MORGAN & CLAYPOOL PUBLISHERS
LIFE SCIENCES

Introduction to Quantitative Cell Biology

Wallace F. Marshall

COLLOQUIUM LECTURES ON
QUANTITATIVE CELL BIOLOGY

Series Editor: *Wallace F. Marshall*

Introduction to Quantitative Cell Biology

Colloquium
Digital Library of Life Sciences

The *Colloquium Digital Library of Life Sciences* is an innovative information resource for researchers, instructors, and students in the biomedical life science community, including clinicians. Each PDF e-book available in the *Colloquium Digital Library* is an accessible overview of a fast-moving basic science research topic, authored by a prominent expert in the field. They are intended as time-saving pedagogical resources for scientists exploring new areas outside of their specialty. They are also excellent tools for keeping current with advances in related fields, as well as refreshing one's understanding of core topics in biomedical science.

For the full list of available titles, please visit:
colloquium.morganclaypool.com

Each book is available on our website as a PDF download. Access is free for readers at institutions that license the *Colloquium Digital Library*.

Please e-mail info@morganclaypool.com for more information.

Colloquium Series on Quantitative Cell Biology

Series Editor

Wallace F. Marchall

Department of Biochemistry and Biophysics, University of California, San Francisco

A fundamental unsolved problem in biology is understanding how a living cell emerges from the multitude of molecular components. While cell biology has made great strides in enumerating all the components of the cell, this is only just the beginning, and the challenge we now face is understanding the cell as a complex, self-organizing system. To meet this challenge, we must take cell biology to a quantitative level, combining mathematical modeling with new methods in measurement and data analysis. The goal of this e-book series is to provide an overview of current approaches and challenges in the emerging field of Quantitative Cell Biology, in a way that will be accessible to readers both from the biological sciences as well as the physical and computational sciences. These state of the art volumes introduce readers to the cutting edge research in the field, including computational modeling and image analysis methods, while also discussing current understanding and open questions in the systems biology of cells. Each book is intended to be useful independent of the others, and the series as a whole will provide a comprehensive introduction for students and researchers who are new to the field.

Published titles

For titles please see the website, www.morganclaypool.com/toc/qcb/1/1.

Introduction to Quantitative Cell Biology
Wallace F. Marshall
www.morganclaypool.com

ISBN: 9781615046683 paperback

ISBN: 9781615046690 ebook

DOI: 10.4199/C00121ED1V01Y201409QCB002

A Publication in the

COLLOQUIUM SERIES ON QUANTITATIVE CELL BIOLOGY

Lecture #1

Series Editor: Wallace F. Marshall, Department of Biochemistry & Biophysics, University of California, San Francisco

Series ISSN

ISSN 2375-7744 print

ISSN 2375-7752 electronic

Introduction to Quantitative Cell Biology

Wallace F. Marshall
Director, Center for Cellular Construction
Professor, Department of Biochemistry and Biophysics
University of California San Francisco
San Francisco, CA

COLLOQUIUM SERIES ON QUANTITATIVE CELL BIOLOGY #1

MORGAN & CLAYPOOL LIFE SCIENCES

ABSTRACT

For the past decade or more, much of cell biology research has been focused on determining the key molecules involved in different cellular processes, an analytical problem that has been amenable to biochemical and genetic approaches. Now, we face an integrative problem of understanding how all of these molecules work together to produce living cells, a challenge that requires using quantitative approaches to model the complex interactions within a cell, and testing those models with careful quantitative measurements.

This book is an introductory overview of the various approaches, methods, techniques, and models employed in quantitative cell biology, which are reviewed in greater detail in the other volumes in this e-book series. Particular emphasis is placed on the goals and purpose of quantitative analysis and modeling, and the special challenges that cell biology holds for understanding life at the physical level.

KEY WORDS

quantitative cell biology, modeling, image analysis, numerical methods, biophysics, parameter estimation, model discrimination, exploratory data analysis, spatial statistics, ARMA models, machine learning, high content image-based screens, coarse-graining, boundary value problems, network models, rule-based modeling, agent based modeling, stochastic modeling, computer aided design, kinetochore, cytokinesis, endocytosis, computational fluid dynamics, level set method, multiscale modeling approaches, morphogens, differential equations, probability theory, statistical analysis

Contents

CHAPTER 1

Overview: What Is Quantitative Cell Biology?

The living cell is a mechanochemical machine of almost unimaginable complexity, but at some level, all of its activities can be reduced to a combination of biochemical processes and physical processes. The challenge in understanding cells is understanding how all the parts work together, and how physics and biochemistry combine to produce life. The field of cell biology started out as a descriptive science, based on visualizing cellular structures by light and electron microscopy. In subsequent decades, work focused on determining the "parts list" of molecules that make up the cell, but this is still an essentially descriptive approach. One might think that determining the functions of the individual molecular components of a cell would provide all the answers, but it did not. The problem is that unlike the comparatively simple processes that biochemistry and molecular biology were able to explain, such as digestion of sugar or activation of gene expression at a particular promoter, cell biology can only be understood at the level of entire systems in which large numbers of molecules work together to produce emergent behaviors. The field of quantitative cell biology seeks to address the need for methods to measure, model, and predict complex behaviors at the level of cells and sub-cellular processes.

1.1 MODELING TO BRIDGE THE GAP IN SCALES

Many of the most interesting unanswered questions in cell biology exist at a mesoscale between the atomic scale (measured in Angstroms) and the spatial scale of light microscopy (measured in microns). At this intermediate spatial scale, direct visualization of key events is extremely difficult because they involve groups of molecules too large to resolve by X-ray crystallography but too small to resolve by light microscopy. Furthermore, cellular processes are inherently cooperative, arising from interactions of large numbers of components and mediated by energy-consuming pathways that allow spontaneous generation of order from disorder. This combination of a gap in direct measurement and the importance of complex interactions of large numbers of components means that modeling and

theoretical approaches are necessary to span this gap in understanding (Mogilner, 2006). The centrality of modeling in this process creates a need for quantitative measurements and methods for data analysis. The combined application of computational modeling and quantitative data analysis to fundamental questions in cell biology has created a new discipline: quantitative cell biology.

1.2 EMERGENT PROPERTIES AND SELF-ORGANIZATION

Traditional concepts and methods of biochemistry and molecular biology are extremely powerful for dissecting mechanism in cases where the observable process is the direct outcome of a single molecule or complex. Consider, for example, DNA replication. Since a single polymerase complex can drive the incorporation of nucleotides, it is possible to isolate the complex and determine its kinetic properties using simple enzymatic assays. There is no need to consider any other components in the cell, since the enzyme in itself is sufficient to produce the phenomenon under study. The key is that a process like DNA replication can be localized to a single point in the cell and assigned to a single enzymatic function. But other processes are not subject to this type of localization. Cell division and motility, for example, are the collective result of hundreds of different molecular plays distributed in broad swathes across the cell. Rather than being a direct outcome of a single enzymatic activity, these phenomena are emergent properties of huge molecular collectives. It is thus very hard, maybe even impossible, to truly understand emergent behaviors by considering one individual molecular species at a time—that is, using the conceptual framework of molecular biology. Does this mean such systems are incomprehensible? Not at all—emergent properties are studied all the time in condensed matter physics, and all it takes is a way to represent the pertinent level of organization at a suitable level of abstraction. For instance, solid–liquid phase transitions can be understood by thinking about order parameters and phase diagrams, without needing to consider the detailed behavior of every single molecule.

Self-organization is one type of emergent property that is highly germane to cell biology. Although the genome of a cell is often likened to a blueprint, it is more like a recipe book that specifies what components should be made, but not how they should be put together. The actual physical structure of the cell apparently results from self-organization of the components. For example, despite the apparent complexity of the mitotic spindles, bipolar spindles are able to self-assemble from components in vitro (Heald et al., 1996), and presumably, these same assembly processes play a key role in spindle assembly in living cells (Pavin and Tolic, 2016). Self-organization of structures is seen in physical systems as well. One of the goals of quantitative cell biology is to view cell structure and behavior as emergent, self-organizing phenomena, and to try to analyze them using similar

mathematical and physical approaches used in condensed matter physics. This program of trying to understand cellular structure and behavior is a modern extension of the goals initially stated almost a century ago by Thompson (1942).

1.3 PREDICTIVE UNDERSTANDING OF CELLULAR SYSTEMS

The goal of studying cell biology is to understand how cells work at a mechanistic level. Mechanisms are commonly described in words or in diagrams. Indeed, it is common practice to formulate a cartoon-like model to summarize our understanding about a given cell biological process. Such cartoons fill modern cell biology textbooks and review articles, and often occur in research articles as a way to encapsulate the conceptual take-home message of the work. But how does one decide whether an appealing-looking cartoon, or a convincingly worded descriptive model, actually corresponds to reality? How do we know whether we really understand how a system works?

What makes science different from other branches of knowledge is that we test our tentative understanding with experiments, which ultimately amounts to asking whether our "model," whether formulated as words or diagrams or in some other form, is sufficient to predict how the system will respond to defined perturbations. This is the very essence of what an experiment is. The trick becomes how to know whether or not a prediction has been satisfied. Sometimes, a prediction can be formulated as something that is very black-and-white: for example, we may predict that if gene X is knocked out, our cells will die. In other cases, however, the prediction may not be something as obvious and easy to decide as life or death.

1.4 HOW DOES QUANTITATIVE CELL BIOLOGY DIFFER FROM SYSTEMS BIOLOGY?

Systems biology and quantitative cell biology share the common feature of adapting tools traditionally employed in the physical sciences to better understand the complexity of living systems, but they differ greatly in the level of biological organization that they address. Systems biology currently places virtually all of its emphasis on molecular genetic regulatory pathways, and rarely if ever addresses the level of the entire cell, or even of sub-cellular components or organelles. As a result, the computational and analytical tools in common use in systems biology, as well as data standards and representations, are not directly applicable to quantitative cell biology data. The fields are therefore distinct.

However, quantitative cell biology has much to learn from systems biology as a field, particularly in the way that the field of systems biology has made use of standardized data formats and modeling languages such as Systems Biology Markup Language (SBML). SBML is a formal language in which mathematical models of biochemical pathways can be represented in a machine-readable form that can then be used by many different software packages, for example, to visualize the pathway in network form, or to run simulations of the dynamic behavior of the network. In many respects, the types of data that systems biology studies, that is, gene expression profiles, are far easier to analyze than Cell Biology data, because the data usually lacks a spatial component. Thus, quantitative cell biology poses unique challenges for understanding, representing, and manipulating data, that systems biology of gene networks does not face. Developing formal languages to represent models of the spatially complex and dynamic behaviors of cells remains a challenging problem.

CHAPTER 2

Quantifying Data

Traditionally, cell biology has relied heavily on image data and biochemical data, such as Western blots, usually with the goal of obtaining qualitative results such as showing that a particular protein localizes near a particular organelle and so on. Such approaches have yielded tremendous advances in our understanding of how cells work. However, there are a number of compelling reasons for wanting cell biological data to be more quantitative in nature, that is, results that are expressed in terms of numbers.

2.1 SUMMARIZING AND VISUALIZING LARGE NUMBERS OF EXAMPLES

Perhaps the simplest reason for wanting numerical measurements is that these measurements allow large numbers of samples to be summarized and aggregated in a single figure. By having numerical measurements obtained as a function of time, for example, one can generate a graph showing the dynamics of the system, thus summarizing the results of many, perhaps hundreds, of time-lapse image series, all in a single plot. Although it might in principle be possible to look at all of the raw image sequences, this becomes prohibitive as the number and size of the files increases.

Moreover, in scientific publications, it is not reasonable to expect every reader to go through the exercise of laboriously viewing hundreds of movies. Nobody will do it. Although some cell biology journals have emphasized the storage and distribution of raw data files, and certainly this is important for specialized purposes, in routine practice such files are essentially useless because they contain too much data. In contrast, simple graphical representations, based on numerical measurements, are far easier to view and draw conclusions from.

As an extension of this application, once a mathematical model is determined to represent key aspects of a system, it is sometimes possible to come up with combinations of system parameters that allow data to be re-plotted in a much simpler form. For example, Figure 1 shows how levels of gene expression change as a function of repressor concentration under many different conditions. The resulting graphs look like a mish-mash of different behaviors. But when the results are replotted