in *Escherichia coli*. *Mol. Syst. Biol.* **13**, 903 (2017).
34. Fang, X. et al. Global transcriptional regulatory network for *Escherichia coli* robustly connects gene expression to transcription factor activities. *Proc. Natl Acad. Sci. USA* **114**, 10286–10291 (2017).
35. Nikolic, N. et al. Autoregulation of mazEF expression underlies growth heterogeneity in bacterial populations. *Nucleic Acids Res.* **46**, 2918–2931 (2018).
36. Carbonell-Ballesterro, M., Garcia-Ramallo, E., Montañez, R., Rodriguez-Caso, C. & Macia, J. Dealing with the genetic load in bacterial synthetic biology circuits: convergences with the Ohm's law. *Nucleic Acids Res.* **44**, 496–507 (2016).
37. Darlington, A. P. S., Kim, J., Jiménez, J. I. & Bates, D. G. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. *Nat. Commun.* **9**, 695 (2018).
38. Qian, Y., Huang, H. H., Jiménez, J. I. & Del Vecchio, D. Resource competition shapes the response of genetic circuits. *ACS Synth. Biol.* **6**, 1263–1272 (2017).
39. Gyorgy, A. et al. Isocost lines describe the cellular economy of genetic circuits. *Biophys. J.* **109**, 639–646 (2015).
40. Venturelli, O. S. et al. Programming mRNA decay to modulate synthetic circuit resource allocation. *Nat. Commun.* **8**, 15128 (2017).
41. Shopera, T., He, L., Oyetunde, T., Tang, Y. J. & Moon, T. S. Decoupling resource-coupled gene expression in living cells. *ACS Synth. Biol.* **6**, 1596–1604 (2017).
Computationally and experimentally identified strategies that reduce indirect coupling between gene circuits within a cell.
42. Cao, Y. et al. Collective space-sensing coordinates pattern scaling in engineered bacteria. *Cell* **165**, 620–630 (2016).
43. Russell, J. B. & Cook, G. M. Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol. Rev.* **59**, 48–62 (1995).
44. Kempes, C. P. et al. Drivers of bacterial maintenance and minimal energy requirements. *Front. Microbiol.* **8**, 31 (2017).
45. Brandman, O. et al. A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* **151**, 1042–1054 (2012).
46. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635 (2014).
Quantification of absolute protein synthesis rates, including L-methionine biosynthesis, revealed that cells optimize protein production to maximize growth efficiency.
47. Eames, M. & Kortemme, T. Cost-benefit tradeoffs in engineered lac operons. *Science* **336**, 911–915 (2012).
48. Borkowski, O. et al. Cell-free prediction of protein expression costs for growing cells. *Nat. Commun.* **9**, 1457 (2018).
Incorporated cell-free estimates of translation efficiencies into models accounting for growth burden to predict efficient construct designs.
49. Basan, M. et al. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature* **528**, 99–104 (2015).
50. Macario, A. J. L. & Conway de Macario, E. Molecular chaperones: multiple functions, pathologies, and potential applications. *Front. Biosci.* **12**, 2588–2600 (2007).
51. Chang, L. et al. High-throughput screen for small molecules that modulate the ATPase activity of the molecular chaperone DnaK. *Anal. Biochem.* **372**, 167–176 (2008).
52. Kragol, G. et al. The antibacterial peptide pyrrolicin inhibits the ATPase actions of DnaK and prevents chaperone-assisted protein folding. *Biochemistry* **40**, 3016–3026 (2001).
53. Klumpp, S., Scott, M., Pedersen, S. & Hwa, T. Molecular crowding limits translation and cell growth. *Proc. Natl Acad. Sci. USA* **110**, 16754–16759 (2013).
54. Kafri, M., Metz-Raz, E., Jona, G. & Barkai, N. The cost of protein production. *Cell Rep.* **14**, 22–31 (2016).
55. Gerosa, L., Kochanowski, K., Heinemann, M. & Sauer, U. Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol. Syst. Biol.* **9**, 658 (2013).
56. Dai, X. et al. Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth. *Nat. Microbiol.* **2**, 16231 (2016).
57. Dai, Z., Huang, M., Chen, Y., Siewers, V. & Nielsen, J. Global rewiring of cellular metabolism renders *Saccharomyces cerevisiae* Crabtree negative. *Nat. Commun.* **9**, 3059 (2018).
58. Martínez, J. L., Bordel, S., Hong, K. K. & Nielsen, J. Ccn4p and the Crabtree effect of yeast: drawing the causal model of the Crabtree effect in *Saccharomyces cerevisiae* and explaining evolutionary trade-offs of adaptation to galactose through systems biology. *FEMS Yeast Res.* **14**, 654–662 (2014).
59. Tokic, M., Hatzimanikatis, V. & Miskovic, L. Large-scale kinetic metabolic models of *Pseudomonas putida* KT2440 for consistent design of metabolic engineering strategies. *Biotechnol. Biofuels* **13**, 33 (2020).
60. Salvy, P. & Hatzimanikatis, V. The ETRFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nat. Commun.* **11**, 30 (2020).
61. Vasilakou, E. et al. Current state and challenges for dynamic metabolic modeling. *Curr. Opin. Microbiol.* **33**, 97–104 (2016).
62. Bertaux, D., Marguerat, S. & Shahrezaei, V. Division rate, cell size and proteome allocation: Impact on gene expression noise and implications for the dynamics of genetic circuits. *R. Soc. Open Sci.* **5**, 172234 (2018).
63. Patanè, A., Santoro, A., Costanza, J., Carapezza, G. & Nicosia, G. Pareto optimal design for synthetic biology. *IEEE Trans. Biomed. Circuits Syst.* **9**, 555–571 (2015).
64. Fung, E. et al. A synthetic gene-metabolic oscillator. *Nature* **435**, 118–122 (2005).
65. Lipson, D. A. The complex relationship between microbial growth rate and yield and its implications for ecosystem processes. *Front. Microbiol.* **6**, 615 (2015).
66. Monod, J. The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371–394 (1949).
67. Nevozhay, D., Adams, R. M., van Itallie, E., Bennett, M. R. & Balázs, G. Mapping the environmental fitness landscape of a synthetic gene circuit. *PLoS Comput. Biol.* **8**, e1002480 (2012).
68. Beeftink, H. H., van der Heijden, R. T. J. M. & Heijnen, J. J. Maintenance requirements: energy supply from simultaneous endogenous respiration and substrate consumption. *FEMS Microbiol. Lett.* **73**, 203–209 (1990).
69. Kovárová-Kovar, K. & Egli, T. Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. *Microbiol. Mol. Biol. Rev.* **62**, 646–666 (1998).
70. Luhring, T. M. & DeLong, J. P. Scaling from metabolism to population growth rate to understand how acclimation temperature alters thermal performance. *Integr. Comp. Biol.* **57**, 103–111 (2017).
71. García-Carreras, B. et al. Role of carbon allocation efficiency in the temperature dependence of autotroph growth rates. *Proc. Natl Acad. Sci. USA* **115**, E7361–E7368 (2018).
72. Lopatkin, A. J. et al. Bacterial metabolic state more accurately predicts antibiotic lethality than growth rate. *Nat. Microbiol.* **4**, 2109–2117 (2019).
Modelling and experiments revealed that antibiotic lethality can be better predicted by the metabolic state of the cell than the growth rate.
73. Maitra, A. & Dill, K. A. Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proc. Natl Acad. Sci. USA* **112**, 406–411 (2015).
74. Aidelberg, G. et al. Hierarchy of non-glucose sugars in *Escherichia coli*. *BMC Syst. Biol.* **8**, 133 (2014).
75. Erickson, D. W. et al. A global resource allocation strategy governs growth transition kinetics of *Escherichia coli*. *Nature* **551**, 119–123 (2017).
Developed a quantitative model of bacterial growth during nutrient transitions using a coarse-grained approach.
76. Hui, S. et al. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* **11**, 784 (2015).
77. Washchina, S., D'Souza, G., Kost, C. & Kaleta, C. Metabolic network architecture and carbon source determine metabolite production costs. *FEBS J.* **283**, 2149–2163 (2016).
78. Madar, D. et al. Promoter activity dynamics in the lag phase of *Escherichia coli*. *BMC Syst. Biol.* **7**, 136 (2013).
79. Berthoumioux, S. et al. Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol. Syst. Biol.* **9**, 634 (2013).
80. Wang, X., Xia, K., Yang, X. & Tang, C. Growth strategy of microbes on mixed carbon sources. *Nat. Commun.* **10**, 1279 (2019).
81. Zelezniak, A. et al. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl Acad. Sci. USA* **112**, 6449–6454 (2015).
82. Boon, E. et al. Interactions in the microbiome: communities of organisms and communities of genes. *FEMS Microbiol. Rev.* **38**, 90–118 (2014).
83. Heyland, J., Blank, L. M. & Schmid, A. Quantification of metabolic limitations during recombinant protein production in *Escherichia coli*. *J. Biotechnol.* **155**, 178–184 (2011).
84. Hottes, A. K. et al. Bacterial Adaptation through Loss of Function. *PLoS Genet.* **9**, e1003617 (2013).
85. Morris, J. J., Lenski, R. E. & Zinser, E. R. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio* **3**, e000361-12 (2012).
86. Tsoi, R. et al. Metabolic division of labor in microbial systems. *Proc. Natl Acad. Sci. USA* **115**, 2526–2531 (2018).
87. Thommes, M., Wang, T., Zhao, Q., Paschalidis, I. C. & Segrè, D. Designing metabolic division of labor in microbial communities. *mSystems* **4**, e00263-18 (2019).
88. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
89. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69–73 (2002).
90. Lopatkin, A. J. et al. Persistence and reversal of plasmid-mediated antibiotic resistance. *Nat. Commun.* **8**, 1689 (2017).
91. Harcombe, W. R. et al. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep.* **7**, 1104–1115 (2014).
92. Knudsen, G. M., Ng, Y. & Gram, L. Survival of bactericidal antibiotic treatment by a persister subpopulation of *Listeria monocytogenes*. *Appl. Environ. Microbiol.* **79**, 7390–7397 (2013).
93. Finkelshtein, A., Roth, D., Jacob, E. Ben & Ingham, C. J. Bacterial swarms recruit cargo bacteria to pave the way in toxic environments. *mBio* **6**, e00074-15 (2015).
94. La Sarre, B., McCully, A. L., Lennon, J. T. & McKinlay, J. B. Microbial mutualism dynamics

- governed by dose-dependent toxicity of cross-fed nutrients. *ISME J.* **11**, 337–348 (2017).
95. Adamowicz, E. M., Flynn, J., Hunter, R. C. & Harcombe, W. R. Cross-feeding modulates antibiotic tolerance in bacterial communities. *ISME J.* **12**, 2723–2735 (2018).
 96. Wilson, C. E. et al. Cooperation and competition shape ecological resistance during periodic spatial disturbance of engineered bacteria. *Sci. Rep.* **7**, 440 (2017).
 97. Song, H., Payne, S., Gray, M. & You, L. Spatiotemporal modulation of biodiversity in a synthetic chemical-mediated ecosystem. *Nat. Chem. Biol.* **5**, 929–935 (2009).
 98. Tanouchi, Y., Pai, A., Buchler, N. E. & You, L. Programming stress-induced altruistic death in engineered bacteria. *Mol. Syst. Biol.* **8**, 626 (2012).
 99. Harrison, E. & Brockhurst, M. A. Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends Microbiol.* **20**, 262–267 (2012).
 100. Blake, W. J., Kærn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
 101. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
 102. Taheri-Araghi, S. et al. Cell-size control and homeostasis in bacteria. *Curr. Biol.* **25**, 385–391 (2015).
 103. Tanouchi, Y. et al. A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature* **523**, 357–360 (2015).
 104. Campos, M. et al. A constant size extension drives bacterial cell size homeostasis. *Cell* **159**, 1433–1446 (2014).
 105. Kleijn, I. T., Krah, L. H. J. & Hermsen, R. Noise propagation in an integrated model of bacterial gene expression and growth. *PLoS Comput. Biol.* **14**, 1–18 (2018).
 106. Hasty, J., Pradines, J., Dolnik, M. & Collins, J. J. Noise-based switches and amplifiers for gene expression. *Proc. Natl Acad. Sci. USA* **97**, 2075–2080 (2000).
 107. Isaacs, F. J., Hasty, J., Cantor, C. R. & Collins, J. J. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl Acad. Sci. USA* **100**, 7714–7719 (2003).
 108. Cerulus, B., New, A. M., Pougach, K. & Verstrepen, K. J. Noise and epigenetic inheritance of single-cell division times influence population fitness. *Curr. Biol.* **26**, 1138–1147 (2016).
 109. Ni, M. et al. Pre-disposition and epigenetics govern variation in bacterial survival upon stress. *PLoS Genet.* **8**, e1003148 (2012).
 110. El Meouche, I., Siu, Y. & Dunlop, M. J. Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells. *Sci. Rep.* **6**, 19538 (2016).
 111. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004).
 112. Blake, W. J. et al. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24**, 853–865 (2006).
 113. Stokes, J. M., Lopatkin, A. J., Lobritz, M. A. & Collins, J. J. Bacterial metabolism and antibiotic efficacy. *Cell Metab.* **30**, 251–259 (2019).
 114. Yang, J. H., Bening, S. C. & Collins, J. J. Antibiotic efficacy — context matters. *Curr. Opin. Microbiol.* **39**, 73–80 (2017).
 115. Brauner, A., Fridman, O., Gefen, O. & Balaban, N. Q. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nat. Rev. Microbiol.* **14**, 320–330 (2016).
 116. Narula, J., Devi, S. N., Fujita, M. & Igoshin, O. A. Ultrasensitivity of the *Bacillus subtilis* sporulation decision. *Proc. Natl Acad. Sci. USA* **109**, E3513–E3522 (2012).
 117. Chastanet, A. et al. Broadly heterogeneous activation of the master regulator for sporulation in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **107**, 8486–8491 (2010).
 118. Schultz, D., Wolynes, P. G., Jacob, E. Ben & Onuchic, J. N. Deciding fate in adverse times: Sporulation and competence in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **106**, 21027–21034 (2009).
 119. Friedman, J., Higgins, L. M. & Gore, J. Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.* **1**, 109 (2017).
 120. Venturelli, O. S. et al. Deciphering microbial interactions in synthetic human gut microbiome communities. *Mol. Syst. Biol.* **14**, e8157 (2018). **Trained a computational model that could predict a synthetic multi-species gut community based on monoculture and pairwise growth data.**
 121. Enke, T. N. et al. Modular assembly of polysaccharide-degrading marine microbial communities. *Curr. Biol.* **29**, 1528–1535.e6 (2019).
 122. Goldford, J. E. et al. Emergent simplicity in microbial community assembly. *Science* **361**, 469–474 (2018).
 123. Lawrence, D. et al. Species interactions alter evolutionary responses to a novel environment. *PLoS Biol.* **10**, e1001330 (2012).
 124. Campbell, K. et al. Self-establishing communities enable cooperative metabolite exchange in a eukaryote. *eLife* **4**, e09943 (2015).
 125. Sanchez, A. & Gore, J. Feedback between population and evolutionary dynamics determines the fate of social microbial populations. *PLoS Biol.* **11**, 1001547 (2013).
 126. Hekstra, D. R. & Leibler, S. Contingency and statistical laws in replicate microbial closed ecosystems. *Cell* **149**, 1164–1173 (2012).
 127. Jackson, M. A. et al. Detection of stable community structures within gut microbiota co-occurrence networks from different human populations. *PeerJ* **6**, e4303 (2018).
 128. Hsu, R. H. et al. Microbial Interaction network inference in microfluidic droplets. *Cell Syst.* **9**, 229–242.e4 (2019).
 129. Mettetal, J. T., Muzey, D., Pedraza, J. M., Ozbudak, E. M. & Van Oudenaarden, A. Predicting stochastic gene expression dynamics in single cells. *Proc. Natl Acad. Sci. USA* **103**, 7304–7309 (2006).
 130. Lord, N. D. et al. Stochastic antagonism between two proteins governs a bacterial cell fate switch. *Science* **366**, 116–120 (2019).
 131. Fridman, O., Goldberg, A., Ronin, I., Shores, N. & Balaban, N. Q. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature* **513**, 418–421 (2014).
 132. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **109**, E2774–E2783 (2012).
 133. Chevereau, G. et al. Quantifying the determinants of evolutionary dynamics leading to drug resistance. *PLoS Biol.* **13**, e1002299 (2015).
 134. Chan, C. T., Lee, J. W., Cameron, D. E., Bashor, C. J. & Collins, J. J. 'Deadman' and 'Passcode' microbial kill switches for bacterial containment. *Nat. Chem. Biol.* **12**, 82–86 (2016).
 135. Liao, M. J., Din, M. O., Tsimring, L. & Hasty, J. Rock-paper-scissors: engineered population dynamics increase genetic stability. *Science* **365**, 1045–1049 (2019).
 136. Scott, M., Klumpp, S., Mateescu, E. M. & Hwa, T. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol. Syst. Biol.* **10**, 747 (2014).
 137. English, M. A. et al. Programmable CRISPR-responsive smart materials. *Science* **365**, 780–785 (2019).
 138. Bothfeld, W., Kapov, G. & Tjo, K. E. J. A glucose-sensing toggle switch for autonomous, high productivity genetic control. *ACS Synth. Biol.* **6**, 1296–1304 (2017).
 139. Perez-Carrasco, R. et al. Combining a toggle switch and a repressilator within the AC-DC circuit generates distinct dynamical behaviors. *Cell Syst.* **6**, 521–530.e3 (2018). **Modelling of the integrated toggle switch and repressilator circuits revealed complex emergent behaviour inaccessible to either circuit individually.**
 140. Luro, S., Potvin-Trottier, L., Okumus, B. & Paulsson, J. Isolating live cells after high-throughput, long-term, time-lapse microscopy. *Nat. Methods* **17**, 93–100 (2020).
 141. Yeung, E. et al. Biophysical constraints arising from compositional context in synthetic gene networks. *Cell Syst.* **5**, 11–24.e12 (2017).
 142. Briat, C., Gupta, A. & Khammash, M. Antithetic integral feedback ensures robust perfect adaptation in noisy bimolecular networks. *Cell Syst.* **2**, 15–26 (2016).
 143. Potvin-Trottier, L., Lord, N. D., Vinnicombe, G. & Paulsson, J. Synchronous long-term oscillations in a synthetic gene circuit. *Nature* **538**, 514–517 (2016).
 144. Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J. & Voigt, C. A. *Escherichia coli* 'Marionette' strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol.* **15**, 196–204 (2019).
 145. Chen, Y. J. et al. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10**, 659–664 (2013).
 146. Niederholtmeyer, H. & Sun, Z. Z. Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife* **4**, e09771 (2015).
 147. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012).
 148. Tomazou, M., Barahona, M., Polizzi, K. M. & Stan, G.-B. Computational re-design of synthetic genetic oscillators for independent amplitude and frequency modulation. *Cell Syst.* **6**, 508–520.e5 (2018).
 149. Kong, W., Meldgin, D. R., Collins, J. J. & Lu, T. Designing microbial consortia with defined social interactions. *Nat. Chem. Biol.* **14**, 821–829 (2018).
 150. Solé, R. V., Montañez, R. & Duran-Nebreda, S. Synthetic circuit designs for earth terraformation. *Biol. Direct* **10**, 37 (2015).
 151. Hiscock, T. W. Adapting machine-learning algorithms to design gene circuits. *BMC Bioinformatics* **20**, 214 (2019).
 152. Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C. & Collins, J. J. Next-generation machine learning for biological networks. *Cell* **173**, 1581–1592 (2018).
 153. Wu, F. et al. A unifying framework for interpreting and predicting mutualistic systems. *Nat. Commun.* **10**, 242 (2019). **Derived a simplified criterion using a support vector machine approach that predicts the outcome of a mutualistic bacterial community regardless of the underlying interactions.**
 154. Zhang, P.-Y. et al. Combined treatment with the antibiotics kanamycin and streptomycin promotes the conjugation of *Escherichia coli*. *FEMS Microbiol. Lett.* **348**, 149–156 (2013).
 155. Schuurmans, J. M. et al. Effect of growth rate and selection pressure on rates of transfer of an antibiotic resistance plasmid between *E. coli* strains. *Plasmid* **72**, 1–8 (2014).
 156. Lopatkin, A. J., Sysyova, T. A. & You, L. Dissecting the effects of antibiotics on horizontal gene transfer: analysis suggests a critical role of selection dynamics. *Bioessays* **38**, 1283–1292 (2016). **Demonstrated that antibiotics do not promote horizontal gene transfer as previously thought by combining quantitative measurements of plasmid conjugation and mathematical modelling of population dynamics.**
 157. Lopatkin, A. J. et al. Antibiotics as a selective driver for conjugation dynamics. *Nat. Microbiol.* **1**, 16044 (2016).
 158. Headd, B. & Bradford, S. A. Physicochemical factors that favor conjugation of an antibiotic resistant plasmid in non-growing bacterial cultures in the absence and presence of antibiotics. *Front. Microbiol.* **9**, 2122 (2018).
 159. Korem, T. et al. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**, 1101–1106 (2015).
 160. Scott, S. R. et al. A stabilized microbial ecosystem of self-limiting bacteria using synthetic quorum-regulated lysis. *Nat. Microbiol.* **2**, 17083 (2017).
 161. Scharli, Y. et al. Synthetic circuits reveal how mechanisms of gene regulatory networks constrain evolution. *Mol. Syst. Biol.* **14**, e8102 (2018).
 162. Rodriguez de Evgrafov, M., Gumpert, H., Munck, C., Thomsen, T. T. & Sommer, M. O. A. Collateral resistance and sensitivity modulate evolution of high-level resistance to drug combination treatment in *Staphylococcus aureus*. *Mol. Biol. Evol.* **32**, 1175–1185 (2015).
 163. Wood, K., Nishida, S., Sontag, E. D. & Cluzel, P. Mechanism-independent method for predicting response to multidrug combinations in bacteria. *Proc. Natl Acad. Sci. USA* **109**, 12254–12259 (2012).
 164. Meredith, H. R., Lopatkin, A. J., Anderson, D. J. & You, L. Bacterial temporal dynamics enable optimal design of antibiotic treatment. *PLoS Comput. Biol.* **11**, e1004201 (2015).
 165. Yang, J. H. et al. A white-box machine learning approach for revealing antibiotic mechanisms of action. *Cell* **177**, 1649–1661.e9 (2019). **Integrated machine learning with metabolic network modelling to uncover novel metabolism-related mechanisms of action for bactericidal antibiotics.**

166. Nichol, D. et al. Antibiotic collateral sensitivity is contingent on the repeatability of evolution. *Nat. Commun.* **10**, 334 (2019).
167. Nguyen, M. et al. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal Salmonella. *J. Clin. Microbiol.* **57**, e01260-18 (2019).
168. Levy, S. F. et al. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181–186 (2015).
169. van der Ark, K. C. H. et al. Model-driven design of a minimal medium for *Akkermansia muciniphila* confirms mucus adaptation. *Microb. Biotechnol.* **11**, 476–485 (2018).
170. Wang, S. et al. Massive computational acceleration by using neural networks to emulate mechanism-based biological models. *Nat. Commun.* **10**, 4354 (2019).
171. Hart, S. F. M. et al. Uncovering and resolving challenges of quantitative modeling in a simplified community of interacting cells. *PLoS Biol.* **17**, e3000135 (2019).
172. Cremer, J., Arnoldini, M. & Hwa, T. Effect of water flow and chemical environment on microbiota growth and composition in the human colon. *Proc. Natl Acad. Sci. USA* **114**, 6438–6443 (2017).
173. Heckmann, D. et al. Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nat. Commun.* **9**, 5252 (2018).
174. Friedman, N. Inferring cellular networks using probabilistic graphical models. *Science* **303**, 799–805 (2004).

Acknowledgements

The authors thank Lingchong You for his feedback on the manuscript. This work was supported by the Defence Threat

Reduction Agency (HDTRA1-15-1-0051), the Broad Institute of MIT and Harvard, and a generous gift from Anita and Josh Bekenstein.

Author contributions

A.J.L. researched data for the manuscript. A.J.L. and J.J.C. outlined the manuscript. A.J.L. wrote the manuscript. A.J.L. and J.J.C. reviewed and edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

Publisher's note

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

© Springer Nature Limited 2020