

Microstates to Macrodynamics: A New Mathematics of Biology

1 The Challenge

Biology is often described as an empirical science with limitless detail, but every instantiation of experiment, observation, and classification is closer akin to an island embedded within an archipelago of unifying theory and commonalities. Indeed, the most lasting innovations in biology are those that bridge the gap between (seemingly) disparate phenomena, crossing the threshold into new paradigms of insight and synthesis. These bridges are theory, their structure is mathematics – whether Darwin’s theory of natural selection, Watson and Crick’s structure of DNA, or Hodgkin-Huxley’s model of the action potential; the present vista of possibilities for advances in mathematical biology extends outward (e.g., protein structure to metabolic pathways to quorum sensing to population dynamics) even as the resolution of observation grows more detailed (e.g., advanced imaging, tagged proteins, and environmental shotgun sequencing).

It is not hyperbole to state that biology is undergoing a renaissance the way physics went through a similar renaissance in the 20th century, though its practitioners are not exclusively biologists; they are mathematicians, physicists, computer scientists, and biochemists, and most importantly, interdisciplinary teams of all of these. The challenge that biology faces is to go beyond identifying life’s basic elements and to explain the emergence of biological organization at multiple scales. Solving such challenges will require new mathematics or broadening of present mathematics that can describe and reconcile the statistics and properties of microstates with the dynamics and behavior of macrostates. In so doing, DNA, proteins, cells, organisms, and environments are objects of primary study, but more importantly, the sites of exchange of information and material across landscapes, surfaces, and networks at vastly different spatiotemporal scales.

To address these challenges, we have assembled a team well-balanced between mathematical theory and empiricism, with a strong representation of those who can serve to interface between the extremes. We will bring new mathematical perspectives to the grand biological challenges, use the stimulus of those challenges to spur on the creation of new mathematics, and aim to discover fundamental laws of biology. The grand challenges in biology we will tackle and the mathematical methodologies we will use to solve them are linked in natural ways (e.g. algebraic statistics and ocean diversity, probability theory and fitness landscapes, differential topology and development, dynamical systems and the evolutionary ecology of viruses) but not limited to arenas of prior success. These problems and methods form a notional pentagon, with biological problems at each vertex (listed in The Team) and mathematical themes (symmetry, duality, and stochasticity) serving as means to unify them. Our aim, in all cases, is to develop mathematical hypotheses (Plato), solve them in appropriate limits (Dirac), and test them in experimental systems. As the program progresses, we expect to fill in this pentagon, establishing links between all its vertices, so that in the end we will be able to interpolate smoothly and efficiently from (e.g.) the level of DNA sequences to gene expression and on to biodiversity dynamics or evolutionary history.

What chance do we have of reaching this goal? Twenty five years ago, Horace Judson, in his famous book on the history of modern molecular biology, wrote that (Judson, 1979)

[The] discoveries have not produced the great practical power that has repeatedly been anticipated for them. Biologists have no atomic power stations and no bombs to point to, or at least not yet. No baby has been cured of a congenital deficiency by insertion of a missing gene into its cells. There is no vaccine against human leukemia, not even a cure for hay fever. Though some of the rewards are at last imminent, most scientists have learned that they must speak guardedly and emphasize to laymen the gaps to be filled in.

Scientists are much more optimistic these days that biological discoveries may well reshape the treatment of disease, understanding of the dynamical mechanisms for the maintenance and generation of diversity, explanations of the development of robustness in complex biological networks, and so on. Yet, a means to predict, design, and modify the behavior of biological systems (from synthetic biology to biomimicry to ecosystem management) is still in its infancy. The reasons are simple: if biology is to move beyond a descriptive science to becoming a predictive science, then a mathematics of non-linear systems, comprised of evolving heterogeneous agents often in conflict and competition with one another, must be discovered and mapped out. Advances in biological data collection as well as innovations in high-throughput computing, constitute a convergent opportunity for such a paradigm shift. The symbiosis between mathematics and physics in the 20th century changed both fields, and quite literally, the world. Biology awaits a similar revolution; this proposal (and the assembled team) is designed to help launch it.

2 The Problems

Microstates to Macrodynamics: One of the fundamental challenges in biology is to explain the link between microscopic building blocks of individual organisms and macroscopic organismal/ecological phenomena within an evolutionary context (Cohen, 2004). Moving from the micro- to the macro-scale requires a description of the non-equilibrium statistical mechanics of large assemblages of individual agents, in which the range of types is continuously evolving through the infusion of new diversity via mutation and other mechanisms. In a sense, stochasticity is embedded in the foundations of multiscale biology and may require concomitant work on novel formulations of appropriate mathematics (Mumford, 1999). An illustrative biological example is the dominant primary producer in the oceans, the cyanobacteria *Prochlorococcus*, for which genetic data along transects are now available, physiological function is partially mapped, and flow cytometry affords data on population levels with depth (Partensky et al., 1999). Still, virtually no synthesis of how the genetic data links to higher structure and organization is forthcoming. Extending the methods of algebraic statistics (Pachter and Sturmfels, 2005) into phylogenodynamics would be a promising start to merge discrete mathematics with theories of dynamical systems to understand the extent of ocean diversity (Venter et al., 2004) and whether this diversity is functionally important. Applying and extending such mathematical developments, or alternatives such as equation-free modeling, where continuous dynamic derivatives are replaced by approximations to the tangent bundle of samples from stochastic microstates (Cisterna et al., 2004), provides the best hope for progress in understanding multiscale biological systems.

Evolution of Robustness: Organisms do not require precisely tuned environmental conditions, they tolerate noisy inputs, sudden change, and competition – robustness is a universal property of life. But, how different are the properties of systems in which robustness is selected for through natural selection from those where robustness is an emergent property? What are the identifying characters of robustness and can we use dynamical measures (e.g., non-stationary time series analysis, topology of data clouds, information theory) to assess the response of organisms to changing conditions, whether endogenous or exogenous? The robustness of the human cardiovascular system, for example, appears to have dynamical signatures that could be used as a diagnostic tool for the treatment of disease (Goldberger et al., 2002). Likewise, bacterial populations can modify their collective behavior in response to starvation/crowded conditions, encoding population-level robustness within autonomous, distributed units (Bassler, 2002; Park et al., 2003). At a (perhaps) deeper level, organisms may modify the topology of their fitness landscapes to tune themselves to robust states, a conjecture related to theories of Highly Optimized Tolerance (Carlson and Doyle, 2000), selection for evolvability in proteins (Earl and Deem, 2004), and studies of the evolution of digital organisms (Lenski et al., 1999). A major innovation in the mathematical foundations of the evolution of robustness would be a watershed moment for biology.

Structure, Function, and Dynamics of Biological Networks: Networks are ubiquitous features of biological systems, whether phylogenetic trees, protein contact maps, or networks of sexual contacts. In recent years, the study of network theory has discovered power-law distributions in non-spatial networks (such as the world wide web) and means to relate dynamical rules of node addition to resultant network topologies (Albert and Barabasi, 2002) – but slower progress has been made in integrating network theory into ontogenetic, spatial, and dynamic phenomena (Newman, 2002). Growth and form offers intriguing opportunities for study, for example, development in humans relies on embedding differentiation into a branching network, but no integrated theory exists of vasculogenesis given topological constraints on growth. In addition, the segmentation of organisms (from insects to vertebrates) relies on linking genetic regulatory networks with large scale patterning (Pourquie, 2003). As such, what bearing do the classification of moves of (or on) manifolds have on empirical studies of development? What topological aspects of hierarchical resource supply networks are consistent/conserved from bacteria to whales? Distinct from these questions, but equally important, is deducing what role neutral or nearly neutral networks – highly connected paths connecting equally fit states (Gavrilets, 2004) – serve in fitness landscapes? Devising a classification of network topologies and homogeneous scaling relationships to transform them is a DARPA-hard problem that crosses scales from cells (metabolic networks) to organisms (intra- and inter-organ connections) to populations (food webs).

Mathematical Issues and Foundations: From a mathematical perspective, many of the biological problems we will consider look like formally similar problems considered at varying scales. Ultimately we expect to develop a theory of phase and configuration spaces for biological systems, whose dynamics could comfortably represent phenomena as superficially disparate as gene expression, embryological development, or the coupled response of ecosystem function to environmental change. To accomplish this we need better techniques for understanding dynamical systems composed of structurally linked components; ideas from

percolation theory (Gimmet, 1999) and singularity theory (Arnol'd, 2000) will be relevant here. Both provide sophisticated approaches to questions of scale change and phase transitions, but they approach the subject from very different directions, and a theory encompassing both points of view may be necessary for biological applications. We hope such methods will clarify the nature of useful order parameters for biological systems. Potential axes might be adaptability vs. stability, invariance vs. diversity, or maintenance vs. development. Ideas from game theory (Hofbauer and Sigmund, 1998) should also play an important role in such quasi-thermodynamical thinking about biology. The effort to develop a new mathematical foundation for the study of biological problems will, as the program progresses, become closer aligned with computational tests and implementations. For example, the modern theory of integrable systems in terms of toric varieties (Kapovich et al., 2000) may provide computationally effective links between the algebraic statistics of DNA sequence space and the level of subcellular biochemical processes (e.g. cell membrane behavior, cytoplasm structure, transcription). The sophistication of statistical approaches in modern biology suggests that advances in mathematics may well operate on models even before they are applied to data, in keeping with our vision of linking hypotheses to asymptotics to real-world complexities.

3 The Team

DNA

Simon Levin, Princeton University. Mathematical biology; complex adaptive systems.
Joshua Weitz, Princeton University. Host-pathogen coevolution; scaling theory.
Lior Pachter, UC-Berkeley. Computational biology; metagenomics.
Bernd Sturmfels, UC-Berkeley. Algebraic statistics; phylogenetics; discrete mathematics.

Fitness Landscapes

Richard Lenski, Michigan State University. Microbial ecology & evolution.
Peter Bates, Michigan State University. Dynamical systems; nonlinear PDEs.
Michael Deem, Rice University. Mathematical immunology; vaccine prediction.
Charles Epstein, University of Pennsylvania. Spectral theory, MR image analysis.

Cell Signaling

Ned Wingreen, Princeton University. Quorum sensing; chemotaxis; cell biophysics.
Herbert Edelsbrunner, Duke University. Computational topology; structural biology.

Development/Structural Stability

Olivier Pourqu e, Stowers Institute. Molecular and genetic basis of development.
Jack Morava, Johns Hopkins University. Topology; quantum field theory.
Sorin Popescu, SUNY Stonybrook. Computational algebra, algebraic geometry.
James Damon, University of North Carolina. Singularity theory; medical imaging.

Physiology

Tim Buchman, Washington University School of Medicine St. Louis. Trauma surgery; physiological variability and health.
Robert Bryant, Duke University. Differential geometry and applications.

4 Statement of Work

Phase 1: We propose to undertake a yearlong effort to identify fundamental problems in mathematical biology that cross many orders of magnitude in spatiotemporal scales. As such, we will undertake work at each vertex, among multiple vertices, as well as discussion and meetings that span all five teams. Activities include single node workshops as well as an all-PI kickoff meeting to begin immediately after initiation of the project, short (day–week) as well as long-term (week–month) visits between different institutions, writing non-technical introductions to biological phenomena as well as summaries of theoretical approaches to be exchanged between mathematicians and biologists, as well as the development of an online repository of relevant scientific papers and preprints to be used as a resource by all team members. Finally, because of the interdisciplinary scope of the proposal and the nature of theoretical work, we will also allow for appropriate quiet time between meetings to familiarize ourselves with new theoretical and empirical terrain.

Phase 2: We will continue to have meetings and workshops with the goal of transferring mathematical language developed in Phase 1 into testable theories of biological phenomenon – this will include an all-PI meeting in the middle of Phase 2. We will assemble datasets to share among the teams – whether genomic data, data on signaling networks, physiological time series, or even population densities for ecosystem level studies. We will implement computational models of theories and propose Gedanken experiments as a means to motivate the identification of model organisms and ecosystems for subsequent testing. We will also begin exploring possible experimental collaborations both within and outside the list of team members identified in the present proposal.

Phase 3: An all-PI meeting with a focus on experimental efforts will kick-off the transition from the development of theory to the implementation of experimental tests. The work involved will range from guiding parameter choices for experimentation, analysis of available datasets (particularly massive genomics datasets), as well as new experimental/empirical tests with tight feedback loops between theorists and empiricists. Each team will develop position papers outlining and explaining the scope of the relevant mathematical innovations and how they will alter accepted conceptions of the structure, function, and dynamics of biological systems. A final all-PI meeting with presentation of findings will coincide with the conclusion of the grant period.

5 Milestones

M1: Identify fundamental biological questions in mathematical language that address problems within each of the identified vertices as well as those that span all scales.

M2: Develop theory that reconciles past observations and has predictive power. Implement computational models of theory that demonstrate the potential of theory to make novel re-assessments of extant datasets and/or novel linkages between existing models.

M3: Test theory in biological systems. Identify measurable microscopic parameters that are coupled to macrostate dynamics and implement experimental tests and biological data analysis techniques designed to evaluate theory at multiple scales.

6 Budget

We estimate the following costs for the implementation of the statement of work, divided into three phases (12 months, 18 months, 18 months) over a four year time-table.

	Year			
	1	2	3	4
Phase 1 (12 months)	2.5 mil			
Phase 2 (18 months)		3.5 mil	1 mil	
Phase 3 (18 months)			4 mil	5 mil

References

- R. Albert and A. L. Barabasi. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, 2002.
- V. Arnol’d. Singularity theory. In *Development of Mathematics 1950–2000*, pages 63–95, Birkhäuser, Basel, 2000.
- B. Bassler. Small talk. cell-to-cell communication in bacteria. *Cell*, 17:421–4, 2002.
- J. M. Carlson and J. Doyle. HOT: Robustness and design in complex systems. *Phys. Rev. Lett.*, 84:2529–32, 2000.
- J. Cisterna, C. W. Gear, S. A. Levin, and I. G. Kevrekidis. Equation-free modelling of evolving diseases: coarse-grained computations with individual-based models. *Proc. Roy. Soc. Lon. A*, 460:2761–79, 2004.
- J. E. Cohen. Mathematics is biology’s next microscope, only better; biology is mathematics’ next physics, only better. *PLoS Biol.*, 2:e439, 2004.
- S. F. Earl and M. W. Deem. Evolvability is a selectable trait. *Proc. Natl. Acad. Sci. USA*, 101:11531–6, 2004.
- S. Gavrillets. *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton, NJ, 2004.
- G. Gimmet. *Percolation*. Springer, Berlin, 1999.
- A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C. K. Peng, and H. E. Stanley. Fractal dynamics in physiology: alterations with disease and aging. *Proc. Natl. Acad. Sci. USA*, 19:99, 2002.

- J. Hofbauer and K. Sigmund. *Evolutionary Games and Percolation Theory*. Cambridge University Press, Cambridge, U.K., 1998.
- H. B. Judson. *The Eighth Day of Creation*. Simon & Schuster, New York, 1979.
- M. Kapovich, J. Millson, and T. Treloar. The symplectic geometry of polygons in hyperbolic 3-space. *Asian J. Math.*, 4:123–164, 2000.
- R. E. Lenski, C. Ofria, T. C. Collier, and C. Adami. Genome complexity, robustness and genetic interactions in digital organisms. *Nature*, 400:618–9, 1999.
- D. Mumford. The dawning of the age of stochasticity. In V. Arnold *et al.*, editor, *Mathematics: Frontiers and Perspectives*. American Mathematical Society, Providence, R.I., 1999.
- M. E. J. Newman. The spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, 2002.
- L. Pachter and B. Sturmfels, editors. *Algebraic Statistics for Computational Biology*. Cambridge University Press, 2005.
- S. Park, P. Wolanin, E. Yuzbashyan, H. Lin, N. Darnton, J. Stock, P. Silberzan, and R. Austin. Influence of topology on bacterial social interaction. *Proc. Nat. Acad. Sci. USA*, 100:13910–5, 2003.
- F. Partensky, W. R. Hess, and D. Vaultot. *Prochlorococcus* a marine photosynthetic prokaryote of global significance. *Micrbiol. Mol. Biol. Rev.*, 63:106–27, 1999.
- O. Pourqu e. The segmentation clock: converting embryonic time into spatial pattern. *Science*, 301:328–30, 2003.
- J. Venter *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304:58–60, 2004.