12-2021

Charting a New Frontier Integrating Mathematical Modeling in Complex Biological Systems from Molecules to Ecosystems

Katharine A. White
University of Notre Dame, kwhite6@nd.edu

Kira D. McEntire
Queens University of Charlotte, mcentirek@queens.edu

Nicole R. Buan
University of Nebraska-Lincoln, nbuan2@unl.edu

Lecia Robinson
Tuskegee University

Elisar Barbar
Oregon State University, Oregon, barbare@oregonstate.edu

Published in *Integrative and Comparative Biology* 61:6 (December 2021), pp. 2255–2266; doi: 10.1093/icb/icab165

Submitted March 15, 2021; revised July 9, 2021; accepted July 15, 2021; published online July 20, 2021.

Charting a New Frontier Integrating Mathematical Modeling in Complex Biological Systems from Molecules to Ecosystems

Katharine A. White,[1] Kira D. McEntire,[2] Nicole R. Buan,[3] Lecia Robinson,[4] and Elisar Barbar[5]

1. Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana, USA
2. Biology Department, Queens University of Charlotte, Charlotte, North Carolina, USA
3. Department of Biochemistry, University of Nebraska–Lincoln, Lincoln, Nebraska, USA
4. Department of Biology, Tuskegee University, Tuskegee, Alabama, USA
5. Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon, USA

Corresponding author – Katharine A. White, email kwhite6@nd.edu

Abstract

Advances in quantitative biology data collection and analysis across scales (molecular, cellular, organismal, and ecological) have transformed how we understand, categorize, and predict complex biological systems. This surge of quantitative data creates an opportunity to apply, develop, and evaluate mathematical models of biological systems and explore novel methods of analysis. Simultaneously, thanks to increased computational power, mathematicians, engineers, and physical scientists have developed sophisticated models of biological systems at different scales. Novel modeling schemes can offer deeper understanding of principles in biology, but there is still a disconnect between modeling and experimental biology that limits our ability to fully realize the integration of mathematical modeling and biology. In this work, we explore the urgent need to expand the use of existing mathematical models across biological scales, develop models that are robust to biological heterogeneity, harness feedback loops within the iterative-modeling process, and nurture a cultural shift toward interdisciplinary and cross-field interactions. Better integration of biological experimentation

and robust mathematical modeling will transform our ability to understand and predict complex biological systems.

Critical time for biological modeling

Biological systems are staggeringly complex. A critical goal of biological research is to untangle this complexity and make predictions about biological systems. Advances in quantitative biology data collection and analysis across scales (molecular, cellular, organismal, and ecological) have transformed how we understand, categorize, and predict complex biological systems. From ecology to single-cell measurements to single-molecule imaging, we can now observe fluctuations in biological data that are intrinsic to the system, may hold key information, and may be biologically meaningful. This surge of quantitative data coincides with increased computational power, creating a unique opportunity to better apply, develop, and evaluate mathematical models of complex biological systems at different scales. *In the context of this paper, we use "model" to refer to either mathematical representations or predictions of biological systems.*

All biological subdisciplines could benefit tremendously from systematic integration of theoretical mathematical modeling approaches and biology. For example, mathematical models can allow biologists to decouple sources of nonbiological noise or experimental error from meaningful biological variability, which would be transformative for uncovering biological roles for stochasticity and heterogeneity. With more resources dedicated to integration of biology and mathematical modeling, we envision a transformational improvement in our ability to both describe and predict complex biology (from molecules to organisms to systems) (Westerhoff and Palsson 2004). To fully realize the integration of mathematical modeling and biological systems,work in four key areas must be emphasized and supported (Fig. 1):

(1) Cross-scales approaches: Mathematical models that can readily scale to other biological systems would be transformative in creating common language to facilitate understanding between fields. Identifying models and biological systems to develop in depth as "anchor" models and systems for validation and characterization is a critical goal.
(2) Increasing complexity: The field needs mathematical models that are robust to complexity within representative biological systems and that can predict accurately how perturbations alter those systems. This would allow biologists to better identify new variables affecting biological outcomes, predict complex biological systems in a rapidly changing world, and even generate entirely virtual biological datasets when samples are scarce.
(3) Iterative feedback loops: Not all biological data is collected in such a way that enables use in model development or validation. Collecting data with a mind on where and how it will be used in modeling is necessary to better integrate biology and mathematical modeling. Furthermore, applying predictive models to direct or identify research

questions would be a long-term payoff of investing in development of robust iterative feedback loops.

(4) Overcome cultural barriers: The current culture of science encourages researchers who work in discipline-specific silos, often to the detriment of research advances. The most effective route to overcome cultural barriers includes supporting interdisciplinary work with long-timeline funding initiatives, providing resources for interdisciplinary curriculum, and providing training initiatives for scientists interested in working at the interdisciplinary interface between mathematics and biology.

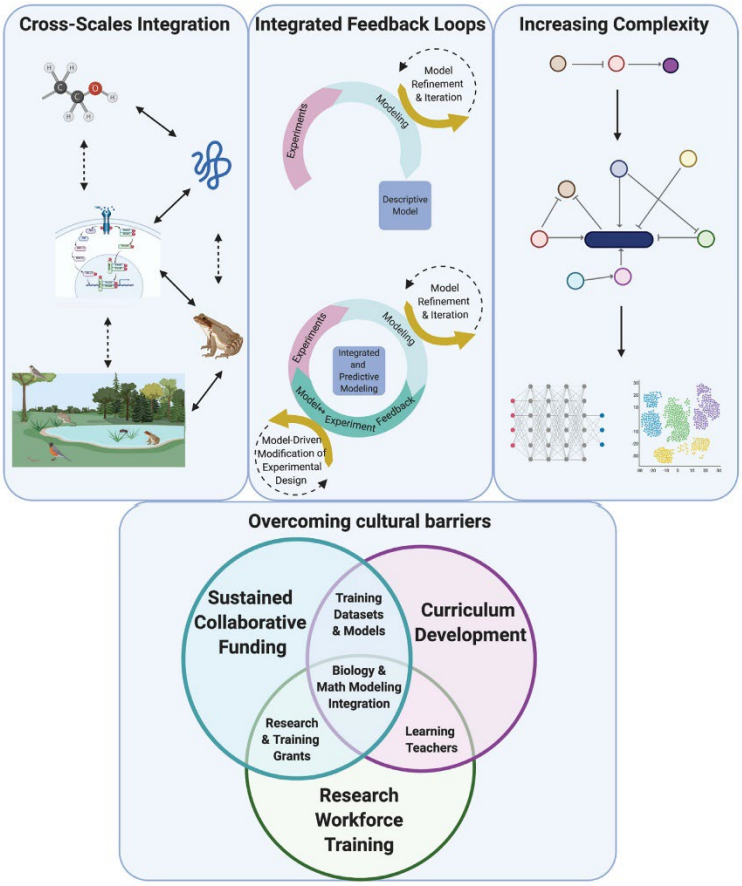

Figure 1. Frontiers for integrating mathematical modeling and complex biological systems. In order to fully realize the integration of mathematical modeling and biological systems, work in four key areas must be emphasized and supported. Cross-scales Integration: Development and validation of mathematical models that can readily scale to other biological systems. Integrated Feedback Loops: Leverage not only model refinement and iteration but also model-driven experimental design to obtain sound, parameterized data for model development. Increasing Complexity: Development of models that are robust to increasingly complex biological systems. Overcoming Cultural Barriers:We must support the integration of mathematical modeling and biology through integrated research workforce training, sustained collaborative funding, and curriculum development.

Achieving just a subset of these goals would transform our ability to understand and predict complex biological systems.We will explore examples and ongoing work needed within each of these goals below.

Cross-scales approaches

Siloing of biology means that some mathematical models that are very well known in certain subfields are often unknown beyond their immediate sphere of application. There are a few significant exceptions that demonstrate the broad applicability of mathematical models across biological scales. Here, we present examples of foundational models developed at the organismal- and ecological-scale (Lotka-Volterra competition models and Hardy-Wienberg population genetics models) that have been successfully applied across biological scales to define, describe, and predict biological outcomes. Then we will outline a few examples of underutilized mathematical models, more common to one biological level than another, and propose how they may be applied in other biological contexts.

Ecologists have used mathematics to describe phenomena for decades, including the Lotka-Volterra competition models (Lotka 1925; Volterra 1926) and the Hardy-Weinberg equilibrium of population genetics (Hardy 1908; Edwards 2008). Because competition for resources takes place at all levels of biology, competition models originally used to describe organismal (predator-prey) competition have been readily adapted to describe chemotaxis in slime molds (Mizukami and Winkler 2017), cancer cell competition (Goyal, Bhardwaj and Prakash 2021), competition for light between trees (Magal and Zhang 2017), and evolution (Aristide and Morlon 2019). In addition to being applied across multiple biological scales, competition models have also been modified to incorporate biological stochasticity (Zhu and Yin 2009) and adapted to accommodate increased biological complexity including interference (Hsu 1982), seasonal succession (Hsu and Zhao 2012), and the changing speed of invasive species movement (Hosono 1998). Competition models are actively used in mathematics to look at the dynamics of the models themselves (introduced in Baigent (2010), e.g., Nathan Ngoteya (2015)), and even to grapple with the challenge of evolutionary factors (Zhang and Lam 2013; Afraimovich et al. 2008).

While other mathematicians and biologists alike have moved to dynamic systems models, the early Lotka-Voltera model is a great example of how biological models can be adapted cross-scales and to model and predict biological systems with increasing complexity. Another classic cross-scales mathematical model adaptation can be found in disease transmission models. For example, SIR models (susceptible-infected-recovered), which have been successfully used to predict both human and animal disease spread (e.g., Almaraz, Gómez-Corral and Rodríguez-Bernal 2016), have also been adapted to model nonbiological systems, such as predicting the economic implications of the COVID-19 pandemic (Ellison 2020). However, while these examples demonstrate cross-scales adaptation of some models, other models can fail when crossing phylogenetic lineages or are not easily integrated with heterogeneous genomic, signaling, and environmental data (LeNovère 2015).

Agent-based models (ABMs) are commonly used in ecology (reviewed in Willem et al. 2017) and are simulation models that offer a way to test our understanding of latent

mechanisms driving visible patterns on many scales. ABMs can have both mobile and static "agents" whose interactions the programmer controls. This allows simulation of organisms (or cells or molecules) and their surrounding environments to understand the drivers of a biological pattern. For example, some patterns might include the local abundance of a species or the way cancerous tumors grow. Since one of the purposes of ABMs is to model latent properties of a system resulting in an observable phenomenon, they can be used at any biological scale, though they have been most prominently used in ecology (Grimm 1999; DeAngelis and Grimm 2014) and disease dynamics (reviewed in Willem et al. 2017).While still underutilized compared to the cross-scales competition models described above, ABMs are starting to be applied in other contexts from modeling cellular membrane formation to evolution (Rangel et al. 2018) and ecosystem processes (Grimm, Ayllón and Railsback 2017). See Models Library of built-in models in the program NetLogo (Wilensky and Resnick 1999).

Since ABMs are a bottom-up approach to scientific questions, one can model physiological or cellular processes that could be the drivers of behavior or even ultimate drivers of population dynamics (e.g., McEntire and Maerz 2019; Sears and Angilletta 2015). In addition to revealing latent properties of a system, the ABM modeling process also often highlights gaps in our understanding of a system. When modeling in a bottom-up approach, one may discover there is a crucial aspect of the system we do not understand. Variance between model predictions and experimental observations reveal latent processes, factors, or parameters that are yet to be studied in detail; thus indicating where future studies should be initiated, often at a different biological level. For example, when modeling salamander activity time based on biophysical models and local weather conditions, the model suggested the salamanders should be active more often and for much longer than is observed in the field (McEntire andMaerz 2019). Further studies and field observations may clarify this discrepancy between the mathematical model and true salamander behavior. Additional studies on an alternative physiological function may instead offer a better mechanism driving salamander activity patterns. Furthermore, we know the model referenced above is not completely capturing the drivers of plant climbing behavior, as field observations recorded climbing outside themodel rule (unpublished data). ABMs are limited by known information, but need to be carefully curated to ensure they are parsimonious.

Not only can these models be applied in multiple contexts, ABMs can be used to *simultaneously* model multiple levels of biology. Many ABMs embed mathematical models or population estimation models as a part of the modeling process. This offers an opportunity to simultaneously model systems that we know are interconnected. For example, when modeling the usage of burned habitats by turkeys, we used descriptive models based on known movement patterns to estimate the probability of movement into a burned area despite being unable to exactly understand an individual turkey's decision-making process (Sullivan et al. 2020). ABMs can thus be used to ask increasingly complex questions or include more variation. However, ABMs are limited by being data demanding, making technological advances all the more important. Another limitation to the wide adoption of versatile models (including ABMs) is that researchers must first know what models are being used at other systems/scales, and must also have the opportunity to collaborate to

apply new modeling approaches to other systems/scales. Such cross-scale adoption of models could provide insights and alternative viewpoints to increase understanding of a wider range of biological systems.

Cross-scale adoption is challenging because it is difficult to keep up with techniques from other fields, and it can be a struggle to understand how to implement a new modeling type. To aid in cross-scales adoption of models, we suggest first determining the type of model needed to answer the research question (such as conceptual, statistical, predictive, and so on), then identify whether a top-down or a bottom-up approach is desired. For example, a scientist interested in adopting ABMs to their question would do well to consider the pros and cons of ABM for their system. They might consider that many traditional models commonly used in ecological studies provide information about a system based on collected data while ABMs can use this data to test their outcomes or as parameters for how the model behaves. Second, ABMs are limited by the data available, which sometimes is difficult to measure.However, ABMs do offer the ability to include stochastic factors and behavior and can also be used to find gaps in knowledge or to estimate latent properties of a system. Once an appropriate modeling strategy has been determined, it is encouraged to consult papers that use the same strategy (regardless of the application/specialty) to see what alternative methodologies are used. Inevitably, applying multiple types of models is essential for understanding and investigating complex biological systems.

Increasing complexity

A challenge in biology is how to build mathematical models that maintain efficacy as they are adapted to describe our increasingly complex understanding of biology. Here, we would like to make a distinction between "complexity" and "complicatedness" following Sun et al. 2016, who distinguished the two as referring to model behavior (complexity) and model structure (complicatedness). When we continue to use the word complex in this paper, we are referring to biological complexity rather than model complicatedness, as the latter can be unhelpful when trying to understand a multifaceted system. Overcomplicated models at any level are unhelpful, and models need to be just complicated enough to appropriately model biological complexity of the system being studied.

At the molecular scale, computational modeling approaches are increasingly important for describing complex cellular function based on physical principles, especially in the field of microbial metabolism (Keseler et al. 2013).The same is true for ecological systems where physical principles can be incredibly important for understanding and making predictions about species interactions with each other and the rapidly changing environment (e.g., Peterman and Gade 2017; Riddell et al. 2017; McEntire and Maerz 2019). Early models of microbial function progressed from strain characterization and phenotyping (Breed, Murray and Hitchens 1944) to the Central Dogma of molecular biology (Crick 1970) that related hereditary information with biochemical function, to genome-scale metabolic models (Duarte 2004) and network models (Perez-Garcia, Lear and Singhal 2016) that describe relationships between organisms. In isolation, these models accurately describe and, in some cases, predict cellular function. For example, the fields of metabolic engineering and synthetic biology are predicated on past successes transporting molecular and biochemical modules

from one organism to another with reliable outcomes. However, we cannot yet predict microbial metabolism from genomic data without extensive experimental curation. Furthermore, modeling multispecies microbial communities is in its infancy, and we are far from being able to forecast an organism's evolutionary trajectory in silico. Though it is possible some of the techniques proposed in the previous section on adapting cross-scale models may help push this idea further.

If we could accurately predict species or community metabolism from genome information, we could rapidly accelerate work focused on the design and growth of natural microbiomes to inhibit pathogens and increase plant, animal, or human health (Ainsworth 2020), or bacteria to clean the environment of pollutants (Ojuederie and Babalola 2017), to name a few examples. Research has shown that microbes interact in complex ways with host organisms and with each other; they can be either harmful or helpful to the plant or animal host depending on environmental and nutritional parameters (Lewin-Epstein and Hadany 2020). Accurately predicting species or community metabolism from genome information could also be critical to conservation efforts when trying to make predictive models about species that are difficult to find, isolate, and/or culture. If it were possible to determine metabolism of any organism from environmental DNA, it could offer better models of habitat and spatial distribution. One example is the bacterium *Escherichia coli* (*E. coli*), an inhabitant of the gastrointestinal tract that benefits the host by synthesizing vitamin cofactors and by contributing to a hostile environment for incoming pathogenic microbes (Cardinale, Joachimiak and Arkin 2013). When sufficient technical prowess and experimental data are available, organisms like *E. coli* can be engineered to attack pathogens (Kurtz et al. 2019) or to deliver drug therapeutics to the host (Claesen and Fischbach 2015). Many microbes are genetically tractable, and researchers at private, academic, and federal research laboratories are engaged in trying to understand how to engineer a wider variety of organisms and how to use cooperative behaviors of microbes for human benefit (Freed et al. 2018). Tremendous resources are being invested by every US federal research agency as we seek to improve the health and function of biological systems at every level.

The challenge in modeling organism function, whether they be microbial or macroscale, is that the biological information-environment space is immense. As foreshadowed in the Central Dogma, accurate description of cellular function requires integrating multiple data types. For instance, to predict the function of a microbe in a community requires knowledge of the environment (temperature, pH, nutrients, and their concentrations; Isaac Newton Institute Fellows et al. 2016), the genetic information carried by the cell (Wu et al. 2009; which may include such things as plasmids, lysogenized viruses, or transposons, etc.), patterns of gene expression and coregulation of suites of genes, the function of the proteins and enzymes encoded by the genes (Bergthorsson, Andersson and Roth 2007), and the probability for genetic exchange (de la Cruz and Davies 2000; Oren et al. 2014).

Population-, community-, or landscape-level models have similar challenges of complex environments and interactions. For practical technical reasons, experiments are often carried out by reducing system complexity to the point that researchers (often undergraduate and graduate students) can observe an unequivocal binary response. For example, in microbial systems we can reduce the complexity by focusing on single homogenous populations of cells under defined culture conditions over short time scales to avoid the

"problem" of evolution (Großkopf et al. 2016) that is observed in longer-term experiments. Similarly in ecological systems, we may reduce questions to a binary of community composition in urban or suburban environments.

Through various experimental approaches, we have made tremendous advances in understanding biological systems from molecular and biochemical functions of discrete metabolic systems to broad scale ecological patterns. However, it is challenging to integrate modeling data and various types of experimental observations. For instance, just because a gene is present in a genome does not mean that it is expressed, and gene expression does not always correlate with protein expression level (Gygi et al. 1999; dos Reis 2003; Colin, Libri and Porrua 2011), resulting in phenotype heterogeneity in a population. As a consequence, researchers familiar with methodologies that produce discrete quantifiable results are sometimes uncomfortable extrapolating into more complex systems research in which results are more often expressed as statistical probability outcomes. However, in order to understand or predict biological function we must develop models that integrate across the information-environment space and can scale with the complexity of the biological system.

A systematic approach is needed to explore the vast biological information-environment space and a community of researchers with organism and system-specific expertise who can seamlessly collaborate to develop new software for modeling in a common platform. The Department of Energy Knowledgebase (KBase; Arkin et al. 2018) is aiming to serve the need for a metabolic modeling "playground" where researchers iteratively explore modeling methods, develop new software applications, refine, and propagate successes to create a platform from which useful software tools can emerge. KBase allows convergence between biology and computer science researchers, allowing them to experiment with scale and complexity to create application workflows and software systems that incorporate increasingly sophisticated machine learning tools grounded in biological and physical theory (Suthers et al. 2021). For instance, constraint-based metabolic modeling that takes chemical mass balance equations and cellular energetics into account is being applied to genome scale models to predict microbial growth and metabolism (Bordbar et al. 2014). Methods from software testing (Cashman et al. 2017) can be appliedwith success to statistically sample the biological information-environment space to reduce experiment time and cost, and concepts from information theory (Sakkaff et al. 2017) may yet provide breakthroughs in describing how cells in a community interact with each other (Ji and Nielsen 2015). These approaches and tools must be translated to a diversity of organisms and biological systems and iteratively tested in order to refine them to the point where they are generalizable (Henry et al. 2010). However, with few exceptions, many tools have not yet closed the iterative feedback loop that exemplifies a build-test cycle necessary to accelerate tool development (Carbonell et al. 2016).

Iterative feedback loops

To understand a biological system fully, one must incorporate both experiments and theory development. This iterative feedback loop method is easily seen when trying to elucidate an atomic-level understanding of the structural, thermodynamic, and kinetic principles

that control function of dynamic protein complexes, which is too daunting of a task to accomplish by experiments alone. Here, we present a case study that demonstrates the need for development of both experimental and modeling approaches for understanding complex biological systems.

Dynamic complexes composed of intrinsically disordered proteins (IDPs) and multiple folded proteins play a fundamental role in many biological processes, from organizing the cell-cycle, tethering cargo to molecular motors (Fejtova et al. 2009; Gupta et al. 2012; Hammer and Wagner 2013), controlling gene regulation (Eastwood et al. 2021), to coordinating the formation of meshlike assemblies associated with phase-separation events (Moutin et al. 2014; Myllykoski et al. 2018), thus having an impact on a range of diseases from cancer (Becker et al. 2018; He et al. 2018: 11) to neurodegenerative pathologies (Chen, Gerwin and Sheng 2009) to viral infection (Kirkham et al. 2015). However, despite their prevalence and importance, our mechanistic understanding of these multivalent assemblies has been severely stymied by their large size and extreme levels of structural and compositional heterogeneity.

Compositional heterogeneity is a hallmark of biological signaling, as many proteins can bind to a different number of interacting partners, all present in equilibrium (Fig. 2). Importantly, there are no experiments that can currently determine the population of each of these species (thermodynamics), their rates of exchange between multiple species (kinetics), and how they look (structure). Thus, mathematical modeling of the structures of this biological mixture is critical to link the functional biological outcomes observed in experiments.

To address these key questions in functional structural biology, a host of innovative methods integrating multiscale computations with a range of experimental modalities are necessary. We highlight here one example from coauthor E. Barbar's lab that demonstrates the importance of integration of experiment and computation. This case study also exemplifies the importance of predictive models for mechanistic understanding of highly complex biological systems that cannot be addressed through experiments alone.

In this example, a heterogeneous assembly of proteins (DNA binding transcription factor called ASCIZ) with an interacting protein (LC8) regulates transcription of ASCIZ and is critical for sensing LC8 concentration in cells, termed the *sensor hypothesis*. In this model, high LC8 occupancy shuts down ASCIZ transcription, and low LC8 occupancy turns on transcription (Fig. 2A), but a range of dynamic, low-occupancy complexes (Fig. 2B) are the dominant species that function as a rheostat to tune the biology, rather than an on/off switch (Jurado et al. 2012; Zaytseva et al. 2014; Clark et al. 2018). Thus, understanding via predictive modeling how these reversibly forming "girders" are generated by a range of IDPs, through dynamic cross-linking with the hub protein LC8, could transform aspects of architectural/ mechanical cell biology (Fig. 2).

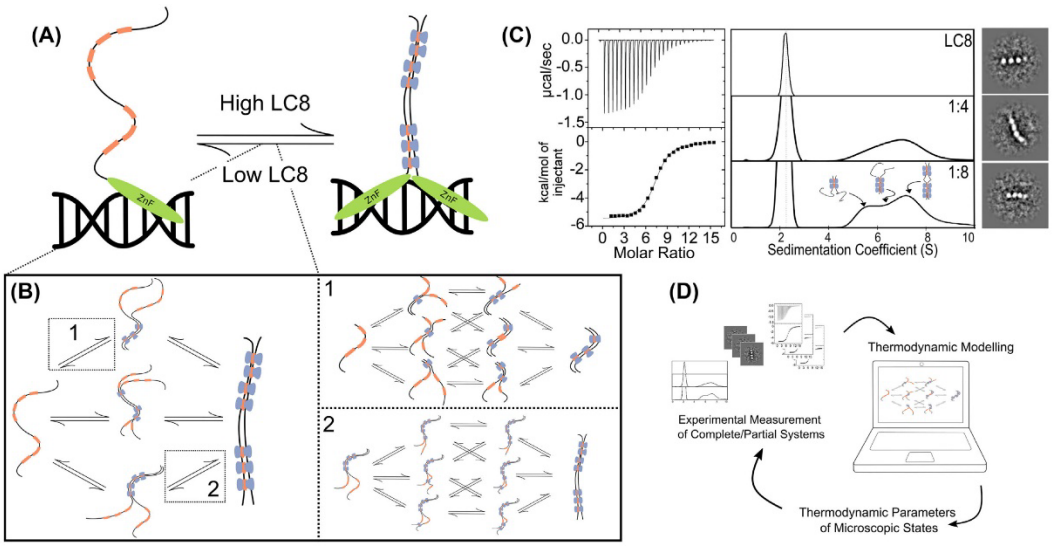

Figure 2. Combining experimental and modeling approaches to investigate molecular-scale binding systems. **(A)** Molecular-scale system: A hub protein dimer (LC8, blue) coordinates with a multidomain protein (ASCIZ) with a DNA binding domain (ZnF, green) and a long disordered tail with multivalent LC8 binding sites (orange). **(B)** Cut-in overview of thermodynamic model of LC8-ASCIZ binding, showing three possible intermediate species. Boxes 1 and 2 show further detail of the equilibria of the overview diagram illustrating the complexity. This is a relatively simple example in terms of multimolecular binding complexities observed in biological molecular interactions **(C)** Multiple experimental techniques to study protein complexes capture only an overall picture of the thermodynamics of binding and not the microstates, heterogeneous species, and achieve only very low-throughput analysis of intermediate states. **(D)** Integration of modeling with experimental data. Model depends on experimental measurements of macroscopic states to compute thermodynamic parameters for microscopic states. Uncertainty in model-derived parameters dictate what further experiments on partial systems are necessary. Those experiments, in turn, can be fed back to the model to improve accuracy and precision of fit parameters.

As with many biological processes, this example illustrates how complex heterogeneity of this system drives diverse functional roles, including modular sensing, responsive feedback regulation, and intermediate/equilibrium binding states. Characterization of biological interactions actually requires an *ensemble perspective* at *multiple scales*—due to heterogeneity across four distinct biological concepts:

(1) the number and orientation of interactors. In this example, one ASCIZ protein can interact with many LC8 molecules at once.

(2) potential correlations between clusters of connected interactors. In this example, the ASCIZ/LC8 complex interacting with additional LC8 monomers.

(3) equilibrium interactions between given interactors. In this example, flux as individual proteins or interactors equilibrate across distinct states in (i) and (ii).

(4) cooperative behavior. In this example, cooperative allostery where binding of LC8 at one site enhances rate of LC8 binding at other sites (Fig. 2B).

Biologists working across scales will read this list of factors/examples and be able to find connections to key biological and interaction concepts that match at their own scale. For example, replace "interactors" above with "species" and you'll find a list of biological concepts that would be easily understood among ecologists (e.g., interspecies interactions, predator/prey segregation, population gain/loss affecting other species, and cooperative feedback from environment (food/water abundance or scarcity, etc.)).

One key benefit of modeling at the molecular/atomic scale is that quantitative measurements of these transient equilibria populations are sometimes not possible using typical binding experiments which yield values for average stoichiometries (Fig. 2C), with no information on individual species nor on the interaction between species distributions. Thus, what is needed is incorporation of a novel, *fully end-to-end automated approach* to multi-species modeling that can robustly account for combinatorial binding equilibria. The ideal analysis, using multiple experiments, could extract experimentally invisible *species-specific* binding parameters, leading to population profiles of dynamic complexes. Also needed is the development of more robust experimental methods that can better quantify or trap transient species to make the unmeasurable measurable.

The biological impact of this case study is the elucidation of the mechanism that underlies the regulatory/sensing abilities of large dynamic complexes and their (dis)assembly pathways. Moreover with multiple iterations of experiment and modeling, we hope to reach a stage where we can predict from the protein sequence: the number of binding partners, their individual binding affinities, the length of the disordered linkers separating them, whether a certain IDP will form a dynamic IDP assembly, phase separate, aggregate, or connect multiple assemblies. These impacts are described at the molecular scale, but similar integration of experiment and modeling feedback at the cellular or organismal scale could transform our understanding of equilibrium dynamics regulating cooperativity and competition during cancer growth or in a complex ecosystem.

Overcoming cultural barriers

Traditional practices within science research promote separatism among disciplines. These separate or independent research processes often lead to inefficiencies in the scientific research community (Gray 2008). Examples of this practice include, but are not limited to, mathematical science researchers tending not to collaborate with biological science researchers and vice versa. Increased collaborations among these disciplines could lead to solving bigger, more complex problems.

We believe the most effective routes to overcoming these cultural barriers include programs that train new generations of scientists and researchers to work at the transdisciplinary interface between mathematics and biology. A well-developed platform for networking and exchange between these fields would provide scientists and researchers with direct access to a wealth of knowledge that was frequently underutilized by their predecessors.

To this end, we propose a four-fold systematic approach that incorporates sustained funding, research training, curriculum, and outreach (Fig. 1).

Sustained funding

Dedicated federal and private funds should be identified and earmarked to expand support for transdisciplinary research and the creation of education centers. In particular, long-term extramural funding opportunities should be provided to establish and maintain infrastructure and research expenses for centers housed within colleges and universities. Federally, first steps are being taken with grant mechanisms such as the NSF Research Coordination Networks that provide funds to foster the creation of collaborative networks. However, these grants lack funding for the research itself which often fails to incentivize forming transdisciplinary research networks. In the private sector, organizations such as the American Cancer Society, Burroughs Wellcome Fund, and the Howard Hughes Medical Institute all agree that philanthropic giving has to play an important role with investigators, building bridges between traditional and emerging fields of research (see Training the Next Generation, https://www.bwfund.org).

Training

Predoctoral and postdoctoral fellowship opportunities should be established to support trainees wanting to work at the mathematics/biology interface. The National Science Foundation Simons MathBioSys Research Center serves as a good example of a model training program. This program matches experimentalists and mathematician mentors with interdisciplinary trainees with an emphasis on building interactional expertise. Another such endeavor is the Institute for Systems Biology (ISB, https://isbscience.org) created in 2000 as the first ever institute created for systems biology. ISB serves as an environment in which scientific collaboration stretches across disciplines and leverages biological approaches to understand mechanisms.

Curriculum

At both the undergraduate and graduate levels, math and biology degree programs should serve to integrate these fields early in a student's post-secondary training. This new transdisciplinary perspective will help expand teaching through implementing new approaches to pedagogy. Faculty development efforts should be implemented to help instructional staff develop their own skills as well. Students at all levels would be encouraged to pursue independent research and engage in laboratory courses, seminar series, and advanced elective experiences. To support transdisciplinary research, we must create a database of training modules that include the datasets, code, and model tutorials. One fantastic example of an accessible and user-friendly model is PhysiCell (Ghaffarizadeh et al. 2018). PhysiCell is a computationally powerful modeling approach for cell-level competition and movement modeling, with both easy-to-follow tutorials as well as crowd-sourced curriculum and educational modules using the platform. Using the success of PhysiCell as a template, a database of training modules that include datasets, code, and model tutorials should be developed.

Outreach

Establishing outreach programs at the middle and high school levels will build up strong cohorts for pursuing transdisciplinary undergraduate degrees. The training module database just described could be easily simplified and adapted for use in middle and high schools to show students how math has improved our understanding of biological questions relevant to society (i.e., climate change, human health, etc.).

Transdisciplinary research can be uncomfortable, difficult, and humbling. Critical to overcoming cultural barriers is emphasizing the creation of a warm and welcoming environment of like-minded researchers who are motivated to learn collectively from diverse perspectives. Special attention needs to be placed on setting a stage that lowers barriers to building a community of "learning teachers" rather than experts. At the same time, creating such an atmosphere would undoubtedly further encourage people from all backgrounds to continue at the interface between biology and math and enhance the diverse and creative potential of the field.

Scientific outcomes

All biological subdisciplines could benefit tremendously from better integrating theoretical modeling approaches as proposed here. The theoretical modeling approaches developed in physics and chemistry disciplines are powerful tools used to elucidate mechanism and can be highly predictive under defined or constrained biological conditions. However, even mathematicians recognize that these constrained conditions rarely happen in biological systems. Conversely, the complexity of biological systems necessitates new ideas on how to express higher-order model behavior and how to scale models to higher levels of complexity. Further development of models that are applicable and validated across biological scales is required in order to fully harness the power of mathematical modeling. Unifying biology with physics, chemistry, and mathematics/ statistics through the use of common model methodologies that apply across scales has the potential to revolutionize our fundamental understanding of biology and biological systems. Finally, building a foundation of integrated feedback loops between model and experiment will serve the goals of both cross-scales applicability and dealing with increased biological complexity. Feedback loops include model refinement and iteration to better reflect the biology, which is routinely accomplished for standard statistical modeling but is essential for integrated predictive modeling. An additional underutilized feedback loop of using the model to guide experimental design is critical for efficient model validation, easier application of models across scales, and for maximizing return on investment for experimental resources.

Accurate models have the power to transform society by serving as a foundation for technological innovation. If biological models begin to approach the predictive accuracy of physical models, we could predict and *design* biology with the ease that we can design a computer. Systematic and iterative refinement of models that describe biological systems forms the bedrock conceptual framework needed to understand molecular, organism, and population behaviors. The development of predictive biological models is also necessary to generate hypotheses, to capture the drivers of biological heterogeneity, and to inspire future discovery. Integrated cross-scale models will inevitably be needed to solve pressing

social challenges such as counteracting the effects of climate or habitat change, reducing the time to harvest food crops to feed a growing world population, curing disease, identifying and preventing the spread of emerging infections threats, and designing biological technologies to generate clean renewable energy. These outcomes cannot be achieved unless we foster cross-disciplinary collaboration, provide long-term funding opportunities at this interface, and train the next generation of scientists to explore a new science frontier at the interface between biology and mathematical modeling.

Author contributions – All authors contributed to drafting and revising this manuscript. All authors approved final submission of the manuscript.

Acknowledgments – This paper was initiated at, and supported by, the US National Science Foundation (NSF) Reintegrating Biology Jumpstart Workshop held in Austin, Texas, in December 2019. We thank Kingshuk Ghosh (University of Denver) for helpful feedback and discussion during early drafts of this perspective piece. Figure 1 was generated using Biorender Software.

Funding – The work was supported by the National Science Foundation [IOS-1938948 to N.R.B.]; the USDA HatchMultistate NC1200 Photosynthetic Processes [NEB-30–133 to N.R.B.]; the Nebraska Center for Energy Sciences Research (Cycle 15) grants [to N.R.B.]; and the National Science Foundation [Award #1617019 to E.B.].

Data availability – No new data were generated or analyzed in support of this research.

References

Afraimovich V, Tristan I, Huerta R, Rabinovich MI. 2008. Winnerless competition principle and prediction of the transient dynamics in a Lotka-Volterra model. Chaos Interdisc J Nonlinear Sci 18:043103.

Ainsworth C. 2020. Therapeutic microbes to tackle disease. Nature 577:S20–2.

Almaraz E, Gómez-Corral A, Rodríguez-Bernal MT. 2016. On the time to reach a critical number of infections in epidemic models with infective and susceptible immigrants. Biosystems 144:68–77.

Aristide L, Morlon H. 2019. Understanding the effect of competition during evolutionary radiations: an integrated model of phenotypic and species diversification. Mooers A (ed.). Ecol Lett 22:2006–17.

Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, et al. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol 36:566–9.

Baigent S. 2010. Lotka-Volterra Dynamics—An introduction. London: Taught Course Center, University College London, p.1–60. (http://www.ltcc.ac.uk/media/london-taught-coursecentre/documents/Bio-Mathematics-(APPLIED).pdf)

Becker JR, Cuella-Martin R, BarazasM, Liu R, Oliveira C, Oliver AW, Bilham K, Holt AB, Blackford AN, Heierhorst J, et al. 2018. The ASCIZ-DYNLL1 axis promotes 53BP1-dependent non-homologous end joining and PARP inhibitor sensitivity. Nat Commun 9:5406.

Bergthorsson U, Andersson DI, Roth JR. 2007. Ohno's dilemma: evolution of new genes under continuous selection. Proc Natl Acad Sci 104:17004–9.

Bordbar A, Monk JM, King ZA, Palsson BO. 2014. Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet 15:107–20.

Breed RS, Murray EG, Hitchens AP. 1944. The outline classification used in the bergey manual of determinative bacteriology. Bacteriol Rev 8:255–60.

Carbonell P, Currin A, Jervis AJ, Rattray NJ, Swainston N, Yan C, Takano E, Breitling R. 2016. Bioinformatics for the synthetic biology of natural products: integrating across the Design–Build–Test cycle. Nat Prod Rep 33:925–32.

Cardinale S, Joachimiak MP, Arkin AP. 2013. Effects of genetic variation on the *E. coli* host-circuit interface. Cell Rep 4:231–7.

Cashman M, Catlett JL, Cohen MB, Buan NR, Sakkaff Z, Pierobon M, Kelley CA. 2017. BioSIMP: using software testing techniques for sampling and inference in biological organisms. Proceedings of the 2017 IEEE/ACM 12th International Workshop on Software Engineering for Science (SE4Science). Buenos Aires, Argentina: IEEE, 2–8.

Chen Y-M, Gerwin C, Sheng Z-H. 2009. Dynein light chain LC8 regulates syntaphilin-mediated mitochondrial docking in axons. J Neurosci 29:9429–38.

Claesen J, Fischbach MA. 2015. Synthetic microbes as drug delivery systems. ACS Synth Biol 4:358–64.

Clark S, Myers JB, King A, Fiala R, Novacek J, Pearce G, Heierhorst J, Reichow SL, Barbar EJ. 2018. Multivalency regulates activity in an intrinsically disordered transcription factor. Elife 7:e40684.

Colin J, Libri D, Porrua O. 2011. Cryptic transcription and early termination in the control of gene expression. Genet Res Int 2011:653494.

Crick F. 1970. Central dogma of molecular biology. Nature 227:561–3.

de la Cruz F, Davies J. 2000. Horizontal gene transfer and the origin of species: lessons from bacteria. Trends Microbiol 8:128–33.

DeAngelis DL, Grimm V. 2014. Individual-based models in ecology after four decades. F1000Prime Rep 6:39.

dos Reis M. 2003. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K–12 genome. Nucleic Acids Res 31:6976–85.

Duarte NC. 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res 14:1298–309.

Eastwood EL, Jara KA, Bornelöv S, Munafò M, Frantzis V, Kneuss E, Barbar EJ, Czech B, Hannon GJ. 2021. Dimerisation of the PICTS complex via LC8/Cut-up drives cotranscriptional transposon silencing in Drosophila. Elife 10:e65557.

Edwards AWF. 2008. G. H. Hardy (1908) and Hardy-Weinberg equilibrium. Genetics 179:1143–50.

Ellison G. 2020. Implications of heterogeneous SIR models for analyses of COVID-19, NBER Work Paper No. 27373, Cambridge (MA): National Bureau of Economic Research.

Fejtova A, Davydova D, Bischof F, Lazarevic V, Altrock WD, Romorini S, Schöne C, Zuschratter W, Kreutz MR, Garner CC, et al. 2009. Dynein light chain regulates axonal trafficking and synaptic levels of Bassoon. J Cell Biol 185:341–55.

Freed E, Fenster J, Smolinski SL, Walker J, Henard CA, Gill R, Eckert CA. 2018. Building a genome engineering toolbox in nonmodel prokaryotic microbes. Biotechnol Bioeng 115:2120–38.

Ghaffarizadeh A, Heiland R, Friedman SH, Mumenthaler SM, Macklin P. 2018. PhysiCell: an open source physics-based cell simulator for 3-D multicellular systems. Poisot T (ed.). PLoS Comput Biol 14:e1005991.

Goyal M, Bhardwaj VK, Prakash A. 2021. Investigating new positive, bounded, and convergent numerical solution for the nonlinear time-dependent breast cancer dynamic competition model. Math Meth Appl Sci 44:4636–53.

Gray B. 2008. Enhancing transdisciplinary research through collaborative leadership. Am J Prev Med 35:S124–32.

Grimm V, Ayllón D, Railsback SF. 2017. Next-generation individual-based models integrate biodiversity and ecosystems: yes we can, and yes we must. Ecosystems 20:229–36.

Grimm V. 1999. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? Ecol Modell 115:129–48.

Großkopf T, Consuegra J, Gaffé J, Willison JC, Lenski RE, Soyer OS, Schneider D. 2016. Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. BMC Evol Biol 16:163.

Gupta A, Diener DR, Sivadas P, Rosenbaum JL, Yang P. 2012. The versatile molecular complex component LC8 promotes several distinct steps of flagellar assembly. J Cell Biol 198: 115–26.

Gygi SP, Rochon Y, Franza BR, Aebersold R. 1999. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19:1720–30.

Hammer JA, Wagner W. 2013. Functions of class V myosins in neurons. J Biol Chem 288:28428–34.

Hardy GH. 1908.Mendelian proportions in a mixed population. Science 28:49–50.

He YJ, Meghani K, Caron M-C, Yang C, Ronato DA, Bian J, Sharma A, Moore J, Niraj J, Detappe A, et al. 2018. DYNLL1 binds to MRE11 to limit DNA end resection in BRCA1-deficient cells. Nature 563:522–6.

Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28:977–82.

Hosono Y. 1998. The minimal speed of traveling fronts for a diffusive Lotka-Volterra competition model. Bull Math Biol 60:435–48.

Hsu S-B, Zhao X-Q. 2012. A Lotka-Volterra competition model with seasonal succession. J Math Biol 64:109–30.

Hsu S-B. 1982. On a resource based ecological competition model with interference. J Math Biol 12: 45–52.

Isaac Newton Institute Fellows, Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, SloanWT, Cordero OX, Brown SP, Momeni B, Shou W, et al. 2016. Challenges in microbial ecology: building predictive understanding of community function and dynamics. ISME J 10:2557–68.

Ji B, Nielsen J. 2015. From next-generation sequencing to systematic modeling of the gut microbiome. Front Genet 6: 219.

Jurado S, Conlan LA, Baker EK, Ng JL, Tenis N, Hoch NC, Gleeson K, Smeets M, Izon D, Heierhorst J. 2012. ATM substrate Chk2-interacting Zn2+ finger (ASCIZ) Is a bi-functional transcriptional activator and feedback sensor in the regulation of dynein light chain (DYNLL1) expression. J Biol Chem 287:3156–64.

Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, et al. 2013. EcoCyc: fusing model organism databases with systems biology. NucleicAcids Res 41:D605–12.

Kirkham JK, Park SH, Nguyen TN, Lee JH, Günzl A. 2016. The dynein light chain LC8 is required for RNA polymerase I mediated transcription in *Trypanosoma brucei*, facilitating assembly and promoter binding of class I transcription factor A. Mol Cell Biol 36:95–107 MCB.00705-15.

Kurtz CB, Millet YA, Puurunen MK, Perreault M, Charbonneau MR, Isabella VM, Kotula JW, Antipov E, Dagon Y, Denney WS, et al. 2019. An engineered *E. coli* Nissle improves hyperammonemia and survival in mice and shows dose-dependent exposure in healthy humans. Sci Transl Med 11: eaau7975.

Le Novère N. 2015. Quantitative and logic modelling of molecular and gene networks. Nat Rev Genet 16: 146–58.

Lewin-Epstein O, Hadany L. 2020. Host-microbiome coevolution can promote cooperation in a rock-paper-scissors dynamics. Proc R Soc B Biol Sci 287:20192754.

Lotka A. 1925. Elements of physical biology. Nature 116:461.

Magal P, Zhang Z. 2017. Competition for light in forest population dynamics: From computer simulator to mathematical model. J Theor Biol 419:290–304.

McEntire KD, Maerz JC. 2019. Integrating ecophysiological and agent-based models to simulate how behavior moderates salamander sensitivity to climate. Front Ecol Evol 7:22.

Mizukami M, Winkler M. 2017. Boundedness and asymptotic stability in a two-species chemotaxis-competition model with signal-dependent sensitivity. Discr Contin Dyn Syst B 22:2301–19.

Moutin E, Compan V, Raynaud F, Clerté C, Bouquier N, Labesse G, Ferguson ML, Fagni L, Royer CA, Perroy J. 2014. The stoichiometry of scaffold complexes in living neurons—DLC2 functions as a dimerization engine for GKAP. J Cell Sci 127:3451–62.

Myllykoski M, Eichel MA, Jung RB, Kelm S, Werner HB, Kursula P. 2018. High-affinity hetero-tetramer formation between the large myelin-associated glycoprotein and the dynein light chain DYNLL1. J Neurochem 147:764–83.

Nathan Ngoteya F. 2015. Sensitivity analysis of parameters in a competition model. Appl Comput Math 4:363.

Ojuederie OB, Babalola OO. 2017. Microbial and plant-assisted bioremediation of heavy metal polluted environments: a review. Int J Environ Res Pub Health 14:1504.

Oren Y, Smith MB, Johns NI, Kaplan-Zeevi M, Biran D, Ron EZ, Corander J, Wang HH, Alm EJ, Pupko T. 2014. Transfer of noncoding DNA drives regulatory rewiring in bacteria. Proc Natl Acad Sci 111:16112–7.

Perez-Garcia O, Lear G, Singhal N. 2016. Metabolic network modeling of microbial interactions in natural and engineered environmental systems. Front Microbiol 7:673.

Peterman WE, Gade M. 2017. The importance of assessing parameter sensitivity when using biophysical models: a case study using plethodontid salamanders. Popul Ecol 59: 275–86.

Rangel TF, Edwards NR, Holden PB, Diniz-Filho JAF, Gosling WD, Coelho MTP, Cassemiro FAS, Rahbek C, Colwell RK. 2018. Modeling the ecology and evolution of biodiversity: biogeographical cradles, museums, and graves. Science 361:eaar5452.

Riddell EA, Apanovitch EK, OdomJP, Sears MW. 2017. Physical calculations of resistance to water loss improve predictions of species range models. Ecol Monogr 87:21–33.

Sakkaff Z, Catlett JL, Cashman M, Pierobon M, Buan NR, Cohen MB, Kelley CA. 2017. End-to-end molecular communication channels in cell metabolism: an information theoretic study. Proceedings of the 4th ACM International Conference on Nanoscale Computing and Communication. Washington DC: ACM, 1–6.

Sears MW, Angilletta MJ. 2015. Costs and benefits of thermoregulation revisited: both the heterogeneity and spatial structure of temperature drive energetic costs. Am Nat 185:E94–E102.

Sullivan DJ, McEntire KD, Cohen BS, Collier BA, Chamberlain MJ. 2020. Spatial scale and shape of prescribed fires influence use by wild turkeys. JWildlife Manag 84:1570–7.

Sun Z, Lorscheid I, Millington JD, Lauf S, MaglioccaNR, Groeneveld J, Balbi S, Nolzen H, Müller B, Schulze J, et al. 2016. Simple or complicated agent-based models? A complicated issue. Environ Model Softw 86:56–67.

Suthers PF, Foster CJ, Sarkar D, Wanga L, Maranas CD. 2021. Recent advances in constraint and machine learning-based metabolic modeling by leveraging stoichiometric balances, thermodynamic feasibility and kinetic law formalisms. Metab Eng 63:13–33.

Volterra V. 1926. Fluctuations in the abundance of a species considered mathematically. Nature 118:558–60.

Westerhoff HV, Palsson BO. 2004. The evolution of molecular biology into systems biology. Nat Biotechnol 22:1249–52.

Wilensky U, Resnick M. 1999. Thinking in levels: a dynamic systems approach to making sense of the world. J Sci Educ Technol 8:3–19.

Willem L, Verelst F, Bilcke J, Hens N, Beutels P. 2017. Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006–2015). BMC Infect Dis 17:612.

Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, et al. 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature 462:1056–60.

Zaytseva O, Tenis N, Mitchell N, Kanno S, Yasui A, Heierhorst J, Quinn LM. 2014. The novel zinc finger protein dASCIZ regulates mitosis in *Drosophila* via an essential role in dynein lightchain expression. Genetics 196:443–53.

Zhang W, Lam JSL. 2013. Maritime cluster evolution based on symbiosis theory and Lotka-Volterra model. Maritime Pol Manag 40:161–76.

Zhu C, Yin G. 2009. On competitive Lotka-Volterra model in random environments. J Math Anal Appl 357: 154–70.