RESEARCH CENTRE

Inria Paris Centre

IN PARTNERSHIP WITH:

Institut Pasteur

2023

ACTIVITY REPORT

Project-Team

INBIO

Experimental and Computational Methods for Modeling Cellular Processes

IN COLLABORATION WITH: Centre de Bioinformatique, Biostatistique et Biologie Intégrative

DOMAIN

Digital Health, Biology and Earth

THEME

Modeling and Control for Life Sciences

Inria

Contents

Project-Team INBIO

Creation of the Project-Team: 2019 November 01

Keywords

Computer sciences and digital sciences

A3.1.1. – Modeling, representation

A3.4.4. – Optimization and learning

A3.4.5. – Bayesian methods

A6.1.1. – Continuous Modeling (PDE, ODE)

A6.1.2. – Stochastic Modeling

A6.1.4. – Multiscale modeling

A6.3.1. – Inverse problems

A6.3.3. – Data processing

A6.4.1. – Deterministic control

A6.4.3. – Observability and Controlability

Other research topics and application domains

B1.1.2. – Molecular and cellular biology

B1.1.7. – Bioinformatics

B1.1.8. – Mathematical biology

B1.1.10. – Systems and synthetic biology

B2.4.2. – Drug resistance

B5.10. – Biotechnology

B9.8. – Reproducibility

1 Team members, visitors, external collaborators

Research Scientist

- Gregory Batt [Team leader, Inria, Senior Researcher, HDR]

Post-Doctoral Fellows

- Angelica Frusteri Chiacchiera [Institut Pasteur, from Jul 2023]
- Allyson Holmes [Institut Pasteur, until Jun 2023]
- Esteban Lebrun [Institut Pasteur, from Oct 2023]

PhD Students

- Alicia Da Silva [Inria, from Oct 2023]
- Henri Galez [Institut Pasteur]
- Viktoriia Gross [Institut Pasteur]

Technical Staff

- Sara Napolitano [Institut Pasteur, Engineer]

Interns and Apprentices

- Konstantin Achkasov [Institut Pasteur, Intern, from Mar 2023 until Aug 2023]
- Hong Duong Ngo [Institut Pasteur, Intern, from Feb 2023 until Aug 2023]

Administrative Assistant

- Nelly Maloisel [Inria]

Visiting Scientists

- Angelica Frusteri Chiacchiera [University di Pavia, until Jun 2023]
- Lorenzo Pazotti [University di Pavia]

External Collaborator

- François Bertaux [Lesaffre]

2 Overall objectives

The main objective of our research is to understand, control, and optimize cellular processes in single cells and at the population level. We combine experimental and theoretical work within a single team.

Our focus is on developing methods and models that take stochasticity of intracellular processes and heterogeneity of cell populations into account. To this end, we use both mixed-effects models as well as continuous-time Markov chains and their diffusion approximations. We develop methods for efficiently calculating with such models and use them to design optimally informative experiments and to reverse engineer unknown cellular processes from experimental data. Furthermore, we deploy models in order to optimally construct and optimally control synthetic gene circuits.

We have recently started to set up our own biology laboratory at Institut Pasteur. We develop novel experimental platforms that are designed to be fully automated, controllable by our own software,

and capabale of updating the experimental plan in response to incoming measurements. Optogenetic actuation of intracellular processes, coupled to real time fluorescence measurements by microscopy or flow cytometry, then allows us to connect cellular processes with models and algorithms in real time.

The spritit of our work is that experimental platforms and circuits should be constructed with our theoretical work in mind, while our mathematical methods should be usable to adress concrete experimental questions in the lab.

3 Research program

3.1 Population dynamics emerging from randomness in single cells

Dynamics of cell populations growing in isolation or as part of some ecological system are often shaped by biochemical processes inside cells, for instance when these processes convey resistance to stressors or trigger cell fate decisions in reponse to environmental conditions. **Understanding how stochastic reaction events inside single cells affect emerging population dynamics**, and how selection effects at the population level feed back to shape single cell characteristics of cells in the population, is one of the key questions in biology. We develop **multi-scale modeling approaches** that allow us to derive emerging population dynamics from mechanistic descriptions of stochastic reaction networks inside single cells. In the past, we have used such approaches to study how stochasticity in restriction-modification systems, acting as simple bacterial innate immune systems, propagates to the ecology of bacteria and bacterial viruses and shapes the dynamics of bacterial populations. More generally, we develop and use these approaches in connection with experimental work in our lab for understanding and **controlling the dynamics of populations** in cases where **controllable system inputs inherently operate at the level of single cells** (e.g. optogenetics) but the output of interest is at the level of populations (e.g. bioproduction).

3.2 Optimal experimental design

One of the major problems in reverse engineering biochemical processes inside cells is that **cellular processes are high-dimensional and complex** with many unknown parameters while the **available data is low dimensional and corrupted by measurement errors**. Such problems can be alleviated by ensuring that the experimental plan is designed to yield data that provides as much information as possible about the unknown model parameters. We develop mathematical approaches and computational tools that can be used to calculate the expected amount of information that can be gained from a given experiment given a specification of either a stochastic model of the system (described above) or a deterministic model based on ordinary differential equations. These information calculation approaches are then coupled to optimization tools and used to plan **maximally informative experiments** in our applications.

3.3 Cybergenetics – real time control of biological processes

Cells have evolved uncountable numbers of feedback circuits to regulate their functionalities in the presence of changing environmental conditions. But can such feedback control also be externalized and placed under control of scientists? Early work on this topic suggested that **optogenetic systems, allowing for external regulation of gene expression**, have the potential to serve as an interface between cells and experimental platform that gives a computer the power to stir the functioning of cells via the application of light. We develop all the tools required to realize automated computer control of intracellular processes. On the experimental side, we develop yeast strains that are equipped with optogenetic promoters to drive various functionalities. On the mathematical side, we develop models and software to equip our experimental platforms with the appropriate programs to realize sucessful feedback control, both at the level of single cells (microscopy) and at the level of populations (bioreactors and plate reader).

3.4 Platforms for automated reactive experiments

The core scientific activity of the team is to connect mathematical methods with biological applications in our lab. The interface between the two sides, that is the experimental platforms, is therefore of crucial importance for the success of our activities. However, platforms that can be purchased by vendors are

typically delivered without the capacity to adapt the experimental plan in response to incoming measurements, a functionality that is crucially needed for deploying our computational methods (e.g. feedback control). Therefore, we develop novel experimental platforms and/or extend existing platforms with additional software and hardware that allows us to perform automated reactive experiments. Concretely, we develop a **microscopy platform and control software for yeast that uses a digital micromirror device to expose single cells to targeted light signals** that can be adjusted in real time in response to measurements taken from the cell. Furthermore, we develop a platform of 16 parallel **small scale automated bioreactors**, each equipped with controllable LEDs to allow for optogenetic gene expression and long-term reactive experiments in tightly controlled conditions. Automation of the platform is achieved via a low-cost open-source **pipetting robot that samples all reactors to a benchtop cytometer** in which single cell gene expression is measured in all sampled cells of all reactors. Finally, we develop **software to take full control of a commercial plate reader with liquid injection capabilities** (Tecan Spark). This platform allows us to use a Raspberry Pi to pilot 96 parallel reactive experiments where optical density is used as a readout of bacterial growth.

4 Application domains

4.1 Preamble

Since most of our research is at the interface of mathematics and biology, there often is no clear split between mathematical reaseach objectives and applications. For instance, feedback control of gene expression is simultaneously a mathematical and an applied problem.

4.2 Understanding resistance and tolerance to antibiotic treatments

The non-susceptibility of pathogenic bacteria to antibiotic treatments is a major health problem. Bacteria might escape treatments in two ways: being resistant or being tolerant. Whereas resistant bacteria can multiply in presence of antibiotics, tolerant bacteria can merely survive. Yet, tolerance is increasingly recognized as a major player in treatment failure. In particular, an increasing fraction of commensal and pathogenic *E coli* bacteria express extended-spectrum β-lactamases and/or carbapenemases. When individual bacteria die as a consequence of antibiotic treatments, these enzymes are released and hydrolyze the antibiotic molecules in the environment, conveying a transient protection to the remaining bacteria that lasts until the enzymes are degraded themselves. Understanding how this collective antibiotic tolerance (CAT) shapes population dynamics is difficult yet important for **optimally killing bacterial populations**: when a second antibiotic dose is applied directly after a first dose it will not be effective since the antibiotics will be degraded by the enzymes released from bacteria killed after the first dose; when the second dose is applied too late the surviving bacterial population will have regrown to a large size. Our plate reader platform allows us to apply complex temporal patterns of antibiotic treatments to bacteria over nearly two days. Parallezing such treatments in the 96 well plates allows us to generate rich data sets and to calibrate population dynamics models that can be used to optimize temporal treatment plans. One of the applied objectives of our team is to use these capacities to study a collection of fully-sequenced clinical isolates treated with a broad range of clinically important antibiotics and grown in various media. Ideally, this will lead to an approach that can be used to assay tolerance to antibiotics in hospitals instead of, or in addition to, standard antibiotic susceptibility tests, detecting resistance.

4.3 Optimization of protein production in yeast

Many proteins are of technological or therapeutical importance. The yeast *S. cerevisiae* is an interesting organism for protein bioproduction since it combines a relatively fast growth rate with good capacities to perform post-translational modifications needed for protein maturation and full functionality. However, imposing a strong demand on protein production to the cell places a significant burden on its physiology, either at the protein production level or at the maturation and secretion levels. Using systems and synthetic biology approaches, we aim at **better understanding the origins of the production bottlenecks** and then using modeling and control approaches, we aim at finding **optimal control solutions** for bioproduction. Three different strategies are envisioned. In the first approach, bioproduction stress

sensors are used to observe in real time the physiological state of the cell, and the demand is externally tuned based on the stress level of the cell population. In the second approach, the stress sensor is used to tune the response capacities of the cell to the external demand, thus creating an internal feedback loop. In the third approach, we control the fraction of the producing cells by engineering an artificial differentiation system that implements the partial differentiation of grower cells into producer cells. The optimization problem is then to find the optimum ratio based on the external environment of the cells.

5 Social and environmental responsibility

5.1 Footprint of research activities

A significant part of our daily research activities involves molecular biology work and consumes plasticware and various chemicals. We also work on lab automation and develop experimental platform to paralellize experiments. However, we work with small biorectors (15 to 50mL), so volumes of cell cultures remain very modest.

We also occasionally use a computer cluster, notably for optimization, but the jobs remain relatively modest on a yearly basis.

Allyson Holmes and Henri Galez are part of a committee working on the quantification of the carbon footprint of the computational works done in the Computational Biology Department of Institut Pasteur.

5.2 Impact of research results

Regarding biological developments, we have two main research directions.

The first one deals with the optimization of bioproduction. Bioproduction is a domain of strategic importance. The field is higly technological and rapidly growing at the global scale. The market for biopharmaceuticals alone, that notably include vaccines and monoclonal antibodies, is estimated to $400B to $500B. France imports > 70% of its vaccines and > 95% of its monoclonal antibodies and lacks sovereignty. Therefore this field has a strong social, medical and economical importance.

The second research direction deals with antibiotic stewardship. The spread of antimicrobial resistance is a both a health and an ecological problem of global impact. Antibiotic stewardship aims at using these drugs in more appropriate ways. To do so, one has to better understand and quantify bacterial response to antibiotic treatments.

Therefore our two main research directions are both tightly connected with important health and social issues.

6 Highlights of the year

6.1 Optimizing protein production using cybergenetics approaches - Publication in Nature Communications

Using our automated experimental platform and an optogenetic induction system for gene expression, we caracterized in depth the impact on cell physiology of the demand for protein secretion in yeast. We found that for hard-to-secrete proteins strong demands led to excessive stresses and to a "burn-out" state for the cell.

This knowledge was then used to propose an external control strategy that should maximize protein secretion by keeping the stress at a maximal acceptable level. This strategy was experimentally validated and led to an improved production in comparison to the full induction strategy.

Our proposed regulation strategy is based on the simple measurement of mean UPR stress levels in cells and is in principle compatible with and complementary to the extensive chassis-engineering optimization strategies found in industrial applications. Moreover, this work also demonstrates that cybergenetic approaches can effectively be used to optimize cellular processes of interests.

7 New results

7.1 Maximizing protein production by keeping cells at optimal secretory stress levels using real-time control approaches

Participants: Sebastian Sosa-Carrillo, Henri Galez, Sara Napolitano, François Bertaux, Gregory Batt.

The production of recombinant proteins is a problem of major industrial and pharmaceutical importance. Secretion of the protein by the host cell considerably simplifies downstream purification processes. However, it is also the limiting production step for many hard-to-secrete proteins. Current solutions involve extensive chassis engineering to favor trafficking and limit protein degradation triggered by excessive secretion-associated stress. In this contribution, we proposed instead a regulation-based strategy in which induction is dynamically adjusted based on the current stress level of the cells. Using a small collection of hard-to-secrete proteins and a bioreactor-based platform with automated cytometry measurements, we demonstrated that the regulation sweet spot is indicated by the appearance of a bimodal distribution of internal protein and of secretory stress levels, when a fraction of the cell population accumulates high amounts of proteins, decreases growth, and faces significant stress, that is, experiences a secretion burn-out. In these cells, adaptations capabilities are overwhelmed by a too strong production. With these notions, we defined an optimal stress level based on physiological readouts. Then, using real-time control, we demonstrated that a strategy that keeps the stress at optimal levels increases production of a single-chain antibody by 70

7.2 Bayesian filtering for model predictive control of stochastic gene expression in single cells

Participants: Zachary Fox, Gregory Batt, Jakob Ruess.

In this contribution, we described a method for controlling the production of protein in individual cells using stochastic models of gene expression. By combining modern microscopy platforms with optogenetic gene expression, experimentalists are able to accurately apply light to individual cells, which can induce protein production. Here we used a finite state projection based stochastic model of gene expression, along with Bayesian state estimation to control protein copy numbers within individual cells. We compared this method to previous methods that use population based approaches. We also demonstrated the ability of this control strategy to ameliorate discrepancies between the predictions of a deterministic model and stochastic switching system.

7.3 Optimal control of bioproduction in the presence of population heterogeneity

Participants: Davin Lunz, Jakob Ruess, J Frederic Bonnans.

Cell-to-cell variability, born of stochastic chemical kinetics, persists even in large isogenic populations. In the study of single-cell dynamics this is typically accounted for. However, on the population level this source of heterogeneity is often sidelined to avoid the inevitable complexity it introduces. The homogeneous models used instead are more tractable but risk disagreeing with their heterogeneous counterparts and may thus lead to severely suboptimal control of bioproduction. In this contribution, we introduced a comprehensive mathematical framework for solving bioproduction optimal control problems in the presence of heterogeneity. We studied population-level models in which such heterogeneity is retained, and proposed order-reduction approximation techniques. The reduced-order models take forms typical of homogeneous bioproduction models, making them a useful benchmark by which to

study the impotance of heterogeneity. Moreover, the derivation from the heterogeneous setting shed light on parameter selection in ways a direct homogeneous outlook cannot, and revealed the source of approximation error. With view to optimally controlling bioproduction in microbial communities, we asked the question: when does optimising the reduced-order models produce strategies that work well in the presence of population heterogeneity? We showed that, in some cases, homogeneous approximations provide remarkably accurate surrogate models. Nevertheless, we also demonstrated that this is not uniformly true: overlooking the heterogeneity can lead to significantly suboptimal control strategies. In these cases, the heterogeneous tools and perspective are crucial to optimise bioproduction.

8 Partnerships and cooperations

8.1 International research visitors

8.1.1 Visits of international scientists

Participants: Lorenzo Pasotti, Angelica Frusteri.

Lorenzo Pasotti, assistant professor at the Department of Electrical, Computer and Biomedical Engineering and at the Centre for Health Technologies of the University of Pavia (Italy), and Angelica Frusteri, a postdoctoral scientist in his group, have been invited for three months in the InBio team.

8.2 National initiatives

- **Institut de Convergence Inception** (2016-2025) on the “Emergence of Diseases in Populations and in Individuals”, coordinated by T. Bourgeron (Institut Pasteur). Partner institutes include Institut Pasteur, Paris Sciences et Lettres, Université de Paris, AP-HP, and research teams from CEA, CNRS, INSERM and INRA.

  The Inception’s goal is to develop a core structure to mobilize data resources, numerical sciences, and fundamental experimental biology in a range of health issues. It uses integrative biology, social science and data science to understand the emergence of diseases in populations and in individuals. Inception provides funding for the PhD work of Viktoriia Gross.

- **PPR Antibiorésistance Anoruti** (2021-2025) on the “Analysis of non-response to antibiotics in vivo: application to Escherichia coli urinary tract infections”, coordinated by I. El Meouche (Inserm).

  The objective of Anoruti is to identify the different factors involved in the fact that some bacteria sensitive to an antibiotic in vitro do not respond to treatment in vivo.

- **PPR Antibiorésistance Seq2Diag** (2021-2026) on “Whole genome sequencing and artificial intelligence to characterize and diagnose antibiotic resistance and capacity to escape treatment”, coordinated by P. Glaser (Institut Pasteur).

  Genomic sequencing has revolutionized microbiological surveillance and molecular epidemiology. The objective of the Seq2Diag project is to provide a proof of concept for its use in hospital and veterinary laboratories as a diagnostic tool for in silico antibiotic sensitivity testing.

- **ANR JCJC SmartSec** (2022-2024) on “Matching maximal host capacities: stress-informed, self-tuning bioproduction circuits”, coordinated by F. Bertaux (Lesaffre) with G. Batt (Inria and Institut Pasteur).

  Bioproduction requires diverting resources normally used by host cells for growth and self-replication towards the production of desired molecules. Achieving maximal resource diversion without compromising the essential functions of the host is of critical importance, but is particularly challenging. To tackle this challenge, SmartSec aims at designing host-aware circuits, with application to the production of secreted proteins.

9 Dissemination

9.1 Promoting scientific activities

9.1.1 Scientific events: selection

Gregory Batt is a member of the International Programme Committee of the 10th International Conference on the Foundations of Engineering in Biology (FOSBE 2024).

9.1.2 Journal

Gregory Batt has been a reviewer for Nature Communications.

Esteban Lebrun has been a reviewer for Lab on a Chip.

Lorenzo Pasotti has been a reviewer for Nature Communications, Nature Chemical Biology, and Microbial Cell Factories.

9.1.3 Invited talks

Gregory Batt gave invited presentations at the International Symposium on Synthetic Biology, at Technishe University Darmstadt (Germany), and at the Symposium on Synthetic and Systems Biology (BioSynSys), at INSA Toulouse (France). He also gave invited seminars at the Microbial Processes and Interactions (MiPI) lab, at Université de Liège - Gembloux Agro-Bio Tech (Belgium), at the Laboratoire Jean Perrin, Sorbonne Université (Paris), and at the Advanced Light Microscopy seminar series, Institut Pasteur (Paris).

Henri Galez gave an invited seminar at Imperial College London (UK).

Sara Napolitano gave an invited talk at the Learning Planet Institute (Paris).

Lorenzo Pasotti gave invited seminar at the Department of Bioengineering, University of Maryland (USA), and at the Department of Biology and Biotechnology, University of Pavia (Italy).

9.1.4 Scientific expertise

Gregory Batt has been an expert for grant evaluation at BPI France.

9.1.5 Research administration

Gregory Batt has been the deputy-diretor and then director of the Computational Biology department at Institut Pasteur. He is also a member of the Bureau du Comité des Equipes Projets at Inria Paris.

9.2 Teaching - Supervision - Juries

Esteban Lebrun gave a course on "Introduction to Synthetic Biology" at the Microbiology Master of Université de Caen Normandie.

Lorenzo Pasotti gave courses on "Bioinformatics and Synthetic Biology" at the Bioengineering Master of University of Pavia and on "Bioinformatics" at the Biotechnology Master of University of Pavia.

9.2.1 Supervision

Gregory Batt is (co-)supervising PhD students:

- PhD in progress: Viktoriia Gross, “An integrative approach to characterize the two sides of enzyme mediated antibiotic escape: resistance and tolerance". Started in Oct. 2020. Supervision by Imane El Meouche, Erick Denamur and Gregory Batt.
- PhD in progress: Alicia da Silva, “A screening and learning approach for protein secretion in yeast". Started in Oct. 2023. Supervision by Senta Blanquet, Etienne Jourdier and Gregory Batt.

Sara Napolitano and Gregory Batt are co-supervising a PhD student:

- PhD in progress: Henri Galez, “Engineering an autocrine-like system for screening libraries of protein secreting strains in yeast". Started in Sept. 2022. Supervision by Sara Napolitano and Gregory Batt.

Henri Galez, Sara Napolitano, and Gregory Batt supervised the internship of Hong Duong Ngo (M2 student, Paris Cité University). Viktoriia Gross and Gregory Batt supervised the internship of Konstantin Achkasov (M1 student, Sorbonne University).

9.2.2 Juries

Gregory Batt has been member of the HDR committee of Manish Kushwaha (Paris-Saclay University) and of the PhD committees of Eve Tasiudi (ETH Zurich), Romain Aubry (Poitiers University), and Esteban Lebrun (Paris-Saclay University).

He has also been member of the thesis advisory committees of Lena Le Quellec (Paris-Saclay University/Institut Pasteur) and Madison Lenormand (Sorbonne University/Institut Pasteur).

10 Scientific production

10.1 Major publications

[1] S. Sosa-Carrillo, H. Galez, S. Napolitano, F. Bertaux and G. Batt. ‘Maximizing protein production by keeping cells at optimal secretory stress levels using real-time control approaches’. In: *Nature Communications* 14.1 (17th May 2023), p. 3028. DOI: 10.1038/s41467-023-38807-9. URL: https://inria.hal.science/hal-03940186.

10.2 Publications of the year

International journals

[2] Z. Fox, G. Batt and J. Ruess. ‘Bayesian filtering for model predictive control of stochastic gene expression in single cells’. In: *Physical Biology* (21st June 2023). DOI: 10.1088/1478-3975/ace094. URL: https://inria.hal.science/hal-04148503.

[3] D. Lunz, J. F. Bonnans and J. Ruess. ‘Optimal control of bioproduction in the presence of population heterogeneity’. In: *Journal of Mathematical Biology* 86.3 (6th Feb. 2023), p. 43. DOI: 10.1007/s00285-023-01876-x. URL: https://inria.hal.science/hal-03445175.

[4] D. Lunz and J. F. Bonnans. ‘Modelling and optimal control of a two-species bioproducing microbial consortium’. In: *SIAM Journal on Applied Mathematics* 83.1 (13th Feb. 2023), pp. 144–171. DOI: 10.1137/22M1476113. URL: https://inria.hal.science/hal-03479385.

[5] S. Sosa-Carrillo, H. Galez, S. Napolitano, F. Bertaux and G. Batt. ‘Maximizing protein production by keeping cells at optimal secretory stress levels using real-time control approaches’. In: *Nature Communications* 14.1 (17th May 2023), p. 3028. DOI: 10.1038/s41467-023-38807-9. URL: https://inria.hal.science/hal-03940186.