Modelling and simulating biological processes in the genomic era: an account of a multidisciplinary thematic school held in Evry (France) in April 2004

François Képès,[1,2] Patrick Amar,[1,3] Georgia Barlovatz,[4] Gilles Bernot,[1,5] Christine Froidevaux,[3] Jean-Louis Giavitto,[5] Janine Guespin,[6] Franck Molina,[7] Vic Norris,[1,8] Vincent Schächter [9] & Philippe Tracqui [10]

[1] *Epigenomics Project / genopole®, Evry, France.*
[2] *ATelier de Génomique Cognitive, CNRS UMR8071 / genopole®, Evry, France; & Centre de Recherche en Épistémologie Appliquée, École Polytechnique, Paris, France; & Dynamique de la Compartimentation Cellulaire, Institut des Sciences du Végétal, CNRS UPR2355, Gif, France.*
[3] *Laboratoire de Recherche en Informatique, CNRS UMR 8623, Université de Paris-Sud, Centre d'Orsay, France.*
[4] *Dynamic / INSERM, Créteil, France.*
[5] *Laboratoire de Méthodes Informatiques, CNRS UMR8042 / genopole®, 523 Terrasses de l'Agora 91000 Evry, France.*
[6] *Laboratoire de Microbiologie du Froid, Université de Rouen, Faculté des Sciences & Techniques, Université de Rouen, 76821 Mont-Saint-Aignan, France.*
[7] *Centre de Pharmacologie et Biotechnologie pour la Santé, CNRS UMR5160 / Faculté de Pharmacie, 34093 Montpellier CEDEX 5, France.*
[8] *Laboratoire des Processus Intégratifs Cellulaires, UPRESA CNRS 6037, Faculté des Sciences & Techniques, Université de Rouen, 76821 Mont-Saint-Aignan, France.*
[9] *Genoscope, CNS / genopole®, Evry, France.*
[10] *TIMC-IMAG, CNRS, Institut A. Bonniot, La Tronche, France.*

INTRODUCTION

The global project of high-throughput biology may be summarized as follows. After genome sequencing comes the annotation using the techniques of 'classical' bioinformatics. It then becomes important to interpret the annotations, to understand the interactions between biological functions, and to predict the outcome of perturbations, while incorporating the results from post-genomics studies. This stage can only be achieved through modeling and simulation of biological processes and should be tightly coupled to bench experimentation. The recent shift in biology, with data accumulating much more rapidly than before at the molecular level, necessarily makes this scientific development a long-term trend.

Two dozen researchers with various scientific backgrounds started in January 2001 to face these challenges in a stimulating year-round workshop that was initiated and supported by genopole® in Evry, France. Some of these scientists were initially more familiar with the field of modelling/simulation, while others were involved in various aspects of (post-) genomics. After 14 months of work, they held a small multidisciplinary seminar in Autrans which brought together 60 participants. Since then, the number of researchers directly involved in workgroups has approximately doubled, and the small seminar has become an oversubscribed thematic school. The estimated size of this rapidly-growing community presently amounts to about 250 people in France.

This year's five-day thematic school comprised lecture sessions, evening discussions around images, and afternoon workshops on specific topics. The five lecture sessions dealt with the epistemology of modelling, simulation and emergent properties, formal models, macromolecular interaction networks, organization and morphogenesis. Each lecture started with a didactic and wide introduction and ended with a research-oriented part. The evening sessions were opportunities to interactively discuss a topic around images from real-world or virtual biology. One afternoon workshop was focused on function-dependent structures (summarized below) and another one on the integration and visualization of genomic data in the framework of human-computer interaction. Another workshop that spread over two afternoons compared various modeling paradigms applied to the same biological object. Poster sessions allowed ample discussion and some posters were highlighted by short talks (summarized below).

The following is an attempt to capture some of the evanescent spirit of this school, and we apologize for any involuntary mis- or under-representation of particular aspects.

Correspondence should be addressed to F. Képès, Epigenomics Project/genopole®, Evry, France.
E-mail: Francois.Kepes@genopole.cnrs.fr

LECTURE SESSIONS

Epistemology of modelling: interdiscipline or indiscipline?

Pierre Sonigo (Institut Cochin, Paris) tackled the central tenet of neo-Darwinism that DNA in the form of a genetic program fully commands cells and organisms

Journal of Biological Physics and Chemistry **4** (2004) 131–136
Received 03 June 2004

and that natural selection acts directly on this program. He used the analogy of the Robot, where there is a master program, and the Forest, where there is not, to highlight two different interpretations of biological systems. He pointed out serious problems with the Robot view. For example, the assumption that selection could act just at the level of the gene is hard to reconcile with the gulf of complexity between the gene and selectable phenotypes. In a forest, selection operates at every level and he argued in favour of the Forest view where molecules are in competition with other molecules, cells with other cells, organisms with other organisms etc. The global structure of the forest emerges from the interactions between individuals obeying local rules, and in the Forest view of the cell emerging from interactions between individual constituents, DNA loses its quasi-divine status. The Forest view has powerful implications for medicine and Sonigo illustrated these with reference to autoimmunity, cancer and differentiation and gave the example of gene therapy, where the introduction of a gene into a cell can be likened to the introduction of a new species into a particular ecosystem.

A reductionist separates his world into system and environment, and in his talk **Marc van Regenmortel** (CNRS/Université de Strasbourg) returned to the theme of the nature of an individual object. Like Sonigo, van Regenmortel insisted on the importance of context. Genes only provide a function in a context. We cannot predict function *de novo* from sequence. We are accustomed to a causality of the form A causes B but the behaviour of a biological system must be explained not by a single cause but by hundreds of causes. Hence the paradigm of cause leading to effect is not very useful for understanding cells. Instead, the function of a constituent with respect to its context becomes important where function is about contributing to the survival or reproduction of a biological system. The same protein structure can be associated with many different functions, so that there is no such thing like a "structure-function relationship". Such a statement from a person who has a strong personal experience in structural biology was certainly striking. Van Regenmortel gave the example of a "specific" binding site on an immunoglobulin in which a binding site is a relational entity and can be considered as an emergent feature.

These two talks were followed by a debate on the above topics, between both speakers and the participants.

Simulation and emergent properties in biology (from microscopic to macroscopic)

Three ways of approaching and studying emergent processes have been confronted during this session.

Jean-Christophe Leloup, in work conducted with **Albert Goldbeter** (Université libre de Bruxelles), showed how the oscillatory circadian rhythm in flies and mammals emerges from a somewhat abstract description of genetic interactions. In *Drosophila*, two genes (PER and TIM) are involved in a negative feedback circuit : the encoded proteins form a complex that represses their transcription. This circuit underlies the circadian clock oscillator and is also found in mammals, where a second pair of genes with a similar behaviour was shown to constitute a second oscillator, intertwined with the first one. These autonomous oscillations are nevertheless coupled to the light cycle of 24 hours. Modelling makes use of a differential approach, where variables correspond to genes and their products, phosphorylated or not. Many more equations are needed to describe the mammalian compared with the fly clock, but the problems are similar. Most of the parameters have not been determined, and must be chosen so that the model fits biological data or diverse behavioural patterns encountered in wild type (the role of light), mutant flies or human pathologies of sleep. In the case of the more complicated mammalian circadian clock, where two negative feedback loops are intertwined, the theoretical model allowed to predict the rôle of each oscillator, and thus to address the dynamical bases of the emergence of oscillation as well as of the physiological disorders related to a perturbation of the clock.

Luc Steels (Sony Computer Science Laboratory, Paris & Université Libre de Bruxelles) addressed the question of self-organization and language evolution. The main hypothesis is that emergent processes are at the root of language evolution. This is in contrast with Chomski's theory of innate structures. Linguistic arguments in favour of the emergent nature of language evolution are rooted in the study of this evolution itself, through language comparison and assessment of actual speech situations, where communication relies on a common (implicit) knowledge of part of the situation. The relationship between words and meaning is therefore very indirect, including several layers of categorization. The implicit plays a major rôle, each language differing from the others as to what is implicit (hence explicit). Starting from linguistic processes that give a hint of emergence, Steels uses multi-agent systems and learning robots endowed with a formal communication protocol to study the emergence of linguistic behaviours similar to real ones. In a community of 2 to 400 agents, what instruction must be given to each agent to obtain an evolution of the communication and a consistency of the language

between all the agents? This has been tested for the emergence of sounds (one sound can spread into the community) or for a real communication where a common ground of knowledge appears during the experiment. One of the open questions is: what is general enough in these models that might also apply to genetics?

Eric Bonabeau (Icosystem Corp., Cambridge, Mass.) discussed the modeling of self-organized systems from the bottom up. The emergent process under scrutiny was organization in social insects, where space must be introduced. Self-organizing processes are complex with many individuals involved, and with a global behaviour that is generally non-predictable, and may even be anti-intuitive. Thus, the only way to model it is to use agent-based modelling, starting from the "bottom" interactions between agents, to obtain, and hopefully understand, the global, emergent, "up" behaviour. 2-D simulations of ant behaviour can be achieved with the very simple rule according to which each ant is attracted by the pheromone released by other ants. Some elementary physical rules, such as the drawback of protruding legs of dead bodies, may be necessary for proper cemetery formations. In every instance, the presence of positive and negative feedback circuits is crucial. Simple 3-D agent-based modelling illuminated how coordination may emerge in the collective construction of wasp nests. In sum, the reverse approach, that consists of finding the elementary rules that correspond to a given structure, is quite appealing, but most difficult. This amounts to searching a very large space of rules, with choice criteria that are difficult to formalize. Genetic algorithms may be used, but the choice is most often intuitive and subjective.

Formal models for biological modelling

Gordon Plotkin (University of Edinburgh) gave the first talk on "biochemical CCS", a formalism derived from process algebras for the modelling of biochemical networks of interactions (metabolic pathway, genetic regulatory network, signalling, etc). The fundamental idea is to see biological processes as calculations and thus, as that view appeared fertile in data processing, to study their naming, their composition and their modularity from an algebraic point of view. In this approach, the elementary processes correspond to transformations (e.g. chemical reactions), various binding processes (formation and dissociation of complexes) and translocations (diffusion). These elementary processes can be formalized in several ways, for example by a transition in a Petri net or by a differential equation. The latter approach is a traditional one, but Plotkin showed with several examples how one can model these elementary processes by a Petri net (in this case, one cannot express the kinetics of the chemical reactions).

François Fages (INRIA, Rocquencourt) presented the BioCham project. BioCham (Biochemical abstract machine) is a specification language using rules to give a precise semantics to the biological network. BioCham appears as an environment integrating a rule-based language dedicated to the specification of the molecules and their interactions, an interrogation language based on CTL (a temporal logic making it possible to simply express a large variety of biological questions about the behaviour of the system) and an interface to NuSMV, a tool for model-checking making it possible to answer CTL requests. The syntax of BioCham makes it possible to define very simply molecules, proteins, their complexation and to annotate them with their functional field. This syntax also corresponds to algebraic operators. The definition of various reactions, transformations, associations/dissociations etc., are represented by rules inspired by transition systems (an approach largely used to formalize the activity of a set of parallel processes in concurrent and distributed systems). The proposed syntax was validated by the development of large-scale examples such as the MAPK signalling pathway and the cell cycle.

Vincent Danos (PPS, CNRS/Université Paris 7) gave a talk on biological combinatorics. After reminding us that a living being requires sugar to process information, and information to process sugar, Danos went on to ask three "big" questions. Can we understand this organization? Which problems is this cellular organization solving? What are the underlying computational principles? Reverse and forward engineering of biological systems need a syntax. Danos discussed syntaxes addressing specific features of cellular computing: "Binding", a calculus for protein-protein and protein-DNA interactions, and "Enfolding", a calculus for membrane fusion and fission, exocytosis and endocytosis. He illustrated "Binding" with the formalization of the lactose operon, and "Enfolding" with membrane traffic. The syntax was validated on the relatively large case of viral invasion.

Christoph Teuscher (EPFL, Lausanne & UCSD, La Jolla) presented examples of computer architectures inspired by biological principles. These new hardware architectures are justified by the needs of new applications, by technological progress, and by new approaches to computation that are largely inspired by biological mechanisms. The first architecture exhibited properties of self-repair and robustness. It comprised

three levels: the molecule (an autonomous module of computation), the cell (a set of interacting molecules) and the organ (a set of interacting cells). The program corresponded to an artificial genome stored identically into each cell. The position of the cell in the machine determines the part of the program (genes) which will be carried out. A prototype was built: BioWall, a machine of 2400 molecules organized into a grid-like cellular automaton. The other types of architecture that were presented were POEtic machines, *i.e.* inspired by Phylogenesis, Ontogenesis and Epigenesis mechanisms. The objective was to develop a hardware architecture being able to evolve, grow, self-repair, adapt and replicate itself. Finally, the Amorphous Blending Membrane project is an attempt at implementing the notion of blending, a fundamental operation for knowledge representation.

Macromolecular networks and statistical inference

Xavier Gidrol (SGF/CEA/Evry) studies genetic networks, with the goal of understanding their dynamics in order to predict stable phenotypes (attractors), and predict in which phenotype the cell will end up. In particular, Gidrol discussed the case of protein Id2, whose overexpression induces cell proliferation. The goal here is to reconstruct the Id2 network in order to acquire knowledge about the gene regulatory network in the neighbourhood of Id2. After reminding us of P. Sorger's aphorism, "Good experimentation is essential to realize the systems biology vision", he proposed several paths to get closer to an exhaustive, high-resolution view of the system. Some of these paths aimed at enlarging the Id2 transcriptional network, both upstream and downstream (using cells-on-chips for instance). Some others addressed the validation issue. This talk was followed by a particularly long and exciting general discussion.

The talk of **Florence d'Alché-Buc** (Programme d'Épigénomique, Evry & Université Paris 6) was aimed at showing the benefits of a machine learning approach for modelling gene regulatory networks. The development of microarray technology makes it possible to compare simultaneously the expression of thousands of genes of a given organism (or tissue) in two conditions and offers new valuable pieces of information to construct gene regulatory networks. D'Alché-Buc introduced the task of constructing this network as a search of parameters for a model of gene interactions to be learned from experimental data. Then she provided a survey of the problems and concepts of statistical learning. The central point was the formulation of the objective of a learning task as an optimization problem. Indeed, once the hypotheses space has been built, learning is tantamount to choosing the best hypothesis in the space. In the second part of her talk, D'Alché-Buc introduced graphical models for reverse modelling. First, she presented the work by Segal et al. who introduced module networks, a probabilistic method for identifying regulatory modules from microarray data. Then she showed how dynamic Bayesian networks used by Perrin et al. for gene networks inference were especially well suited to tackle the stochastic nature of gene regulation and gene expression measurement. D'Alché-Buc concluded her talk by listing a number of perspectives, including the elaboration of a set of benchmark problems.

Wolfgang Banzhaf (Memorial University of Newfoundland, Canada) reminded us of the ubiquity of networks. To understand them, we should investigate their similarities and differences, the mechanisms for their generation and the interrelations between their structure and dynamics. A few steps have been made thanks to global characterization (scale free/small world topology) or local characterization (network motifs). Networks remain our best bet to catch emergent phenomena. Banzhaf then presented a method to compare natural regulatory networks from *E. coli* and *S. cerevisiae*, and artificial ones. The artificial regulatory network comprises 32 genes and is generated by a neutralist duplication and divergence process. Banzhaf extracted basic structural elements of the complex networks. Very few motifs occurred with significantly higher probability than in random networks. There was a clear relation between the natural distribution of motifs and the distribution in artificial networks generated by a duplication and divergence process. No evolutionary selection pressure had been applied to the artificial system, though. Thus, it can be stated that the distribution outcome is more a reflection of the mechanism of its generation than a result of evolutionary pressures.

Alessandro Vespignani (LPT, CNRS/Université Paris-Sud, Centre d'Orsay) introduced networks as a system that allows its abstract/mathematical representation as a graph. He argued on the ubiquity and variety of networks in physics, cybernetics, sociology, biology, etc. These networks are interconnected to form complex networks. They could form layers or have interdependencies. Vespignani provided an introduction to the general properties and possible generative mechanisms of such complex networks. By contrast to a complex system, a complicated system was described as a set of many elements assembled following a predefined blueprint imposed from outside. Consequently, the

system shows expected properties and performs predefined tasks. On the other hand, complex systems composed of many interacting units show dynamical evolution and the capacity for self-organization. The outcome is a non-trivial architecture, unexpected emergent properties and cooperative phenomena. This was illustrated with protein-protein interaction networks (PIN) obtained from the two-hybrid technique. Finally, Vespignani proposed to use the global information provided by the PIN to functionally annotate proteins. This approach was based on maximizing unclassified interacting proteins with shared functionality and minimizing classified interacting proteins with different functionalities. It may serve as a general method to obtain statistical prediction of protein function from their interactions with reliability around 80% for highly interacting proteins.

Organization and morphogenesis

James Tabony (RDC/DSV/CEA, Grenoble) addressed the physicochemical processes underlying *in vitro* microtubule self-organization. His experimental work was interpreted in the framework of dissipative systems, for which theoreticians such as I. Prigogine have predicted that macroscopic self-organization can arise from a non-linear coupling of reactive processes with molecular diffusion. Tabony found that the *in vitro* formation of microtubules from tubulin shows this type of behaviour. These preparations spontaneously self-organize by way of reaction and diffusion, and the morphology that develops depends upon the presence of a weak external factor, such as gravity or a magnetic field, at a critical bifurcation time early in the process. Thus, the presence of an external symmetry-breaking factor, such as gravity, can determine the morphology that subsequently develops. Once assembled from tubulin, microtubules grow and shrink from opposite ends. The shrinking end of a microtubule leaves behind itself a chemical trail of high tubulin concentration. Neighbouring microtubules preferentially grow into these regions, thus progressively leading to self-organization in a manner that shows analogies with the way that ants self-organize. Numerical simulations of the reaction-diffusion process based on the chemical dynamics of a population of microtubules successfully predict the main features of the experimental behaviour. Evidence is presented that processes of this type occur *in vivo* during embryogenesis of *Drosophila* egg.

In his talk, **Vincent Hakim** (LPS, ENS, Paris) examined the synchronization properties of neuron networks. He investigated how the instantaneous firing rate of a neuron can be modulated by a noisy input. The results show that the firing-rate modulation is shaped by the subthreshold resonance. For weak noise, the firing-rate modulation shows a minimum near the preferred subthreshold frequency. For higher noise, such as that prevailing *in vivo*, the firing-rate modulation peaks near the preferred subthreshold frequency. Using numerical simulations of conductance-based neurons and analytical calculations of one-variable nonlinear integrate-and-fire neurons, the dependence of this synchronization on the modulated noise has been analyzed. These results were discussed in connection with intrinsic neuron properties, including the characteristics of fast sodium channel that could determine the speed with which neurons respond to noisy inputs.

WORKSHOPS

One of the afternoon sessions was convened by **Michel Thellier** (Académie des Sciences & Université de Rouen) and **Patrick Amar** (Programme d'Épigénomique, Evry & Université Paris-Sud). Thellier introduced the notion of *function-dependent structures (FDS)*: dynamical structures that are created and maintained due to their functioning. To illustrate this concept, he used the example of the self-assembly of successive enzymes in a metabolic pathway. Using partial differential equations, he demonstrated that when the affinities between the successive enzymes are large enough, assemblies appear. Under specific conditions, the system can exhibit some regulatory behaviours such as sigmoidicity. He showed another kind of FDS which inhibits the overall process when the last product is not released by sequestering the enzymes. Amar presented a simulation program, *Hsim*, designed to study the dynamics, and in particular the assembly and disassembly of large numbers of molecules in a virtual cell. He described the program as a stochastic automaton coupled to a modelling language. The language allows the definition of molecular types, the description of interaction rules between pairs of neighboring molecules of given types, and the description of an initial state of the system. The simulator was demonstrated on a metabolic pathway with a chain of five enzymes that progressively self-assemble and efficiently transform the initial substrate to the final product, and then dissociate when all the initial substrate has been transformed. He went on to demonstrate the growth of actin filaments in the virtual cell, where filaments tend to align along the cell axis, an emergent feature.

SHORT TALKS AND POSTERS

Some posters were highlighted by short talks.

S. Randall Thomas (CHU Necker, Paris) illustrated the application and the usefulness of mathematical modelling for physiologists. Actually, the complexity of the interactions among coupled flows at the level of kidney organisation and the very inaccessibility of the inner structure, prevent *in vivo* intervention and requires theoretical analysis to quantitatively formulate working hypotheses and predict new experiments. He described tools accessible on the web that provide a hierarchical collection of models at different scales and a quantitative database, « QKDB ». The database is an open source built for and with the participation of the kidney modelling community. These tools are being developed using generic approaches, with a view toward easy adaptation to other fields.

Jacques-Deric Rouault (NAMC, Université Paris-Sud, Centre d'Orsay) showed how DS^2 models (Dynamical Systems evolving in a Dynamical Structure) can be used to predict patterns appearing on *Drosophila*. He demonstrated his approach with a model using a few rectangular cells and a cell multiplication procedure in 2D space.

Pierre Mazière (CNRS/Faculté de Pharmacie, Montpellier) described a new integrative language, called BioΨ, developed to describe biological functions with respect to five parameters: schedule, specification, localization, biochemical state and kinetics. Four scales of observation were chosen, functional motifs, functional domains, molecular entities and functional modules. The first level is built on about a hundred Basic Elements of Action whose combinations are sufficient to account for the *ca* 3000 Enzyme Code descriptors. The application of the method was illustrated by the description of the insulin receptor system.

Cecilia Garmendia-Torres (IGM, Université Paris-Sud, Centre d'Orsay) showed how Msn2, a transcriptional activator involved in stress response in *S. cerevisiae*, shuttles periodically between the cytoplasm and the nucleus. She presented a model in which nuclear Msn2, after a delay, triggers the process which leads to its exit from the nucleus. She has determined that the part of Msn2 which confers the periodic migration contains a nuclear exit signal and a nuclear localization signal. The latter, once grafted onto a reporter protein bearing a stress-independent nuclear exit signal, is sufficient to trigger its oscillations.

Paul François (LPS, ENS, Paris) is interested in modular regulatory networks. He described an evolutionary *in silico* procedure that creates small protein-protein and protein-DNA networks performing basic tasks such as toggle switches or oscillators. Selection made use of genetic algorithms. One of the interesting outcomes was the observation that protein-protein interactions often dominated protein-DNA interactions in the solutions achieved by artificial evolution.

CONCLUSION

Among the messages taken away, a couple were particularly strong. Firstly, post-genomic biology will be dominated by multi-disciplinary teams formed both from specialists and from a new breed of interdisciplinary students. Breeding should be facilitated by schools such as the present one. Secondly, the gap between biological and physical approaches to complex systems is being bridged. New concepts are being generated and we are facing exciting research challenges. Thirdly, the dialogue between simulation and bench experimentation should be strongly emphasized in the near future. More than ever before, the goal is now to foster collaborative interactions towards building together a better understanding of life.

FURTHER READING

An Evry seminar book was edited by Patrick Amar, Jean-Paul Comet, François Képès and Vic Norris. It contains the meeting abstracts, papers and courses that relate to the above topics, and detailed accounts of the thematic school. Just like the Autrans and Dieppe proceedings books, it is available from Dr Hélène Pollard, Directrice Genopole-Recherche (Helene.Pollard@genopole.com).

ACKNOWLEDGMENTS

The authors thank Hélène Pollard, Catherine Meignen and Christelle Maimon for their constant and friendly support. FK is grateful to Jeremy Ramsden for inviting the present report. We wish to acknowledge the sponsoring of genopole®, CNRS & Société Francophone de Biologie Théorique.