Integrative cell models for disease intervention

Matt Scott (University of Waterloo),
Peter Swain (University of Edinburgh),
Hans Othmer (University of Minnesota)

June 10 – June 15, 2018

Biological processes do not occur in isolation and their appropriate execution requires communication and coordination across the cell. We know that signals are conveyed via interactions between proteins and between proteins and DNA, but such regulatory interactions are not the sole drivers of cellular responses. Changes in the physiological composition of the cell and particularly in the levels of common resources, such as energy and raw materials, also provides a higher level of regulation. The extent of the control provided by this potentially primordial regulation has only recently been appreciated. Yet its effects are felt widely, ranging from the development of antibiotic resistance to the production of chemicals in the biotechnology industry. The broad ambition of this workshop was to foster the development of a new mathematical framework for modelling cellular processes that includes global regulation, either implicitly or explicitly, and so enable quantitative prediction particularly focused on disease intervention.

Overview

Biological cells, whether as single-cell organisms or as part of a tissue, comprise many component subsystems, such as the nucleus and the mitochondria, and execute many processes, such as nutrient uptake, metabolism, biosynthesis, and gene expression, which interact to define the state and function of the cell. These components and processes typically are strongly interacting, but the traditional approach has been to study cellular systems under the implicit assumption that the component or process in question can be understood in isolation from the remaining ones, which merely provide a constant, background environment. This assumption was necessary to make progress in early investigations, but recent research has now shown that the assumption is weak, at best. In systems ranging from microbes to human cancers, the environment and growth rate can affect the level of expression both of single genes, even those with 'constant' unregulated gene expression, and of hundreds of genes across the entire genome.

Taken together, these developments question our understanding of cellular regulation and whether the biochemical networks in cells can be usefully decomposed into separate modules, the current paradigm in mathematical and systems biology. This fundamental limitation of existing methods becomes particularly acute in the mathematical modelling of disease states. The interdependence of cellular processes confounds disease modelling in two important ways. First, diseases can induce system-wide changes in gene expression making it difficult to distinguish proximal from distal causes. Second, a lack of modularization in cellular function makes the success of intervention strategies difficult to predict. Often brute-force elimination of what is thought to be the primary molecular cause of disease has no effect: the cell simply compensates through auxiliary and highly redundant channels. What is needed is an understanding of disease that transcends molecular mechanism to provide a system-level view of the phenomenon of how the complex interconnected networks integrate different signals to produce the observed outputs. Gaining this understanding requires mathematical models and new analytical techniques in order to integrate the effects of various levels of regulation and physiology.

Cell physiology imposes strong constraints on how the cell allocates its internal resources. Broadly speaking, to support rapid growth rates, the cell must devote a large fraction of its synthetic machinery to make more of that synthetic machinery. This allocation necessarily comes at the expense of producing other cellular constituents, such as proteins responsible for nutrient assimilation and for repair and maintenance of

DNA. As a consequence, in addition to the local, point-to-point concept of regulation, there is a more diffuse and global layer arising from constraints that exist because of the allocation of cellular resources. This primordial regulation has a breadth and impact that could completely dominate local interactions. In other words, the information flow in a cell can be carried by 'signals' as commonly understood – concentrations of specific molecules – but also, perhaps more generally, by macrovariables of the physiological status of the cell. An important example arises in cancer, where the metabolic state both affects and is controlled by up-regulation of transporters and enzymes and by the micro-environment in which a cell finds itself. This can lead to supracellular integration of signals via a process called symbiotic metabolism, in which hypoxic cells produce more lactate that is used by non-hypoxic cells, thereby freeing glucose for cells in hypoxic regions of a tumour. The realization that this global level of control is important in cellular functioning calls for a new class of integrative models for cellular systems.

Quantitatively understanding such physiology-dependent disease modeling poses substantial challenges, both experimentally and particularly mathematically. We do not know what level of description of the 'rest of the cell' is appropriate; we do not know how to correctly connect existing models of different cellular phenomena; we do not know the best type of data to parameterize such modelling or have algorithms in place to carry out this parameterization; we do not even know which cellular processes are core and should be modelled.

Two distinct approaches to this context-dependent modelling are being developed. The first, which we will call coarse-grained modelling, has focused on phenomenological models that aim to capture sufficient information to answer questions of interest with a minimal mathematical description. These approaches have uncovered what appear to be fundamental growth 'laws' demonstrating apparent simplicity – such as linear dependencies on the growth rate – despite the underlying biological complexity [1]. The second methodology, which we will call fine-grained modelling, provides context by including as much of the cell as possible in the model. A recent triumph is the 'whole cell' model of *Mycoplasm genitalium*, a pathogenic bacterium with an unusually small genome, in which, in an engineering feat, models that operate at different spatial scales and over different time scales were linked together and coupled to data to produce a simulation that includes the majority of the biochemical processes we currently know exist in that cell type [2].

The goal of this workshop was to bring together practitioners of both the fine and coarse-grained methodologies along with numerical analysts and those working at the interface of context-dependent modelling with medicine and biotechnology. Our focus was on how physical and biological trade-offs at one level of organization constrain behaviour at other levels. By exploring how both methodologies might be integrated, how models at different scales may be formally combined, and how models might best be simulated and quantitatively compared with data, we aimed to provide the momentum and ideas to develop a framework for physiologically-dependent disease modelling that will drive the field of cellular mathematical biology for the foreseeable future. Different modelling strategies, ranging from Boolean to continuum models, were represented by participants.

The workshop was timely. Experimental advances are generating data on how cellular physiology changes because of both cell growth and human intervention, such as the additions of antimicrobials and of synthetic circuits. New biophysics now allows physiological variables, such as pH, ATP and NADH levels, and membrane potential, to be measured in single cells. Further, automation has led to high-throughput technology that measures physiological variables, such as biomass, pH, dissolved oxygen, fluorescent molecules, levels of NADH and NADPH, and volume, in tens of parallel bioreactors with either continuous or fed-batch growth or at the single-cell level in a so-called 'mother machine' [3]. Mathematical models do not yet exist to allow quantitative comparison with such data: a gap that this workshop aimed to fill.

In addition, context-dependent effects are hindering our understanding of the development of antimicrobial resistance and of tumour development. Models that include cellular physiology can predict growth rate, the standard measure of fitness for microbes. Such information is crucial if we are to reduce antibiotic and anticancer resistance because we have to understand how the molecular mechanisms that confer resistance also determine fitness. A better understanding of how cellular context affects disease and disease intervention at various levels will lead to significant advances in medicine, biology and the biotechnology industry, and our workshop was one step in that direction.

Presentation highlights

Matt Scott (U Waterloo) kicked off the workshop with an overview of physiological constraints on bacterial response to antibiotics [1]. Two cases studies were examined: the first connecting mechanistic parameters to the efficacy of chemical antibiotics [4]; the second connecting pre-infection growth state to susceptibility to phage infection. **Ting Lu** (U Illinois - UC) continued the theme of physiological constraints, but in the context of rational design of synthetic genetic elements in bacteria [5]. The goal is to standardize components to the extent that the design-build-test cycle so familiar in engineering can be efficiently implemented in synthetic biological systems. **David Umulis** (Purdue) showcased a variety of sophisticated mathematical and image-analysis tools he has created for tracking development in zebra fish embryos. These tools were decisive in discriminating a variety of proposed mechanisms for gene expression patterning in the developing embryo [6]. **Tomas Gedeon** (Montana State University) spoke about a generalized theoretical framework for converting large, poorly characterized mathematical models into tractable quasi-Boolean networks [7]. Within this framework, he is able to rapidly identify parameter regimes that are necessary to produce, for example, hysteretic trajectories in the full model [8]. Back-to-back talks in the evening rounded out the first day by returning to operating constraints in bacterial physiology. **Bas Teusink** (VU Amsterdam) presented joint theoretical and experimental work linking elementary-flux-mode analysis of metabolic networks with minimal-constraint criteria in flux balance analysis. Under general hypotheses, he was able to prove that the optimal number of elementary flux modes operating in steady-state are less-than-or-equal-to the number of constraints operating on the system [9]. **Frank Bruggeman** (VU Amsterdam) followed with an extension of the elementary-flux analysis to growth transitions. His object of study was the control structure that steers a bacterial system toward an optimal growth state. He was able, along with Planqué, to prove that the number of metabolites that bind regulator proteins is greater-to-or-equal to the number of environmental parameters to which the system is robust [10].

On the second day, **Edda Klipp** (Humboldt Berlin) spoke about her ambitious research program to create a comprehensive mathematical model of yeast growth, comprised of modular sub-systems characterizing metabolism, gene expression, cell cycle, volume growth, membrane transport and signalling [11, 12]. Moving from microorganisms to multicellular systems, **Hans Othmer** (U Minnesota) presented a mechanistic model for intercellular interactions in the wing disk development of *Drosophila*. Othmer and coworkers have constructed a detailed kinetic model to inform future experiments. Future work seeks to show that the model recovers both observed scale invariance and uniform growth across the disk. **Erik van Nimwegen** (Biozentrum, Basel) presented a large collection of single-cell data that challenges existing views of regulation in bacteria. First, the novel idea that regulators perform a 'noise-propagation' role in addition to their conditioned response [15]. Second, that the lag-time to respond to a nutrient shift (lactose) exhibits a long-tailed distribution at the single-cell level [14]. According to van Nimwegen, both noise-propagation and long-tailed response distributions serve an evolutionary purpose. **Carla Bosia** (Politecnico di Torino) presented experimental and theoretical work on the growth of human leukemia cells. Bosia uses inoculum size as a means of adjusting proliferation rate, but does not observe the same empirical relations coupling ribosome abundance and proliferation rate that are observed in bacterial systems. Returning to bacterial physiology, **Teuta Pilizota** (U Edinburgh) described a methodology using the flagellar motor of *E. coli* as an endogenous 'voltmeter' to track intracellular energy levels. This exciting technique was used to quantify the difference between indole and UV radiation damage.

The third day began with **Jia Gou** (U Minnesota) who spoke about theoretical work to both analyze and visualize the variety of possible patterns forming when three ligands compete for a single receptor. **Eldon Emberly** (Simon Fraser) presented examples of spatial organization in bacteria. The first focused on entropy-driven localization of protein aggregates [18, 16]. The second, more recent work using a diffusion-driven ratchet-walker model to analyze intracellular transport via the ParAB system [17]. The final talks of the day were devoted to antibiotic resistance in bacteria. **Mary Dunlop** (Boston U) presented exciting experimental data [20, 21] and mathematical modeling [19] that suggested that transient resistance to antibiotics (via stochastic induction of the MarAB system) predisposes populations to develop permanent resistance. **Rosalind Allen** (U Edinburgh) followed with experimental data and mathematical modeling showing the physiological consequences of cell-wall targeting antibiotics. Though still in the early stages, Allen's phenomenological model was able to recapitulate many of the unusual features of this much-prescribed class of antibiotics.

Kamila Larripa (Humboldt State) started off the fourth day reporting on an industry-driven effort to improve treatment of myeloma (a blood cancer). Larripa and collaborators have developed a mathematical framework to explore effects of specific drugs and their targets. **Luca Ciandrini** (U Montpellier) presented a mechanistic mathematical model of protein translation, based on prior work characterizing ribosome procession as a totally-asymmetric-simple-exclusion process. The work suggests that circularization of mRNA transcripts can lead to increased local concentration of recycling ribosomes and that this effect can be well-quantified by a simple model [22]. **Meriem el Karoui** (U Edinburgh) considered in detail the mechanistic and stochastic elements of DNA damage repair in *E. coli*. First, the DNA damage response appears to inherit growth-rate dependence from the physiology of the organism. Second, repair proteins repair in low basal abundance, yet the variability is far less than expected [23] – the reason for this is unknown. **Bruno Martins** (Cambridge) used experimental characterization of the circadian rhythms of the cyanobacterium *Synechococcus elongatus* to illustrate that population heterogeneity can obscure underlying strategies of cell-size homeostasis [25, 24]. Martins has developed a simple mathematical model to quantitatively account for the population heterogeneity. **Ariel Amir** (Harvard) concluded the fourth day with an ambitious survey of past and present mathematical models of bacterial growth and division. In contrast to classical work by Powell, Amir determines conditions under which variance in the division time can increase population growth rate [28, 26, 27].

Jim Greene (Rutgers) began the final day of the workshop with a focus on induced drug-resistance in chemotherapy. A simplified mathematical model was used to explore the control structure that optimizes treatment efficacy. Echoing many of themes of previous speakers, **Stefan Klumpp** (U Goettingen) provided an overview of work coupling bacterial physiology to population heterogenity. The survey included efforts to model the effect of cell growth as an additional pseudo-regulator [29], modulation of the inferred drug efficacy depending on whether the cells are grown in bulk or in a single-cell 'mother machine' '[30], and the physiological characterization of a *Bacillus subtilus* strain with reduced genome size. **Vahid Shahrezaei** (Imperial College London) presented several efforts to extend coarse-grained physiological models to fission yeast, as well as efforts to provide greater detail in the models [31, 32]. In the final talk, **Peter Swain** (U Edinburgh) outlined a graph-theoretical approach to enumerate and visualize the outcome of competitive evolutionary dynamics. The methodology enables incomplete information to be leveraged to obtain increasingly accurate prediction of the outcome of the competitive evolutionary dynamics [33].

Outcomes and open questions

The two major outcomes of the workshop were the recognition of the surprising similarity between bacterial and cancer models and to identify clear gaps in our knowledge of cell physiology.

As organizers, we deliberately chose participants who likely knew less than 20% of the other speakers. As a result, from side-by-side talks about cancer biology and antimicrobial resistance, it became clear that much of the framework of the mathematical modelling is common between these two very different systems.

Nonetheless, the incorporation of physiological constraints into mechanistic mathematical models is far more advanced in bacterial studies, whereas little in the way of robust empirical relationships linking proliferation rate and the macromolecular state have been established in cancer cell lines. Carla Bosia's work illustrates that one of the major obstacles is determining a reliable method to modulate proliferation rate.

The second major outcome was a broad recognition of how little is known about the coupling between cell growth and DNA replication, even in bacteria. In 2015, several groups used single cell data to establish that in *E. coli*, the size at birth (s_b) is related to the size at division (s_d) via an *adder* mechanism, $s_d = s_b + \Delta$, where Δ is a constant (though growth-rate dependent) increment of mass. The mechanism underlying this behaviour remains unresolved. A third of the participants formed an impromptu focus session to clarify the current state of the field. Unpublished data from the van Nimwegen group in particular suggests that the simple 'adder' model is inadequate.

Altogether, the open questions that emerged are:

1. How can we reliably modulate the proliferation rate of human cell lines in a way to sustain exponential growth over many generations?

2. Are the complex adaptations of cancer cell lines (e.g. the Warburg effect) the result of physiological

constraints (as appears to be the case in bacteria [34])?

3. How are cell division and DNA replication coordinated in proliferating cells? (This has been an open question for fifty years.)

Participant response

Several of the participants remarked that this workshop was the most productive conference or workshop that they had ever attended. Some of these are reproduced below:

> I wanted to let you know how much I enjoyed the BIRS workshop. Thanks to that meeting I have already initiated two new collaborations (with Erik [van Nimwegen (Basel)] and Ariel [Amir (Harvard)]) and I have learned an enormous amount about both the models currently used for bacterial diseases and also about the potential similarities with models for cancer treatment and drug resistance in cancer (i.e. Kamila [Larripa (Humboldt)] and Jim's [Greene (Rutgers)] talks). I particularly enjoyed the format of the meeting with long talks that allowed for a lot of interactions between the speaker and the audience as well as enough free time to work with specific colleagues and to enjoy the magnificent surroundings. I think that the fact that we were a relatively small group also helped a lot with making the meeting very interactive: I think I had productive discussion with almost all the participants. Last but not least I was very impressed that the gender balance in the meeting was very good which in my experience is rarely the case. The facilities were excellent and the provision of small room with boards for discussion was particularly useful. I also enjoyed the swimming pool! In summary, this was the best meeting I have attended in a number of years. Thanks again for organizing this workshop!
>
> *– Meriem el Karoui (U Edinburgh)*

> This was one of the best conferences I attended, for several reasons:
>
> - The topic was focused enough such that my work has significant overlap with many of the participants. Nevertheless I have not met many of them previously and therefore this was extremely stimulating scientifically, both in terms of feedback and ideas related to my own work as well as learning about recent progress in the field.
>
> - The program was organized in a way that allowed for many off-line discussions, which were often as useful as the formal talks. I hope to keep in touch with new people I met at BIRS, and I think that there is also a good chance that this conference will trigger new collaborations.
>
> *– Ariel Amir (Harvard)*

> The BIRS conference was a wonderful experience. I found it extremely useful to hear about work slightly peripheral to my own area of research, especially from senior and esteemed mathematicians. I work in cancer biology, but a number of mathematical techniques and models being used in bacterials systems are directly applicable to my project– this was wonderful to discover. This has opened up a number of potential directions I plan to pursue, and I have also found some wonderful collaborators. The group was extremely accomplished and I benefitted not just from listening to formal talks, but from the questions I received about my own work. Meals and breaks provided wonderful opportunities for discussion, and the small group size really encouraged this. Big thanks to the organizers and to BIRS for providing such a wonderful scientific program!
>
> – Kamilla Larripa (Humboldt)

References

[1] M. Scott *et al.*, Interdependence of cell growth and gene expression: Origins and consequences, *Science* **330** (2010), 1099–1102.

[2] J. R. Karr *et al.*, A Whole-Cell Computational Model Predicts Phenotype from Genotype, *Cell* **150** (2012), 389–401.

[3] P. Wang *et al.*, Robust Growth of Escherichia coli, *Current Biology* **20** (2010), 1099-1103.

[4] P. Greulich *et al.*, Growth-dependent bacterial susceptibility to ribosome-targeting antibiotics, *Molecular systems biology* **11** (2015), 796.

[5] C. Liao, A. Blanchard and T. Lu, An integrative circuit-host modeling framework for predicting synthetic gene network behaviors, *Nature Microbiology* **2** (2017), 1658–1666.

[6] J. Zinski *et al.*, Systems biology derived source-sink mechanism of BMP gradient formation, *eLIFE* **6** (2017), e22199.

[7] B. Cummins *et al.*, DSGRN: Examining the Dynamics of Families of Logical Models, *Frontiers in physiology* **9** (2018), 549.

[8] T. Gedeon *et al.*, Identifying robust hysteresis in networks, *PLoS Computational Biology* **14** (2018), e1006121.

[9] D. H. de Groot *et al.*, Maximal growth rate requires minimal metabolic complexity, *BioRxiv* 167171.

[10] M. T. Wortel *et al.*, Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield, *PLoS Computational Biology* **14** (2018), e1006010.

[11] K. Tummler, C. Kuhn and E. Klipp, Dynamic metabolic models in context: biomass backtracking, *Integrative Biology* **7** (2015), 940–951.

[12] S. Gerber *et al.*, A thermodynamic model of monovalent cation homeostasis in the yeast Saccharomyces cerevisiae., *PLoS Computational Biology* **12** (2016), e1004703.

[13] J. Gou, L. Lin and H. G. Othmer, A model for autonomous and non-autonomous effects of the Hippo pathway in Drosophila, *bioRxiv* 128041.

[14] M. Kaiser *et al.*, Monitoring single-cell gene regulation under dynamically controllable conditions with integrated microfluidics and software, *Nature Communications* **9** (2018), 212.

[15] L. Wolf, O. K. Silander and E. van Nimwegen, Expression noise facilitates the evolution of gene regulation, *eLIFE* **4** (2015), e05856.

[16] L. Jindal and E. Emberly, Operational Principles for the Dynamics of the In Vitro ParA-ParB System, *PLoS Computational Biology* **11** (2015), e1004651.

[17] M. R. Keramati *et al.*, Confinement-dependent localization of diffusing aggregates in cellular geometries, *Physical Review E* **91** (2015), 012705.

[18] K. Scheu *et al.*, Localization of aggregating proteins in bacteria depends on the rate of addition, *Frontiers in Microbiology* **5** (2014), 418.

[19] J. Garcia-Bernardo and M. J. Dunlop, Tunable Stochastic Pulsing in the Escherichia coli Multiple Antibiotic Resistance Network from Interlinked Positive and Negative Feedback Loops, *PLoS Computational Biology* **9** (2013), e1003229.

[20] I. El Meouche, Y. Siu and M. J. Dunlop, Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells, *Scientific Reports* **6** (2016), 19538.

[21] A. M. Langevin and M. J. Dunlop, Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells, *Scientific Reports* **6** (2016), 19538.

[22] L. D. Fernandes, A. P. S. de Moura and L. Ciandrini, Gene length as a regulator for ribosome recruitment and protein synthesis: theoretical insights, *Scientific Reports* **7** (2017), 17409.

[23] B. Okumus *et al.*, Mechanical slowing-down of cytoplasmic diffusion allows in vivo counting of proteins in individual cells, *Nature Communications* **7** (2016), 11641.

[24] B. M. C. Martins *et al.*, Frequency doubling in the cyanobacterial circadian clock, *Molecular Systems Biology* **12** (2016), 896.

[25] B. M. C. Martins and J. O. W. Locke, Microbial individuality: how single-cell heterogeneity enables population level strategies, *Current Opinion in Microbiology* **24** (2015), 104–112.

[26] J. Lin and A. Amir, *BioRxiv* (2018), 255950.

[27] P.-Y. Ho, J. Lin and A. Amir, Modeling Cell Size Regulation: From Single-Cell-Level Statistics to Molecular Mechanisms and Population-Level Effects, *Annual Review of Biophysics* **47** (2018), 251–271.

[28] J. Lin and A. Amir, The Effects of Stochasticity at the Single-Cell Level and Cell Size Control on the Population Growth, *Cell Systems* **5** (2017), 358.

[29] S. Klumpp and T. Hwa, Bacterial growth: global effects on gene expression, growth feedback and proteome partition, *Current Opinion in Biotechnology* **28** (2014), 96–102.

[30] A. Roy and S. Klumpp, Simulating Genetic Circuits in Bacterial Populations with Growth Heterogeneity, *Biophysical Journal* **114** (2018), 484–492.

[31] V. Shahrezaei and S. Marguerat, Connecting growth with gene expression: of noise and numbers, *Current Opinion in Microbiology* **25** (2015), 127–135.

[32] F. Bertaux, S. Marguerat and V. Shahrezaei, Division rate, cell size and proteome allocation: impact on gene expression noise and implications for the dynamics of genetic circuits, *Royal Society Open Science* **5** (2018), 172234.

[33] C. Josephides and P.S. Swain, Division rate, cell size and proteome allocation: impact on gene expression noise and implications for the dynamics of genetic circuits, *Nature Communications* **8** (2017), 685.

[34] M. Basan *et al.*, Overflow metabolism in E. coli results from efficient proteome allocation, *Nature* **528** (2015), 99–104.