



# Computer-Aided Whole-Cell Design: Taking a Holistic Approach by Integrating Synthetic With Systems Biology

## OPEN ACCESS

### Edited by:

Dong-Yup Lee,

Sungkyunkwan University,  
South Korea

### Reviewed by:

Hyun Uk Kim,

Korea Advanced Institute of Science  
and Technology, South Korea  
Meiyappan Lakshmanan,  
Bioprocessing Technology Institute  
(A\*STAR), Singapore

### \*Correspondence:

Lucia Marucci  
lucia.marucci@bristol.ac.uk

Matteo Barberis  
m.barberis@surrey.ac.uk;  
matteo@barberislab.com

Jonathan Karr  
karr@mssm.edu

Oliver Ray  
csxor@bristol.ac.uk

Paul R. Race  
Paul.Race@bristol.ac.uk

Claire Grierson  
claire.grierson@bristol.ac.uk

Elbio Rech  
elibio.rech@embrapa.br

Richard Seabrook  
richard.seabrook@bristol.ac.uk

Christopher Woods  
Christopher.Woods@bristol.ac.uk

† These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Synthetic Biology,  
a section of the journal  
Frontiers in Bioengineering and  
Biotechnology

**Received:** 29 May 2020

**Accepted:** 21 July 2020

**Published:** 07 August 2020

**Lucia Marucci**<sup>1,2,3\*</sup>, **Matteo Barberis**<sup>4,5,6\*†</sup>, **Jonathan Karr**<sup>7\*†</sup>, **Oliver Ray**<sup>8\*†</sup>,  
**Paul R. Race**<sup>3,9\*†</sup>, **Miguel de Souza Andrade**<sup>10,11</sup>, **Claire Grierson**<sup>3,12\*</sup>,  
**Stefan Andreas Hoffmann**<sup>13</sup>, **Sophie Landon**<sup>1,3</sup>, **Elbio Rech**<sup>10\*</sup>, **Joshua Rees-Garbutt**<sup>3,12</sup>,  
**Richard Seabrook**<sup>14\*</sup>, **William Shaw**<sup>15</sup> and **Christopher Woods**<sup>3,16\*</sup>

<sup>1</sup> Department of Engineering Mathematics, University of Bristol, Bristol, United Kingdom, <sup>2</sup> School of Cellular and Molecular Medicine, University of Bristol, Bristol, United Kingdom, <sup>3</sup> Bristol Centre for Synthetic Biology (BrisSynBio), University of Bristol, Bristol, United Kingdom, <sup>4</sup> Systems Biology, School of Biosciences and Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom, <sup>5</sup> Centre for Mathematical and Computational Biology, CMCB, University of Surrey, Guildford, United Kingdom, <sup>6</sup> Synthetic Systems Biology and Nuclear Organization, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands, <sup>7</sup> Icahn Institute for Data Science and Genomic Technology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States, <sup>8</sup> Department of Computer Science, University of Bristol, Bristol, United Kingdom, <sup>9</sup> School of Biochemistry, University of Bristol, Bristol, United Kingdom, <sup>10</sup> Brazilian Agricultural Research Corporation/National Institute of Science and Technology – Synthetic Biology, Brasília, Brazil, <sup>11</sup> Department of Cell Biology, Institute of Biological Sciences, University of Brasilia, Brasilia, Brazil, <sup>12</sup> School of Biological Sciences, University of Bristol, Bristol, United Kingdom, <sup>13</sup> Manchester Institute of Biotechnology, The University of Manchester, Manchester, United Kingdom, <sup>14</sup> Elizabeth Blackwell Institute for Health Research (EBI), University of Bristol, Bristol, United Kingdom, <sup>15</sup> Department of Bioengineering, Imperial College London, London, United Kingdom, <sup>16</sup> School of Chemistry, University of Bristol, Bristol, United Kingdom

Computer-aided design (CAD) for synthetic biology promises to accelerate the rational and robust engineering of biological systems. It requires both detailed and quantitative mathematical and experimental models of the processes to (re)design biology, and software and tools for genetic engineering and DNA assembly. Ultimately, the increased precision in the design phase will have a dramatic impact on the production of designer cells and organisms with bespoke functions and increased modularity. CAD strategies require quantitative models of cells that can capture multiscale processes and link genotypes to phenotypes. Here, we present a perspective on how whole-cell, multiscale models could transform design-build-test-learn cycles in synthetic biology. We show how these models could significantly aid in the design and learn phases while reducing experimental testing by presenting case studies spanning from genome minimization to cell-free systems. We also discuss several challenges for the realization of our vision. The possibility to describe and build whole-cells *in silico* offers an opportunity to develop increasingly automatized, precise and accessible CAD tools and strategies.

**Keywords:** whole-cell models, synthetic biology, systems biology, multiscale models, bioengineering, biodesign

## INTRODUCTION

Whole-cell models (WCMs) are state-of-the-art Systems Biology formalisms: they aim at representing and integrating all cellular functions within a unique computational framework, ultimately enabling a holistic, and quantitative understanding of cell biology (Tomita, 2001; Karr et al., 2015a). Quantitative and high-throughput *in silico* experiments generated from WCMs promise to significantly shorten the distance between hypothesis/design formulation and testing (Carrera and Covert, 2015).

While simplified models for specific cellular functions were first developed over 30 years ago [e.g., gene expression regulation (McAdams and Arkin, 1997), signaling (Morton-Firth and Bray, 1998) and metabolic pathways (Cornish-Bowden and Hofmeyr, 1991), cell growth (Shu and Shuler, 1989) and the cell cycle (Goldbeter, 1991; Tyson, 1991; Novak and Tyson, 1993)], the first WCM, the E-Cell model, was only derived in the 1990s for *Mycoplasma genitalium*, which has the smallest genome among freely living organisms (Tomita et al., 1999). The so-called virtual self-surviving cell (SSC) model is partially stochastic; it includes only a subset of protein-coding genes and enables dynamic simulations which encompass various subcellular processes, including enzymatic reactions, complex formation and substance translocation. In parallel, the first genome-scale metabolic models (GSMs) were developed by Palsson's group (Varma and Palsson, 1994) using flux balance analysis (FBA) in the 1990s.

More recently, hundreds of GSMs have been reconstructed for different organisms, with an increasing number of represented genes (McCloskey et al., 2013; Yilmaz and Walhout, 2017; Mendoza et al., 2019). GSMs have been complemented with a mathematical description of other processes, such as transcription, translation, and signaling (Lee et al., 2008; Thiele et al., 2009). Less than a decade ago a more complete, hybrid WCM, representing all genes and molecular functions known for an organism, was reported by Covert's group (Karr et al., 2012). In this pioneering work, Karr and colleagues integrated 28 sub-models to represent one cell cycle of *M. genitalium*; each sub-model is represented with a distinct formalism, including ordinary differential equations (ODEs), FBA, stochastic simulations and Boolean rules.

Substantial research and effort are still needed to improve WCMs' descriptive power and to increase the complexity of organisms they can represent. Developing a WCM is a challenging task, which requires the collection of extensive experimental data, integration of sub-cellular models and *in silico/in vivo* model validation. A complete WCM should ideally integrate multiscale interactions at the cellular level (Karr et al., 2012; King et al., 2016) while accounting for the overall cellular structure (Betts and Russell, 2007), the dynamic structure of molecular interactions (Noske et al., 2008; McGuffee and Elcock, 2010; Yu et al., 2016), and the spatial compartment of the subcellular components (Ander et al., 2004; Takahashi et al., 2005; Thul et al., 2017). Ensuring an accurate representation of all of the cellular processes across organisms of increasing complexity is highly challenging (Bouhaddou et al., 2018; Singla et al., 2018; Szigeti et al., 2018). It is therefore not surprising that, to date, only the *M. genitalium* and, very recently, *E. Coli* (Macklin

et al., 2020). WCMs have been released, although several other WCMs are currently under development<sup>1</sup>. We refer the reader to recent efforts which provide an overview of the state-of-the-art in the development of WCMs (Goldberg et al., 2018; Feig and Sugita, 2019).

Here, we focus on the enormous potential we believe WCMs have for design-build-test cycles integrating synthetic with systems biology (Figure 1). While the applications are diverse, they share a high degree of complexity which would require extensive trial and error experimental cycles in the absence of robust computational design algorithms based on predictive models. We conclude by considering relevant challenges that must be addressed by interdisciplinary communities to fully realize our vision, discussing future directions for integrating WCMs through synthetic and systems biology.

## WHOLE-CELL DESIGN STRATEGIES IN SYNTHETIC BIOLOGY

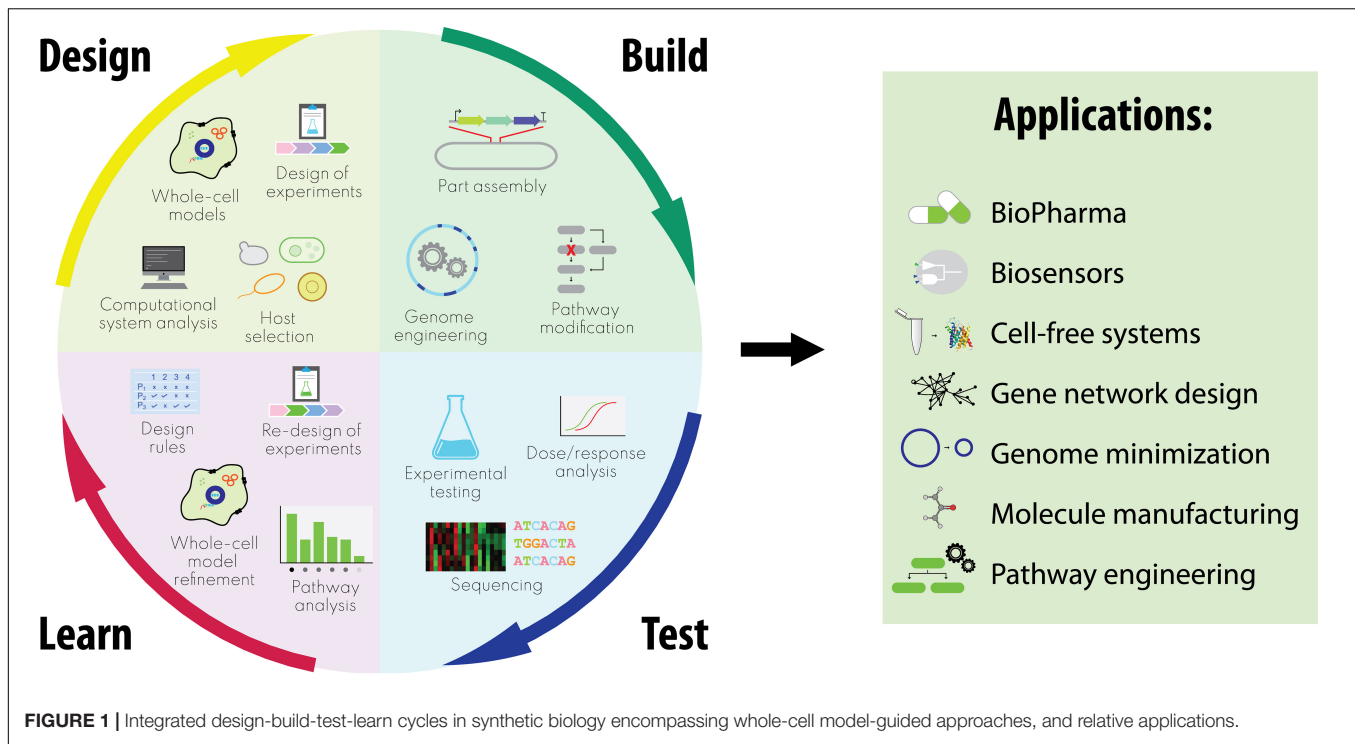
### Model Granularity of Gene Network (re)Design

Mathematical models can be instrumental for the (re)design of network circuits that recapitulate definite biological functions. Knowledge of regulatory mechanisms in biological pathways has been gained by considering living systems as a composition of functional modules, which are investigated through minimal computer models. Examples include controllable oscillators (Marucci et al., 2009; Purcell et al., 2010, 2013; Tomazou et al., 2018), circadian clocks (Gerard et al., 2009; Ananthasubramaniam et al., 2020), signaling networks (Prescott and Abel, 2017), the metabolism (Castellanos et al., 2004; Pandit et al., 2017), and transcriptional regulation (Carrera et al., 2009). Existing minimal and detailed computer models span a broad range of granularity in their biochemical details. However, one may expect that, if the core design of a minimal and a detailed model is similar, their general properties will match.

The understanding of a living organism at a system's level may be reached through decomposing it into functional modules or modular circuits (Hartwell et al., 1999; Kitano, 2002; Ravasz et al., 2002). The capability to sustain viability through autonomously generated offspring is essential. It is therefore a feature that WCMs shall account for through modeling cell division, which is intimately integrated with various layers of cellular regulation (metabolism, signaling, gene regulation, transcription, etc.). A number of minimal models have been developed for the eukaryotic cell cycle by Barberis', Tyson's and Novák's groups (Battogtokh and Tyson, 2004; Barberis et al., 2012; Gerard et al., 2013, 2015; Linke et al., 2017; Mondeel et al., 2020).

Currently, the majority of multiscale models (not WCMs) lack components able to bridge cellular networks or function (cell cycle, metabolism, signaling, gene regulation, etc.). Identification of hubs, i.e., elements with high connectivity in the cellular environment that integrate cellular networks, is a critical feature

<sup>1</sup><https://www.wholecell.org/>



of WCMs. Transcription factors have recently been identified as hubs that integrate multiscale networks, potentially connecting the cell cycle to metabolism (Mondeel et al., 2019), and can be among the parts of a system that influence its state as a whole. Multiscale frameworks coupling networks of differing granularity are being developed, by identifying the relevant regulations occurring among common network nodes and through the use of different mathematical formalisms (van der Zee and Barberis, 2019). These and other strategies are also being developed to integrate networks of cellular functional modules (Prescott et al., 2015). Together with the identification of networks underlying the cell's autonomous oscillations, these strategies can rationalize the proper timing of offspring generation accounted by WCMs.

Designing synthetic gene networks by modeling and integrating them within WCM formalisms [as in Purcell et al. (2013)] could be critical to investigate how gene expression correlates with codon usage, explore possible cell burden effects (Borkowski et al., 2016), and predict modularity of synthetic gene networks and tools to modulate gene expression across different chassis (Way et al., 2014; Pedone et al., 2019; Gomide et al., 2020).

## Design and Engineering of Reduced Genomes

Minimal genomes can be defined as reduced genomes containing only the genetic material which is essential for a cell to reproduce (Glass et al., 2017). Studying and engineering minimal genomes can be instrumental both to understand the most essential tasks a cell must perform to sustain life, and to obtain optimal chassis for synthetic biology applications, with reduced cell burden and superior robustness (Moya et al., 2009;

Hutchison et al., 2016; Ceroni and Ellis, 2018; Mol et al., 2018; Landon et al., 2019).

Exhaustive experimental characterization of a minimized genome is unfeasible: even for an organism as small as *M. genitalium* (0.58 mb and 525 genes), there are thousands of possible combinations of gene knockouts to be performed. Of note, this figure is most probably underestimated, accounting for the fact that the order in which gene deletions are performed can alter the resulting phenotypes (Gawand et al., 2015). Genome-scale computational models of cells could be instrumental to fully understand the dynamic and context-dependent nature of gene essentiality (Rancati et al., 2018), and to rationally design minimized genomes *in silico*. Computer-aided minimal genome engineering could significantly reduce the time and cost to reduce genomes compared to current approaches based on extensive experimental iterations (Posfai et al., 2006; Iwadate et al., 2011; Hirokawa et al., 2013; Hutchison et al., 2016; Zhou et al., 2016; Reuss et al., 2017; Breuer et al., 2019).

To the best of our knowledge, two top-down genome reduction approaches have been proposed so far based on genome-scale models. The MinGenome algorithm applies a mixed-integer linear programming (MILP) algorithm to a GSMM of *Escherichia coli*, using information pertaining to essential genes and synthetic lethal pairs within the optimization (Wang and Maranas, 2018). In contrast, Minesweeper and GAMA are top-down genome minimization algorithms based on the *M. genitalium* WCM. They exploit a divide-and-conquer approach and a biased genetic algorithm, respectively, to iteratively simulate reduced genomes (Rees-Garbutt et al., 2020); their *in silico* predictions have not been tested in the laboratory yet.

GSMM-based genome reduction algorithms such as MinGenome or analogous, adaptable metaheuristic techniques [e.g., (Burgard et al., 2003; Tang et al., 2015; Mutturi, 2017)] are currently more broadly applicable across organisms given the large availability of these formalisms. Still, as more WCMs become available, we expect WCM-based genome reduction algorithms to provide superior predictions of cellular processes and genetic interactions, thanks to their richness of multiscale cellular process representation.

## Design and Prototyping of Cell-Free Systems

Cell-free transcription/translation systems, based on crude cellular extracts, are a valuable platform to address fundamental biological questions in a controllable and reproducible way. In recent years, the decrease of costs associated with this technology and significant improvements in synthesis yield capabilities (Calhoun and Swartz, 2005) have made cell-free systems increasingly popular in synthetic biology for the prototyping and testing of engineered biological parts (McCloskey et al., 2013; Reuss et al., 2017; Yilmaz and Walhout, 2017; Mendoza et al., 2019) and networks (Noireaux et al., 2003; Siegal-Gaskins et al., 2014; Takahashi et al., 2015). As the possible applications of cell-free systems grow [see (Silverman et al., 2020) for a recent review], mathematical models are being developed to quantitatively formalize how biological processes perform within cell-free platforms (Koch et al., 2018).

So far, deterministic models (ODEs, or constraint-based) have been proposed to describe specific processes within cell-free platforms such as transcription and translation (Karzbrun et al., 2011; Stogbauer et al., 2012; Siegal-Gaskins et al., 2014), resource competition (Underwood et al., 2005; Borkowski et al., 2018; Matsuura et al., 2018; Moore et al., 2018), and metabolism (Vilkhovoy et al., 2018). The integration of mathematical formalisms across scales for cell-free platforms, building toward WCMs, could be highly beneficial to both facilitate *de novo* design of circuits, and to quantitatively compare *in vitro* cell-free products with their *in vivo* counterparts.

## Whole-Cell Biosensor Design and Testing

Biosensors are analytical devices which can convert a biochemical reaction into a measurable signal. The recognition unit in a biosensor can be composed of whole cells, nucleic acids, enzymes, proteins, antibodies or combinations thereof. Synthetic biology has significantly accelerated biosensor development; new generation whole-cell biosensors (i.e., sensors implemented throughout living cells) have been engineered, allowing, for example: arsenic detection (Diesel et al., 2009), detection of pollutants and antibiotics (van der Meer and Belkin, 2010), microbial detection in industrial settings (Lu et al., 2013) and *in vivo* diagnostic applications [e.g., detection of environmental signals in the gut (Kotula et al., 2014) and diagnosis of liver

metastases (Danino et al., 2015); see (Slomovic et al., 2015) for an overview].

The application of WCMs to the design, prototyping and testing of whole-cell biosensors could suggest rational approaches to tune their sensitivity, stability, and dynamic range while facilitating the choice of the ideal chassis and, if needed, guide its re-engineering to optimize biosensor performance (Hicks et al., 2020). If WCMs become available for different chassis and entire organisms, they could also support the design of optimized targeted delivery of genetically encoded biosensors.

## Industrial Implications of Whole-Cell Models

Although the intellectual merit of pursuing a computer-aided whole-cell design approach is unquestioned, it is clear that the success of this endeavor will ultimately be judged by its impact on science, medicine, and industry. The increasing drive of computer-aided designs (CADs) toward “green” chemistry approaches, allied to increases in gene synthesis speed and capability and associated cost reductions, are making biosynthesis an increasingly appealing route for the manufacture of high-value chemicals (El Karoui et al., 2019). This includes a plethora of opportunities across the pharmaceutical, agrochemical, commodity chemical, and materials sectors, amongst others.

A major challenge, however, remains the development of robust, scalable microbial chassis, whose metabolic processes can be predictably tuned for a desired outcome (Xu et al., 2020). Currently, chassis choice is largely restricted to a subset of genetically tractable microorganisms, whose physiology and performance during fermentation are well understood, and for whom effective molecular genetic tools required for their manipulation exist. Chassis optimization to date has relied exclusively on incremental, stepwise improvements in desired host strain characteristics, including growth rate, feedstock utilization, and product yield (Calero and Nikel, 2019). For these reasons, the process of chassis optimization remains prohibitively slow and expensive, accounting in part for the paucity of high-value small molecules that are currently manufactured using synthetic biology processes. Targeted manipulations often lead to unanticipated off-target effects, linked to the co-dependency of metabolic processes, which generally function in concert within interdependent cellular networks (Woolston et al., 2013): perturbations may compromise rather than enhance desirable characteristics, leading to undesired outcomes. Clearly, robust, predictable WCMs represent an attractive solution to the problem of chassis optimization, affording a catch-all tool that can be used to unpick dependencies and ensure that performance criteria can be met.

Additionally, the complexities associated with population heterogeneity during chassis fermentation must be resolved (Danchin, 2012). For fermentation-based industrial processes to be tractable, product yields must be sufficiently high to make biosynthesis financially viable. The emergence of “cheaters” or slow-growers within microbial populations should be tackled

with tunable regulatory processes that operate throughout populations. The introduction of such characteristics is a major challenge to conventional chassis design approaches. WCM-driven approaches could more easily implement and test these processes.

Critical to the success of a computer-aided whole-cell design approach is the quality of the employed model (Fernandez-Castane et al., 2014). Microbial systems with small genomes represent a compelling entry point for study, with model development possibly being facilitated by ongoing studies focused on establishing the core constituents of a functional genome. These studies are in part driven by genome minimization experiments, which in turn can be used to further refine model performance. Importantly, fundamental gaps remain in our understanding of microbial metabolic processes, and this will unquestionably hinder progress (Price et al., 2018). However, the capacity of WCMs to predict previously unidentified metabolic dependencies should be viewed as an acid test of model validity. Indeed, GSMs often fail due to their inability to account for metabolic dependencies, a feature which has led to skepticism within industrial circles, questioning the value of such models. Whole-cell approaches offer a mechanism to circumvent this issue. This is of particular significance when developing chassis for “non-natural” products whose chemistries sit outside those of metabolites found in nature (Calero and Nikel, 2019). Expanding the metabolic capacity of chassis organisms to deliver such novel products risks introducing additional complexities, including excessive depletion of core metabolite pools or the generation of toxic products or intermediates. Design approaches driven by WCMs are uniquely placed to identify such issues and provide a route to their circumvention.

The capacity to design-in explicit control over cellular behavior is also critical for industrial adoption of model-derived chassis. It can be argued that the ability to regulate cellular processes is as important as defining the processes themselves. Tunable regulatory systems must afford a degree of both intrinsic and extrinsic control. Synthetic biology-based approaches for constructing genetic circuitry are now placing us on a path to broad-reaching cellular regulation, though issues still exist. These systems are often insufficiently orthogonal, with bespoke designs required for different chassis due to variations in core metabolic process (Pandit et al., 2017). Again, whole-cell design approaches offer a solution to this issue, as such systems can be predefined and tested for functionality *in silico* prior to undertaking costly lab experimentation.

## WHAT'S NEXT? GOING BEYOND THE PROTOTYPE

In recent years, advances in genomic measurement technologies for data generation, the establishment of data repositories, and the development of WCM simulation platforms have significantly facilitated the derivation of WCMs [see (Goldberg et al., 2018) for a review]. Nevertheless, the implementation of WCM-based design-build-test cycles for genome-scale

engineering requires further challenges to be addressed (Bartley et al., 2020).

If a model has to be used for the design and prototyping of an engineered living system, the model needs to be reliable. Even for a simple organism, the number of kinetic parameters raises as the complexity and the level of detail of a mathematical model increase; constraining parameters thus becomes harder and requires extensive experimental data. Mathematical models can be used to produce predictions of missing data, however, they often abstract physical processes using simplifying assumptions which might hold in specific conditions (Babtie and Stumpf, 2017). To set the 1,462 quantitative parameters of the *M. genitalium* WCM, values from related organisms were incorporated due to a lack of organism-specific data (Macklin et al., 2014); a combination of parameter values reported from previous experiments and numerical optimization on a reduced model was performed. While, ideally, we would like to measure all kinetic parameters directly from experiments, we still lack the ability to measure each state in individual cells over time, and across all possible environmental conditions. A combination of direct experimental estimation and parameter inference will likely be needed for genome-scale formalisms and WCMs.

Sensitivity analysis, usually performed by perturbing parameters to understand how uncertainties affect the model outputs (Erguler and Stumpf, 2011), can become extremely computationally expensive when applied to genome-scale models. Alternatively, statistical approaches such as those based on Bayesian methods (Vernon et al., 2018) or the Fisher information matrix (Rand, 2008) could be carefully carried out at least at the sub-model level, and possibly scaled up to WCMs. The Reverse Engineering Assessments and Methods (DREAM8) parameter estimation challenge (Karr et al., 2015b) was organized to develop new parameter estimation techniques specific for WCMs. It suggested possible interesting avenues for WCM parameterization (i.e., model reduction and a combination of differential evolution and random forests), and highlighted that the availability of comprehensive data is critical to ensure the model is practically identifiable (Ashyraliyev et al., 2009), and to calibrate WCMs.

Researchers have started to collect data needed for WCM development into public repositories [e.g., (Wittig et al., 2012; Kolesnikov et al., 2015; Sajed et al., 2016; UniProt Consortium, 2018; Caspi et al., 2020)]; still, the data needed to derive and fit WCMs are dispersed across many repositories and publications and often not annotated or normalized, ultimately requiring a massive manual effort. Federated archives of repositories, such as the PDB-Dev system to deposit Integrative/Hybrid models and corresponding data (Burley et al., 2017), also exist and might be well placed to archive and disseminate both data and models, while enabling different researchers to attempt alternative modeling/parameterization approaches. Cover's group developed the WholeCellKB database (Karr et al., 2013) to organize the quantitative measurements (over 1,400) from which the *M. genitalium* WCM was derived; it would be ideal to enable automatic access and querying in such databases.

To enhance WCM reproducibility and collaboration, new standards and simulations software are also needed (Medley et al., 2016). Researchers should invest efforts to use and expand the capabilities of standard formats such as the Systems Biology Markup Language (SBML) (Hucka et al., 2003) and the Systems Biology Graphical Notation (SBGN) (Le Novère et al., 2009) to be suitable for WCMs. For example, several aspects of the *M. genitalium* WCM cannot be represented by SBML, such as the multi-algorithmic nature of the model (Waltemath et al., 2016). Further development of standard modeling formats is needed to enable reproducible WCM simulations, e.g., by including in the SBML Hierarchical Model Composition package ontologies which could represent the algorithm needed for specific sub-models (Courtot et al., 2011). In the context of synthetic biology applications, we believe it would be appropriate and beneficial to report and deposit data related to various iterations of WCM-generated *in silico* predictions, *in vivo* testing and possible model/design refinement; this would establish the predictive power of WCMs and illuminate steps to make design-build-test-learn cycles more effective.

It is also important to consider the structural uncertainties in the model, which depend on model assumptions. While, for certain sets of models (e.g., small ODE systems for signaling pathways), likelihood- and Bayesian-based approaches have been proposed for model selection (Wilkinson, 2007; Kirk et al., 2013) and semidefinite programming for model invalidation (Anderson and Papachristodoulou, 2009), no suitable techniques for WCMs have been proposed to date.

We foresee that automation will play a fundamental role in the derivation of WCMs for eukaryotic organisms and in their application to design complex processes. Ideally, we would like to introduce automation at different stages, such as data extraction from the literature, model derivation, and model/data integration both within the model fitting and validation steps, and when comparing *in silico* design prediction with *in vivo* tests (Bartley et al., 2020). This, in turn, will require the adoption of standards for both data and model repositories. Also, laboratory automation, coupled to WCM-based CAD, is expected to transform design-build-test cycles. As the use of robotics becomes increasingly common in both academia and industry, the throughput and reproducibility of experiments needed for both WCM derivation and validation can be significantly increased, and protocol sharing across research communities facilitated (Jessop-Fabre and Sonnenschein, 2019).

To assist the adoption of WCMs for synthetic biology applications, high-performance parallelized computer clusters are required to run the models with lengthy runtimes, coordinate the corresponding databases, parameterize and validate the models, and then to integrate WCMs in design cycles in combination with optimization algorithms (Macklin et al., 2014; Chalkley et al., 2019).

The implementation of standardized tools to share data and simulate WCMs would, in turn, facilitate model validation. This should involve the definition of proper metrics and formal model verification techniques such as those developed for SBML-encoded models (Kwiatkowska et al., 2011).

## (RE)THINKING SYSTEM APPROACHES: A COLLABORATIVE EFFORT

In addressing the aforementioned challenges, we believe there is a tremendous opportunity to rethink approaches used so far to generate genome-scale models, including WCMs, and to integrate with broader communities including software engineers, computer scientists, structural biologists, bioinformaticians, and systems and synthetic biologists.

We do anticipate that, as diverse communities synergize on WCM-related research, different kinds of formalisms might be integrated within genome-scale models. Symbolic reasoning provides a range of expressive and intuitive logical frameworks that could potentially complement and help glue together sub-models at different scales. Such methods are routinely applied to complex systems in the electronics and software industries, and have been applied to biological systems for nearly a decade (Iyengar, 2011). Recent work showed the feasibility of applying logic programming methods to signaling pathways (Ray et al., 2011), metabolic networks (Bragagli and Ray, 2015) and automating a mechanistic philosophy of scientific discovery in simulated organisms (Rozanski et al., 2015); it should be feasible to integrate such sub-models within a WCM framework.

We believe there is scope to further increase the descriptive and predictive ability of WCMs across spatial and temporal scales by integrating the structural biology and the molecular modeling communities to carefully consider not only the biochemical, but also the physical, molecular and structural components of cells. The development of the so-called “physical” WCMs [see (Feig and Sugita, 2019) and (Feig and Sugita, 2013) for comprehensive reviews] is an emerging field, with the first models describing minimal cellular environments in full atomistic detail (Feig et al., 2015; Yu et al., 2016). With the final aim to integrate biochemical and physical WCMs within a multiscale framework (Sali et al., 2015), we need approaches which can cope with the limitations of atomistic models of biomolecules (mainly in terms of computational resources), possibly exploiting coarse-grained (Ando and Skolnick, 2010; Hyeon and Thirumalai, 2011) or continuum (Solernou et al., 2018) approaches.

By collaborating with software engineers, we need to develop tools which can enable, and possibly automate, the integration of different data types across scales, model derivation, fitting and validation, and visualization and interpretation of results (Szigeti et al., 2018).

Moreover, rule-based models might become the new standard to represent each molecular species with the required level of granularity and multi-algorithmic sub-models (e.g., FBA and stochastic dynamical models). Frameworks where intuitive logic is coupled to rule-based models have started to be developed recently (van der Zee and Barberis, 2019).

As we produce ever-increasing amounts of experimental data and increasingly sophisticated computational tools to realize detailed and complex representations of actual cells, approaches instead focusing on deliberately abstract and parsimonious simulations of artificial cellular systems provide

a valuable change of perspective. Such “toy models” might be a valuable tool to test different algorithms for model derivation and fitting, while offering an opportunity to engage with broader research communities and with the public (Castiglione et al., 2014).

Finally, we believe there is tremendous potential for applying machine learning techniques to both WCM derivation and their applications in synthetic biology. Two recent works (Lin et al., 2017; Ma et al., 2018) showed that deep neural networks are well placed to reconstruct the architecture of living systems [namely, the hierarchical organization of nuclear transcriptional factors in the nucleus (Lin et al., 2017) and of a basic eukaryotic cell (Ma et al., 2018)] and predict cell states and phenotypes. In both cases, the configuration of network layers and thus the biological structure were formulated using extensive prior knowledge, ultimately enabling fully “visible” systems, where all the internal biological states can be interrogated mechanistically (Yu et al., 2018). Machine learning could be beneficial to systematically process large *in vivo* and *in silico* whole-cell datasets, for example by applying Bayesian inference, to integrate data from diverse sources and supplement sparse data (Perdikaris and Karniadakis, 2016), and to help to automatically classify WCM simulations and link phenotypes to genotypes (Alber et al., 2019). Ensemble methods, which combine multiple independent models into a single predictive model for increasing the overall robustness of predictions, might also be adopted to develop subcellular formalisms and support their integration across chassis (Camacho et al., 2018). Additionally, machine learning might assist in WCM parameter identification, for example applying Bayesian parameter estimation (Vyshemirsky and Girolami, 2008), regression models and reinforcement learning techniques (Alber et al., 2019). Optimal experimental design techniques might also offer a valuable methodology to select the best experimental datasets for both model identification and validation (Smucker et al., 2018).

## DISCUSSION

We have shown that WCMs are likely to be instrumental to inform design-build-test cycles across synthetic biology applications. WCMs can accelerate the realization of “designer” cells and organisms tailored to specific functions, reducing experimental iterations and increasing the predictive power of computational formalisms used so far.

In the (re)design of cellular network functionalities, it is therefore important to quantitatively analyze and predict, through dedicated modeling strategies, the dynamics of interactions between various layers of cellular regulation. Thus, WCMs should take into account how different cellular layers are integrated, and how regulatory feedback among these layers occurs in time. These challenges must be tackled through integrative computational and experimental collaborative efforts aimed, respectively, toward: (i) engineering *in vivo* network designs which, through predictive systems biology, may be able to autonomously oscillate, sustaining generation of offspring, and (ii) extraction, visualization and functional exploration of

regulatory interactions among cellular layers through novel multiscale modeling frameworks.

As synthetic biology moves toward the (re)engineering of entire genomes and multicellular systems, interdisciplinary communities need to collaborate for the development of tools that are required to improve the predictive power of WCMs. Although challenges remain, it is clear that the adoption of model-based methods has the potential to transform both basic research and the current bioproduction development process, leading to marked improvements in host performance and product yield on an industrial scale.

Ultimately, as the development of human genome-scale kinetic models becomes more feasible (Bordbar et al., 2015; Szigeti et al., 2018), we anticipate that whole-cell formalisms will become an indispensable tool to study human variation, and design treatments and synthetic cellular screening systems.

## AUTHOR CONTRIBUTIONS

LM, MB, JK, OR, and PR wrote the manuscript. MS prepared the figure. All other authors participated to discussion within the workshop, helped with editing, and/or provided feedback.

## FUNDING

LM was funded by the Engineering and Physical Sciences Research Council (EPSRC, grants EP/R041695/1 and EP/S01876X/1) and Horizon 2020 (CosyBio, grant agreement 766840); MB was funded by the Systems Biology Grant of the University of Surrey; JK was funded by the National Institutes of Health (award R35GM119771); PR was funded by the EPSRC (EP/R020957/1) and the Biotechnology and Biological Sciences Research Council (BBSRC, BB/T001968/1); CG, LM, and PR were funded by the BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre (BB/L01386X/); SL and JR-G were funded by the EPSRC Future Opportunity Ph.D. scholarships; ER was funded by the INCT BioSyn (National Institute of Science and Technology in Synthetic Biology), CNPq (National Council for Scientific and Technological Development), CAPES (Coordination for the Improvement of Higher Education Personnel), Brazilian Ministry of Health, and FAPDF (Research Support Foundation of the Federal District), Brazil.

## ACKNOWLEDGMENTS

This work captures discussions between participants at the “Computer-Aided Whole-Cell Design and Engineering” Workshop held on the 02-03 July 2019 at the University of Bristol, United Kingdom, and funded by the Engineering and Physical Sciences Research Council (EPSRC) within the remit of the Big Ideas initiative. We sincerely thank Dr. Kathleen Sedgley for her support with the workshop organization, and Dr. Thomas Gorochowski for participating in discussions.

## REFERENCES

- Alber, M., Buganza Tepole, A., Cannon, W. R., De, S., Dura-Bernal, S., Garikipati, K., et al. (2019). Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ Digit. Med.* 2:115.
- Ananthasubramaniam, B., Schmal, C., and Herzog, H. (2020). Amplitude effects allow short jet lags and large seasonal phase shifts in minimal clock models. *J. Mol. Biol.* 432, 3722–3737. doi: 10.1016/j.jmb.2020.01.014
- Ander, M., Beltrao, P., Di Ventura, B., Ferkinghoff-Borg, J., Foglierini, M., Kaplan, A., et al. (2004). SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. *Syst. Biol.* 1, 129–138. doi: 10.1049/sb:20045017
- Anderson, J., and Papachristodoulou, A. (2009). On validation and invalidation of biological models. *BMC Bioinform.* 10:132. doi: 10.1186/s12918-017-0484-132
- Ando, T., and Skolnick, J. (2010). Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18457–18462. doi: 10.1073/pnas.1011354107
- Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J. A., and Blom, J. G. (2009). Systems biology: parameter estimation for biochemical models. *FEBS J.* 276, 886–902. doi: 10.1111/j.1742-4658.2008.06844.x
- Babbie, A. C., and Stumpf, M. P. H. (2017). How to deal with parameters for whole-cell modelling. *J. R. Soc. Interf.* 14:237.
- Barberis, M., Linke, C., Adrover, M. A., Gonzalez-Novo, A., Lehrach, H., Krobitsch, S., et al. (2012). Sic1 plays a role in timing and oscillatory behaviour of B-type cyclins. *Biotechnol. Adv.* 30, 108–130. doi: 10.1016/j.biotechadv.2011.09.004
- Bartley, B. A., Beal, J., Karr, J. R., and Strychalski, E. A. (2020). Organizing genome engineering for the gigabase scale. *Nat. Commun.* 11:689.
- Battogtokh, D., and Tyson, J. J. (2004). Bifurcation analysis of a model of the budding yeast cell cycle. *Chaos* 14, 653–661. doi: 10.1063/1.1780011
- Betts, M. J., and Russell, R. B. (2007). The hard cell: from proteomics to a whole cell model. *FEBS Lett.* 581, 2870–2876. doi: 10.1016/j.febslet.2007.05.062
- Bordbar, A., McCloskey, D., Zielinski, D. C., Sonnenschein, N., Jamshidi, N., and Palsson, B. O. (2015). Personalized whole-cell kinetic models of metabolism for discovery in genomics and pharmacodynamics. *Cell Syst.* 1, 283–292. doi: 10.1016/j.cels.2015.10.003
- Borkowski, O., Bricio, C., Murgiano, M., Rothschild-Mancinelli, B., Stan, G. B., and Ellis, T. (2018). Cell-free prediction of protein expression costs for growing cells. *Nat. Commun.* 9:1457.
- Borkowski, O., Ceroni, F., Stan, G. B., and Ellis, T. (2016). Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr. Opin. Microbiol.* 33, 123–130. doi: 10.1016/j.mib.2016.07.009
- Bouhaddou, M., Barrette, A. M., Stern, A. D., Koch, R. J., DiStefano, M. S., Riesel, E. A., et al. (2018). A mechanistic pan-cancer pathway model informed by multi-omics data interprets stochastic cell fate responses to drugs and mitogens. *PLoS Comput. Biol.* 14:e1005985. doi: 10.1371/journal.pcbi.1005985
- Bragagli, S., and Ray, O. (2015). “Nonmonotonic learning in large biological networks” in *Inductive Logic Programming. Lecture Notes in Computer Science*, Vol. 9046, eds J. Davis and J. Ramon (Cham: Springer).
- Breuer, M., Earnest, T. M., Merryman, C., Wise, K. S., Sun, L., Lynott, M. R., et al. (2019). Essential metabolism for a minimal cell. *eLife* 8:e36842.
- Burgard, A. P., Pharkya, P., and Maranas, C. D. (2003). OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84, 647–657. doi: 10.1002/bit.10803
- Burley, S. K., Kurisu, G., Markley, J. L., Nakamura, H., Velankar, S., Berman, H. M., et al. (2017). PDB-Dev: a prototype system for depositing integrative/hybrid structural models. *Structure* 25, 1317–1318. doi: 10.1016/j.str.2017.08.001
- Calero, P., and Nikel, P. I. (2019). Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* 12, 98–124. doi: 10.1111/1751-7915.13292
- Calhoun, K. A., and Swartz, J. R. (2005). Energizing cell-free protein synthesis with glucose metabolism. *Biotechnol. Bioeng.* 90, 606–613. doi: 10.1002/bit.20449
- Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C., and Collins, J. J. (2018). Next-Generation machine learning for biological networks. *Cell* 173, 1581–1592. doi: 10.1016/j.cell.2018.05.015
- Carrera, J., and Covert, M. W. (2015). Why build whole-cell models? *Trends Cell Biol.* 25, 719–722. doi: 10.1016/j.tcb.2015.09.004
- Carrera, J., Rodrigo, G., and Jaramillo, A. (2009). Model-based redesign of global transcription regulation. *Nucleic Acids Res.* 37:e38. doi: 10.1093/nar/gkp022
- Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., Midford, P. E., et al. (2020). The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res.* 48, D445–D453.
- Castellanos, M., Wilson, D. B., and Shuler, M. L. (2004). A modular minimal cell model: purine and pyrimidine transport and metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6681–6686. doi: 10.1073/pnas.0400962101
- Castiglione, F., Pappalardo, F., Bianca, C., Russo, G., and Motta, S. (2014). Modeling biology spanning different scales: an open challenge. *Biomed. Res. Int.* 2014:902545.
- Ceroni, F., and Ellis, T. (2018). The challenges facing synthetic biology in eukaryotes. *Nat. Rev. Mol. Cell Biol.* 19, 481–482. doi: 10.1038/s41580-018-0013-2
- Chalkley, O., Purcell, O., Grierson, C., and Marucci, L. (2019). The genome design suite: enabling massive in-silico experiments to design genomes. *bioRxiv* [Preprint]. doi: 10.1101/681270
- Cornish-Bowden, A., and Hofmeyr, J. H. (1991). MetaModel: a program for modelling and control analysis of metabolic pathways on the IBM PC and compatibles. *Comput. Appl. Biosci.* 7, 89–93. doi: 10.1093/bioinformatics/7.1.89
- Courtot, M., Juty, N., Knupfer, C., Waltemath, D., Zhukova, A., Dräger, A., et al. (2011). Controlled vocabularies and semantics in systems biology. *Mol. Syst. Biol.* 7:543. doi: 10.1038/msb.2011.77
- Danchin, A. (2012). Scaling up synthetic biology: do not forget the chassis. *FEBS Lett.* 586, 2129–2137. doi: 10.1016/j.febslet.2011.12.024
- Danino, T., Prindle, A., Kwong, G. A., Skalak, M., Li, H., Allen, K., et al. (2015). Programmable probiotics for detection of cancer in urine. *Sci. Transl. Med.* 7:289ra84. doi: 10.1126/scitranslmed.aaa3519
- Diesel, E., Schreiber, M., and van der Meer, J. R. (2009). Development of bacteria-based bioassays for arsenic detection in natural waters. *Anal. Bioanal. Chem.* 394, 687–693. doi: 10.1007/s00216-009-2785-x
- El Karoui, M., Hoyos-Flight, M., and Fletcher, L. (2019). Future trends in synthetic biology-a report. *Front. Bioeng. Biotechnol.* 7:175. doi: 10.3389/fbioe.2018.00175
- Erguler, K., and Stumpf, M. P. (2011). Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Mol. Biosyst.* 7, 1593–1602.
- Feig, M., Harada, R., Mori, T., Yu, I., Takahashi, K., and Sugita, Y. (2015). Complete atomistic model of a bacterial cytoplasm for integrating physics, biochemistry, and systems biology. *J. Mol. Graph. Model.* 58, 1–9. doi: 10.1016/j.jmgm.2015.02.004
- Feig, M., and Sugita, Y. (2013). Reaching new levels of realism in modeling biological macromolecules in cellular environments. *J. Mol. Graph. Model.* 45, 144–156. doi: 10.1016/j.jmgm.2013.08.017
- Feig, M., and Sugita, Y. (2019). Whole-cell models and simulations in molecular detail. *Annu. Rev. Cell Dev. Biol.* 35, 191–211. doi: 10.1146/annurev-cellbio-100617-062542
- Fernandez-Castane, A., Feher, T., Carbonell, P., Pauthenier, C., and Faulon, J. L. (2014). Computer-aided design for metabolic engineering. *J. Biotechnol.* 192(Pt B), 302–313.
- Gawand, P., Said Abukar, F., Venayak, N., Partow, S., Motter, A. E., and Mahadevan, R. (2015). Sub-optimal phenotypes of double-knockout mutants of *Escherichia coli* depend on the order of gene deletions. *Integr. Biol.* 7, 930–939. doi: 10.1039/c5ib00096c
- Gerard, C., Gonze, D., and Goldbeter, A. (2009). Dependence of the period on the rate of protein degradation in minimal models for circadian oscillations. *Philos. Trans. A Math. Phys. Eng. Sci.* 367, 4665–4683. doi: 10.1098/rsta.2009.0133
- Gerard, C., Tyson, J. J., Coudreuse, D., and Novak, B. (2015). Cell cycle control by a minimal Cdk network. *PLoS Comput. Biol.* 11:e1004056. doi: 10.1371/journal.pone.0004056
- Gerard, C., Tyson, J. J., and Novak, B. (2013). Minimal models for cell-cycle control based on competitive inhibition and multisite phosphorylations of Cdk substrates. *Biophys. J.* 104, 1367–1379. doi: 10.1016/j.bpj.2013.02.012
- Glass, J. I., Merryman, C., Wise, K. S., Hutchison, C. A. III, and Smith, H. O. (2017). Minimal Cells-Real and imagined. *Cold Spring Harb. Perspect. Biol.* 9:a023861. doi: 10.1101/cshperspect.a023861



- Goldberg, A. P., Szigeti, B., Chew, Y. H., Sekar, J. A., Roth, Y. D., and Karr, J. R. (2018). Emerging whole-cell modeling principles and methods. *Curr. Opin. Biotechnol.* 51, 97–102. doi: 10.1016/j.copbio.2017.12.013
- Goldbeter, A. (1991). A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9107–9111. doi: 10.1073/pnas.88.20.9107
- Gomide, M. S., Sales, T. T., Barros, L. R. C., Limia, C. G., de Oliveira, M. A., Florentino, L. H., et al. (2020). Genetic switches designed for eukaryotic cells and controlled by serine integrases. *Commun. Biol.* 3:255.
- Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. (1999). From molecular to modular cell biology. *Nature* 402(Suppl.), C47–C52.
- Hicks, M., Bachmann, T. T., and Wang, B. (2020). Synthetic biology enables programmable cell-based biosensors. *Chemphyschem* 21:131. doi: 10.1002/cphc.201901191
- Hirokawa, Y., Kawano, H., Tanaka-Masuda, K., Nakamura, N., Nakagawa, A., Ito, M., et al. (2013). Genetic manipulations restored the growth fitness of reduced-genome *Escherichia coli*. *J. Biosci. Bioeng.* 116, 52–58. doi: 10.1016/j.jbiosc.2013.01.010
- Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., et al. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524–531.
- Hutchison, C. A. III, Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351:aad6253.
- Hyeon, C., and Thirumalai, D. (2011). Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* 2:487.
- Iwadate, Y., Honda, H., Sato, H., Hashimoto, M., and Kato, J. (2011). Oxidative stress sensitivity of engineered *Escherichia coli* cells with a reduced genome. *FEMS Microbiol. Lett.* 322, 25–33. doi: 10.1111/j.1574-6968.2011.02331.x
- Iyengar, S. (2011). *Symbolic Systems Biology: Theory and Methods*. Burlington, MA: Jones and Bartlett Learning.
- Jessop-Fabre, M. M., and Sonnenschein, N. (2019). Improving reproducibility in synthetic biology. *Front. Bioeng. Biotechnol.* 7:18. doi: 10.3389/fbioe.2018.0018
- Karr, J. R., Sanghvi, J. C., Macklin, D. N., Arora, A., and Covert, M. W. (2013). WholeCellKB: model organism databases for comprehensive whole-cell models. *Nucleic Acids Res.* 41, D787–D792.
- Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B. Jr., et al. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell* 150, 389–401. doi: 10.1016/j.cell.2012.05.044
- Karr, J. R., Takahashi, K., and Funahashi, A. (2015a). The principles of whole-cell modeling. *Curr. Opin. Microbiol.* 27, 18–24. doi: 10.1016/j.mib.2015.06.004
- Karr, J. R., Williams, A. H., Zucker, J. D., Raue, A., Steiert, B., Timmer, J., et al. (2015b). Summary of the DREAM8 parameter estimation challenge: toward parameter identification for whole-cell models. *PLoS Comput. Biol.* 11:e1004096. doi: 10.1371/journal.pone.1004096
- Karzbrun, E., Shin, J., Bar-Ziv, R. H., and Noireaux, V. (2011). Coarse-grained dynamics of protein synthesis in a cell-free system. *Phys. Rev. Lett.* 106:048104.
- King, Z. A., Lu, J., Drager, A., Miller, P., Federowicz, S., Lerman, J. A., et al. (2016). BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* 44, D515–D522.
- Kirk, P., Thorne, T., and Stumpf, M. P. (2013). Model selection in systems and synthetic biology. *Curr. Opin. Biotechnol.* 24, 767–774. doi: 10.1016/j.copbio.2013.03.012
- Kitano, H. (2002). Computational systems biology. *Nature* 420, 206–210.
- Koch, M., Faulon, J.-L., and Borkowski, O. (2018). Models for cell-free synthetic biology: make prototyping easier, better, and faster. *Front. Bioeng. Biotechnol.* 6:182. doi: 10.3389/fbioe.2018.00182
- Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y. A., Williams, E., et al. (2015). Array Express update—simplifying data submissions. *Nucleic Acids Res.* 43, D1113–D1116.
- Kotula, J. W., Kerns, S. J., Shakel, L. A., Siraj, L., Collins, J. J., Way, J. C., et al. (2014). Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4838–4843. doi: 10.1073/pnas.1321321111
- Kwiatkowska, M., Norman, G., and Parker, D. (eds) (2011). *PRISM 4.0: Verification of Probabilistic Real-Time Systems 2011*. Berlin: Springer.
- Landon, S., Rees-Garbutt, J., Marucci, L., and Grierson, C. (2019). Genome-driven cell engineering review: in vivo and in silico metabolic and genome engineering. *Essays Biochem.* 63, 267–284. doi: 10.1042/ebc20180045
- Le Novere, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., et al. (2009). The systems biology graphical notation. *Nat. Biotechnol.* 27, 735–741.
- Lee, J. M., Gianchandani, E. P., Eddy, J. A., and Papin, J. A. (2008). Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput. Biol.* 4:e1000086. doi: 10.1371/journal.pcbi.1000086.g002
- Lin, C., Jain, S., Kim, H., and Bar-Joseph, Z. (2017). Using neural networks for reducing the dimensions of single-cell RNA-Seq data. *Nucleic Acids Res.* 45:e156. doi: 10.1093/nar/gkx681
- Linke, C., Chasapi, A., Gonzalez-Novo, A., Al Sawad, I., Tognetti, S., Klipp, E., et al. (2017). A Clb/Cdk1-mediated regulation of Fkh2 synchronizes CLB expression in the budding yeast cell cycle. *NPJ Syst. Biol. Appl.* 3:7.
- Lu, T. K., Bowers, J., and Koeris, M. S. (2013). Advancing bacteriophage-based microbial diagnostics with synthetic biology. *Trends Biotechnol.* 31, 325–327. doi: 10.1016/j.tibtech.2013.03.009
- Ma, J., Yu, M. K., Fong, S., Ono, K., Sage, E., Demchak, B., et al. (2018). Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* 15, 290–298. doi: 10.1038/nmeth.4627
- Macklin, D. N., Ahn-Horst, T. A., Choi, H., Ruggero, N. A., Carrera, J., Mason, J. C., et al. (2020). Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation. *Science* 369:eaav3751. doi: 10.1126/science.aav3751
- Macklin, D. N., Ruggero, N. A., and Covert, M. W. (2014). The future of whole-cell modeling. *Curr. Opin. Biotechnol.* 28, 111–115. doi: 10.1016/j.copbio.2014.01.012
- Marucci, L., Barton, D. A., Cantone, I., Ricci, M. A., Cosma, M. P., Santini, S., et al. (2009). How to turn a genetic circuit into a synthetic tunable oscillator, or a bistable switch. *PLoS One* 4:e8083. doi: 10.1371/journal.pone.0008083
- Matsuura, T., Hosoda, K., and Shimizu, Y. (2018). Robustness of a reconstituted *Escherichia coli* protein translation system analyzed by computational modeling. *ACS Synth. Biol.* 7, 1964–1972. doi: 10.1021/acssynbio.8b00228
- McAdams, H. H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 94, 814–819.
- McCloskey, D., Palsson, B. O., and Feist, A. M. (2013). Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol. Syst. Biol.* 9:661. doi: 10.1038/msb.2013.18
- McGuffee, S. R., and Elcock, A. H. (2010). Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* 6:e1000694. doi: 10.1371/journal.pcbi.1000694
- Medley, J. K., Goldberg, A. P., and Karr, J. R. (2016). Guidelines for reproducibly building and simulating systems biology models. *IEEE Trans. Biomed. Eng.* 63, 2015–2020. doi: 10.1109/tbme.2016.2591960
- Mendoza, S. N., Olivier, B. G., Molenaar, D., and Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biol.* 20:158.
- Mol, M., Kabra, R., and Singh, S. (2018). Genome modularity and synthetic biology: engineering systems. *Prog. Biophys. Mol. Biol.* 132, 43–51. doi: 10.1016/j.pbiomolbio.2017.08.002
- Mondeel, T., Holland, P., Nielsen, J., and Barberis, M. (2019). ChIP-exo analysis highlights Fkh1 and Fkh2 transcription factors as hubs that integrate multi-scale networks in budding yeast. *Nucleic Acids Res.* 47, 7825–7841. doi: 10.1093/nar/gkz603
- Mondeel, T., Ivanov, O., Westerhoff, H. V., Liebermeister, W., and Barberis, M. (2020). Clb3-centered regulations are recurrent across distinct parameter regions in minimal autonomous cell cycle oscillator designs. *NPJ Syst. Biol. Appl.* 6:8.
- Moore, S. J., MacDonald, J. T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., et al. (2018). Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4340–E4349.
- Morton-Firth, C. J., and Bray, D. (1998). Predicting temporal fluctuations in an intracellular signalling pathway. *J. Theor. Biol.* 192, 117–128. doi: 10.1006/jtbi.1997.0651
- Moya, A., Gil, R., Latorre, A., Pereto, J., Pilar Garcillan-Barcia, M., and de la Cruz, F. (2009). Toward minimal bacterial cells: evolution vs. design. *FEMS Microbiol. Rev.* 33, 225–235. doi: 10.1111/j.1574-6976.2008.00151.x

- Mutturi, S. (2017). FOCuS: a metaheuristic algorithm for computing knockouts from genome-scale models for strain optimization. *Mol. Biosyst.* 13, 1355–1363. doi: 10.1039/c7mb00204a
- Noireaux, V., Bar-Ziv, R., and Libchaber, A. (2003). Principles of cell-free genetic circuit assembly. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12672–12677. doi: 10.1073/pnas.2135496100
- Noske, A. B., Costin, A. J., Morgan, G. P., and Marsh, B. J. (2008). Expedited approaches to whole cell electron tomography and organelle mark-up in situ in high-pressure frozen pancreatic islets. *J. Struct. Biol.* 161, 298–313. doi: 10.1016/j.jsb.2007.09.015
- Novak, B., and Tyson, J. J. (1993). Numerical analysis of a comprehensive model of M-phase control in *Xenopus oocyte* extracts and intact embryos. *J. Cell Sci.* 106(Pt 4), 1153–1168.
- Pandit, A. V., Srinivasan, S., and Mahadevan, R. (2017). Redesigning metabolism based on orthogonality principles. *Nat. Commun.* 8:15188.
- Pedone, E., Postiglione, L., Alicino, F., Rocca, D. L., Montes-Olivas, S., Khazim, M., et al. (2019). A tunable dual-input system for on-demand dynamic gene expression regulation. *Nat. Commun.* 10:4481.
- Perdikaris, P., and Karniadakis, G. E. (2016). Model inversion via multi-fidelity Bayesian optimization: a new paradigm for parameter estimation in haemodynamics, and beyond. *J. R. Soc. Interf.* 13:20151107. doi: 10.1098/rsif.2015.1107
- Posfai, G., Plunkett, G. III, Feher, T., Frisch, D., Keil, G. M., Umenhoffer, K., et al. (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science* 312, 1044–1046. doi: 10.1126/science.1126439
- Prescott, A. M., and Abel, S. M. (2017). Combining in silico evolution and nonlinear dimensionality reduction to redesign responses of signaling networks. *Phys. Biol.* 13:066015. doi: 10.1088/1478-3975/13/6/066015
- Prescott, T. P., Lang, M., and Papachristodoulou, A. (2015). Quantification of interactions between dynamic cellular network functionalities by cascaded layering. *PLoS Comput. Biol.* 11:e1004235. doi: 10.1371/journal.pone.1004235
- Price, M. N., Wetmore, K. M., Waters, R. J., Callaghan, M., Ray, J., Liu, H., et al. (2018). Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* 557, 503–509. doi: 10.1038/s41586-018-0124-0
- Purcell, O., Jain, B., Karr, J. R., Covert, M. W., and Lu, T. K. (2013). Towards a whole-cell modeling approach for synthetic biology. *Chaos* 23:025112. doi: 10.1063/1.4811182
- Purcell, O., Savery, N. J., Grierson, C. S., and di Bernardo, M. (2010). A comparative analysis of synthetic genetic oscillators. *J. R. Soc. Interf.* 7, 1503–1524. doi: 10.1098/rsif.2010.0183
- Rancati, G., Moffat, J., Typas, A., and Pavelka, N. (2018). Emerging and evolving concepts in gene essentiality. *Nat. Rev. Genet.* 19, 34–49. doi: 10.1038/nrg.2017.74
- Rand, D. A. (2008). Mapping global sensitivity of cellular network dynamics: sensitivity heat maps and a global summation law. *J. R. Soc. Interf.* 5(Suppl. 1), S59–S69.
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabasi, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555. doi: 10.1126/science.1073374
- Ray, O., Soh, T., and Inoue, K. (2011). “Analysing pathways using ASP-based approaches,” in *Proceedings of the 2010 Conference on Algebraic and Numeric Biology*, Berlin.
- Rees-Garbutt, J., Chalkley, O., Landon, S., Purcell, O., Marucci, L., and Grierson, C. (2020). Designing minimal genomes using whole-cell models. *Nat. Commun.* 11:836.
- Reuss, D. R., Altenbuchner, J., Mader, U., Rath, H., Ischebeck, T., Sappa, P. K., et al. (2017). Large-scale reduction of the *Bacillus subtilis* genome: consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res.* 27, 289–299. doi: 10.1101/gr.215293.116
- Rozanski, R., Ray, O., King, R., and Bragaglia, S. (2015). “Automating development of metabolic network models,” in *Computational Methods in Systems Biology. CMSB 2015. Lecture Notes in Computer Science*, Vol. 9308, eds O. Roux and J. Bourdon (Cham: Springer).
- Sajed, T., Marcu, A., Ramirez, M., Pon, A., Guo, A. C., Knox, C., et al. (2016). ECMDDB 2.0: A richer resource for understanding the biochemistry of *E. coli*. *Nucleic Acids Res.* 44, D495–D501.
- Sali, A., Berman, H. M., Schwede, T., Trewthella, J., Kleywegt, G., Burley, S. K., et al. (2015). Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure* 23, 1156–1167. doi: 10.1016/j.str.2015.05.013
- Shu, J., and Shuler, M. L. (1989). A mathematical model for the growth of a single cell of *E. coli* on a glucose/glutamine/ammonium medium. *Biotechnol. Bioeng.* 33, 1117–1126. doi: 10.1002/bit.260330907
- Siegel-Gaskins, D., Tuza, Z. A., Kim, J., Noireaux, V., and Murray, R. M. (2014). Gene circuit performance characterization and resource usage in a cell-free “breadboard”. *ACS Synth. Biol.* 3, 416–425. doi: 10.1021/sb400203p
- Silverman, A. D., Karim, A. S., and Jewett, M. C. (2020). Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet.* 21, 151–170. doi: 10.1038/s41576-019-0186-3
- Singla, J., McClary, K. M., White, K. L., Alber, F., Sali, A., and Stevens, R. C. (2018). Opportunities and challenges in building a spatiotemporal multi-scale model of the human pancreatic beta cell. *Cell* 173, 11–19. doi: 10.1016/j.cell.2018.03.014
- Slomovic, S., Pardee, K., and Collins, J. J. (2015). Synthetic biology devices for in vitro and in vivo diagnostics. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14429–14435. doi: 10.1073/pnas.1508521112
- Smucker, B., Krzywinski, M., and Altman, N. (2018). Optimal experimental design. *Nat. Methods* 15, 559–560.
- Solernou, A., Hanson, B. S., Richardson, R. A., Welch, R., Read, D. J., Harlen, O. G., et al. (2018). Fluctuating finite element analysis (FFEA): a continuum mechanics software tool for mesoscale simulation of biomolecules. *PLoS Comput. Biol.* 14:e1005897. doi: 10.1371/journal.pone.1005897
- Stogbauer, T., Windhager, L., Zimmer, R., and Radler, J. O. (2012). Experiment and mathematical modeling of gene expression dynamics in a cell-free system. *Integr. Biol.* 4, 494–501.
- Szigei, B., Roth, Y. D., Sekar, J. A. P., Goldberg, A. P., Pochiraju, S. C., and Karr, J. R. (2018). A blueprint for human whole-cell modeling. *Curr. Opin. Syst. Biol.* 7, 8–15. doi: 10.1016/j.coisb.2017.10.005
- Takahashi, K., Arjunan, S. N., and Tomita, M. (2005). Space in systems biology of signaling pathways—towards intracellular molecular crowding in silico. *FEBS Lett.* 579, 1783–1788. doi: 10.1016/j.febslet.2005.01.072
- Takahashi, M. K., Chappell, J., Hayes, C. A., Sun, Z. Z., Kim, J., Singhal, V., et al. (2015). Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription-translation (TX-TL) systems. *ACS Synth. Biol.* 4, 503–515. doi: 10.1021/sb400206c
- Tang, P. W., Chua, P. S., Chong, S. K., Mohamad, M. S., Choon, Y. W., Deris, S., et al. (2015). A review of gene knockout strategies for microbial cells. *Recent. Pat. Biotechnol.* 9, 176–197. doi: 10.2174/1872208310666160517115047
- Thiele, I., Jamshidi, N., Fleming, R. M., and Palsson, B. O. (2009). Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput. Biol.* 5:e1000312. doi: 10.1371/journal.pcbi.10000312
- Thul, P. J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., et al. (2017). A subcellular map of the human proteome. *Science* 356:6340.
- Tomazou, M., Barahona, M., Polizzi, K. M., and Stan, G. B. (2018). Computational Re-design of synthetic genetic oscillators for independent amplitude and frequency modulation. *Cell Syst.* 6:50.
- Tomita, M. (2001). Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol.* 19, 205–210. doi: 10.1016/s0167-7799(01)01636-5
- Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T. S., Matsuzaki, Y., Miyoshi, F., et al. (1999). E-CELL: software environment for whole-cell simulation. *Bioinformatics* 15, 72–84. doi: 10.1093/bioinformatics/15.1.72
- Tyson, J. J. (1991). Modeling the cell division cycle: cdc2 and cyclin interactions. *Proc. Natl. Acad. Sci. U.S.A.* 88, 7328–7332. doi: 10.1073/pnas.88.16.7328
- Underwood, K. A., Swartz, J. R., and Puglisi, J. D. (2005). Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. *Biotechnol. Bioeng.* 91, 425–435. doi: 10.1002/bit.20529
- UniProt Consortium, T. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 46, 2699. doi: 10.1093/nar/gky092
- van der Meer, J. R., and Belkin, S. (2010). Where microbiology meets microengineering: design and applications of reporter bacteria. *Nat. Rev. Microbiol.* 8, 511–522. doi: 10.1038/nrmicro2392
- van der Zee, L., and Barberis, M. (2019). Advanced modeling of cellular proliferation: toward a multi-scale framework coupling cell cycle to metabolism

- by integrating logical and constraint-based models. *Methods Mol. Biol.* 2049, 365–385. doi: 10.1007/978-1-4939-9736-7\_21
- Varma, A., and Palsson, B. O. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* 60, 3724–3731. doi: 10.1128/aem.60.10.3724-3731.1994
- Vernon, I., Liu, J., Goldstein, M., Rowe, J., Topping, J., and Lindsey, K. (2018). Bayesian uncertainty analysis for complex systems biology models: emulation, global parameter searches and evaluation of gene functions. *BMC Syst. Biol.* 12:1. doi: 10.1186/s12918-017-0484-3
- Vilkhovoy, M., Horvath, N., Shih, C. H., Wayman, J. A., Calhoun, K., Swartz, J., et al. (2018). Sequence specific modeling of *E. coli* cell-free protein synthesis. *ACS Synth. Biol.* 7, 1844–1857. doi: 10.1021/acssynbio.7b00465
- Vyshemirsky, V., and Girolami, M. (2008). BioBayes: a software package for bayesian inference in systems biology. *Bioinformatics* 24, 1933–1934. doi: 10.1093/bioinformatics/btn338
- Waltemath, D., Karr, J. R., Bergmann, F. T., Chelliah, V., Hucka, M., Krantz, M., et al. (2016). Toward community standards and software for whole-cell modeling. *IEEE Trans. Biomed. Eng.* 63:14.
- Wang, L., and Maranas, C. D. (2018). MinGenome: an in silico top-down approach for the synthesis of minimized genomes. *ACS Synth. Biol.* 7, 462–473. doi: 10.1021/acssynbio.7b00296
- Way, J. C., Collins, J. J., Keasling, J. D., and Silver, P. A. (2014). Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell* 157, 151–161. doi: 10.1016/j.cell.2014.02.039
- Wilkinson, D. J. (2007). Bayesian methods in bioinformatics and computational systems biology. *Brief Bioinform.* 8, 109–116. doi: 10.1093/bib/bbm007
- Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., et al. (2012). SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Res.* 40, D790–D796.
- Woolston, B. M., Edgar, S., and Stephanopoulos, G. (2013). Metabolic engineering: past and future. *Annu. Rev. Chem. Biomol. Eng.* 4, 259–288. doi: 10.1146/annurev-chembioeng-061312-103312
- Xu, X., Liu, Y., Du, G., Ledesma-Amaro, R., and Liu, L. (2020). Microbial chassis development for natural product biosynthesis. *Trends Biotechnol.* 38, 779–796. doi: 10.1016/j.tibtech.2020.01.002
- Yilmaz, L. S., and Walhout, A. J. (2017). Metabolic network modeling with model organisms. *Curr. Opin. Chem. Biol.* 36, 32–39. doi: 10.1016/j.cbpa.2016.12.025
- Yu, I., Mori, T., Ando, T., Harada, R., Jung, J., Sugita, Y., et al. (2016). Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *eLife* 5:e19274.
- Yu, M. K., Ma, J., Fisher, J., Kreisberg, J. F., Raphael, B. J., and Ideker, T. (2018). Visible machine learning for biomedicine. *Cell* 173, 1562–1565. doi: 10.1016/j.cell.2018.05.056
- Zhou, J., Wu, R., Xue, X., and Qin, Z. (2016). CasHRA (Cas9-facilitated homologous recombination assembly) method of constructing megabase-sized DNA. *Nucleic Acids Res.* 44, e124. doi: 10.1093/nar/gkw475

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Citation:* Marucci L, Barberis M, Karr J, Ray O, Race PR, de Souza Andrade M, Grierson C, Hoffmann SA, Landon S, Rech E, Rees-Garbutt J, Seabrook R, Shaw W and Woods C (2020) Computer-Aided Whole-Cell Design: Taking a Holistic Approach by Integrating Synthetic With Systems Biology. *Front. Bioeng. Biotechnol.* 8:942. doi: 10.3389/fbioe.2020.00942

Copyright © 2020 Marucci, Barberis, Karr, Ray, Race, de Souza Andrade, Grierson, Hoffmann, Landon, Rech, Rees-Garbutt, Seabrook, Shaw and Woods. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.