Modelling and simulation of biological processes in the context of genomics

Vic Norris[1], Patrick Amar[3], Gilles Bernot[2], Jean-Louis Giavitto[2], Christophe Godin[4], Janine Guespin[5], Hélène Pollard[6], Philippe Tracqui[7] and François Képès[8]

[1]*Laboratoire des Processus Intégratifs Cellulaires, UMR CNRS 6037, Faculté des Sciences & Techniques, Université de Rouen, 76821 Mont-Saint-Aignan, France*
[2]*Laboratoire de Méthodes Informatiques, CNRS UMR 8042, Université d'Evry, 91025 Evry cedex, France*
[3]*Laboratoire de Recherche en Informatique, Université Paris-Sud, Orsay, France*
[4]*CIRAD, Laboratoire de modélisation des plantes, Montpellier, France*
[5]*Laboratoire de Microbiologie du Froid, Faculté des Sciences & Techniques, Université de Rouen, 76821 Mont-Saint-Aignan, France*
[6]*Genopole Recherche, 2 rue Gaston Crémieux, 91057 Evry cedex, France*
[7]*Laboratoire des Techniques Imagerie Modélisation Cognition, CNRS UMR 5525, Faculté de Médecine, 38706 La Tronche, France*
[8]*ATelier de Genomique Cognitive, CNRS UMR 8071/genopole®, 523 Terrasses de l'Agora, 91000 Evry, France*

The post-genomic era of biology is characterized by a deluge of molecular data about the cell. This "new" biology has its own vocabulary of genome, transcriptome, proteome and even metabolome, interactome and lipidome. The challenge is to make sense out of this information by coming up with a new integrated picture of the cell which takes into account that it is simultaneously an autocatalytic set, a tensegrity structure, a network or a set of networks with particular connectivities and feedback characteristics, a set of codes and decoding devices, a self-organizing system based on phase transitions, membrane physics, water structures and a host of physico-chemical properties of ions, polymers and other cellular constituents, a multi-level society adapted to the vagaries of an environment that can vary rapidly from heaven to hell … and so on. Taking on this challenge therefore requires the introduction of concepts unfamiliar to biologists and the development of new ones, the formulation of new hypotheses and their testing via simulations or wet experiments. Taking on this challenge therefore requires specialists from across the sciences to learn each other's language so as to collaborate effectively on defined projects.

Just such a multidisciplinary group of scientists has been meeting regularly at Genopole, a leading centre for genomics in France. This, the *epigenomics* group, is divided into four subgroups. The *consensus* subgroup has as one of its present objectives the interpretation of transcriptome data from micro-arrays obtained, for example, from the exposure of cells to low level pollutants or from cells as they progress through the cell cycle. The *membranes and intracellular structures* subgroup focuses on membrane deformations involved in the functioning of the Golgi body, in cell division or in attachment to surfaces, on the dynamics of the cytoskeleton, and on the dynamics of *hyperstructures* (which are extended, multi-molecule assemblies that serve a particular function). The *organization* subgroup has adopted a systems biology approach with the application and development of new programming languages to describe biological systems, which it has been applying to problems in the growth and differentiation of plants and in the structure and functioning of mitochondria. The *observability* subgroup addresses the question of which models are coherent and how can they best be tested by applying a formal system, originally used for testing computer languages, to an epigenetic model for mucus production by *Pseudomonas aeruginosa*, the bacterium involved in cystic fibrosis.

The work of these subgroups underpinned the first conference organized in Autrans in 2002. This work also underpinned the conference in Dieppe which, as reported here, brought together biologists, physical chemists, physicists, statisticians and computer scientists from both inside and outside the epigenomics group and gave leading specialists the opportunity to address an audience of doctoral and post-doctoral students as well as colleagues from other disciplines.

The consensus subgroup: interpretations of high throughput biological data

The introductory tutorial by François Képès aimed at demonstrating through two examples how a qualitative understanding of cellular dynamics could be used to ask fertile questions in biology and how bioinformatics could be particularly useful in providing the answers. The first example was the translational control of membrane protein assembly which has led to the "+70 pause" hypothesis. The second example was the regular positioning of coregulated genes along yeast chromosomes which has led to the "solenoidal DNA"

Correspondence should be addressed to F. Képès, ATGC, CNRS ESA 8071 / genopole®, 523 Terrasses de l'Agora, 91000 Evry, France. E-mail: Francois.Kepes@genopole.cnrs.fr

Journal of Biological Physics and Chemistry **3** (2003) 106–110
Received 21 July 2003

hypothesis of chromosome structure. This hypothesis-driven approach that feeds on large bodies of data was contrasted with the data-driven approach which is increasingly proposed despite its relative lack of success. The seminar by Eduardo P. C. Rocha illustrated nicely the hypothesis-driven approach. Starting from well-formulated hypotheses, he showed how the asymmetrical nature of the replication of the leading and lagging strands can induce biases in gene distribution and nucleotide composition that even affect the amino acid compositions of proteins. These constraints, which structure bacterial genomes, are opposed by the intense shuffling between the high number of repeats present in some genomes. Rocha then demonstrated how this trade-off between order and disorder has shaped modern-day genomes. Bernard Vandenbunder's tutorial covered a wide range of concepts and practices in the fields of transcriptomics and of networks of transcriptional interactions. He rightly insisted both on the necessity for multidisciplinary work in these fields and on the importance of a dialogue between experimentalists using local or global approaches. "Horizontal" exploration includes the inference of regulatory networks, modelling, analysis of subnetworks and bench experimentation. "Vertical" exploration requires a proper consideration of multilevel events of chromatin structure (which, as explained by Arndt Benecke below, has its own code), and of the stochastic character of gene expression. Vincent Schächter and Bertrand Séraphin discussed the intricacies of proteomics from two different perspectives. Schächter tackled the two-hybrid approach, which provides data on binary protein-protein interactions. He compared various technologies and results in detail. Séraphin described the tandem affinity purification (TAP) method, which uses tags to purify protein complexes and mass spectrometry to identify them, and compared it with the only other available technology. Both speakers pointed to the conceptual and technological pitfalls of each method, thus outlining very useful guidelines for the proteomics-oriented modeller.

Marie Dutreix and Christine Froidevaux have been using micro-array transcriptional analysis to detect the effects of low level exposure to radiation and pollutants on the yeast, *Saccharomyces cerevisiae*. They analysed their data using the RELIEF technique which is based on the level of activation of transcription within each class of instances versus variation between classes. They then compared the results of the RELIEF analysis with those from a standard analysis of variance and another standard deviation-based technique. Although these analyses did give some different results, they were in agreement in implicating genes associated with the functions of the mitochondrial membrane such as oxidative phosphorylation and ATP synthesis.

The membranes and intracellular structures subgroup: membranes and hyperstructures

Chaouqui Misbah devoted his talk to the dynamics of vesicles and membranes as observed *in vitro*. Firstly, he explained the physics of vesicles tumbling over or sticking to a surface in a flow of liquid. Critical dynamic regimes can be observed that depend on the ratio between the external fluid viscosity and the internal vesicle viscosity. The vesicle undergoes shearing stress and competition between adhesion onto the surface and a lift force. As the flow increases, the vesicle is deformed and can be detached from the surface. Above a threshold of the viscosity ratio, another dynamical regime appears in which the vesicles roll and tumble. Secondly, he discussed the fluctuations and deformations induced in phospholipid membranes by the binding of macromolecules. By considering the binding energy of the membrane and the possible diffusion of macromolecules within the membrane, he showed how a self-sustained increase of membrane curvature can be obtained by local recruitment of macromolecules. This phenomenon could provide a theoretical framework for explaining vesicle formation from almost planar membranes, for example from the endoplasmic reticulum.

The lecture of Georgia Barlovatz-Meimon was devoted to the links between cell migration and extracellular matrix (ECM) remodelling. The underlying global regulation scheme is the feedback loop—biochemical and biomechanical signals from the microenvironment induce cellular responses that in turn modify the characteristics of the cell's environment. The plasminogen activator system (Pas) was given as an example of such coupling between the cell and the extracellular matrix. In this system, the type I inhibitor of the plasminogen activator (PAI-1) exists both in a diffusible form and as linked to an ECM protein such as vitronectin. This dual rôle was used to explain how the anti-proteolytic activity of PAI-1 on the ECM is nevertheless compatible with high levels of PAI-1 being a strong indicator of tumour invasiveness. The importance of PAI-1 to cell shape was shown by studying the distribution of F-actin in cells cultured on rigid supports coated with PAI-1 where cells adopted a morphology suitable for rapid migration.

Multi-agent systems are useful for simulating certain self-organizing aspects of biological systems.

Abdallah Zemirline and Pascal Ballet described multi-agent systems in which, firstly, the environment is a 3-D space, secondly, the agents exist in several categories and, thirdly, each category has a small set of rules that define the behaviour of the agent. These agents can represent biological objects and they interact with one another according to the composition and decomposition operations defined for dynamic graphs. They have developed software, BioDyn, in which a multi-agent system is combined with a dynamical mass spring system. This allows, for example, antibody binding or membrane deformations to be simulated and they showed how this could be done by building composite agents out of simpler agents via both top-down and bottom-up approaches. To investigate the assembly and disassembly of large intracellular structures, Patrick Amar has developed a program that is a hybrid between multi-agent systems and cellular automata (which are simpler than agents, only interact locally and move on a grid). He explained that a molecule in his program can interact with its neighbours in four ways – association, dissociation, reaction and catalysis – and can diffuse from one voxel to the next. The numbers and biochemical characteristics of molecules, the size and geometry of the simulated cell, and the time-scale for diffusion and reactions are biologically realistic. He showed how his program could be used to investigate the polymerization of one of the principal constitutents of the eukaryotic cytoskeleton, actin, into filaments as well as the interaction of these filaments with the membrane.

The membranes and intracellular structures subgroup: nuclear processes and nuclear hyperstructures

The lecture of Danielle Hernandez-Verdun was about the relationship between nuclear functions and nuclear organisation. She focused on the nucleolus and the pre-nucleolar bodies (PNBs) which are pre-aggregates of DNA and some of the proteins involved in transcription initiation. She presented the nucleolus and PNBs as highly dynamic structures; indeed, the nucleolus is not defined by a surrounding membrane but, to a large extent, only exists when engaged in its function of making ribosomes. PNBs appear during telophase and travel to transcription sites to form the nucleolus. The dynamics of PNBs was illustrated by videomicroscopy sequences that showed the oscillating features of PNB formation and propagating concentration waves. The possible rôles of PNBs were also reviewed and in particular their possible effects on cell-cycle regulation via modulation of cyclin kinase activity.

Deciphering the code written into chromatin structure and dynamics is one of the great questions of biology. Arndt Benecke showed that chromatin has both positive and negative regulatory effects on gene expression and argued that hypercycles of coactivator and corepressor action on the chromatin constitute this code. The numerous enzymatic modifications of an individual nucleosome change its state in a manner that is *a priori* independent from the underlying DNA sequence. He suggested that a chromatin modification code interpreted by transcriptional coregulators might also regulate all DNA-based nuclear processes including functional nuclear organization in the shape of actively functioning chromatin hyperstructures.

Bertrand Séraphin reviewed RNA splicing mechanisms and focused on the recognition of introns and regulation of alternative splicing. The gene encoding troponin T was taken as example since its sequence, which includes five optional exons, allows a large number of different proteins to be generated. In the second part of his talk, he presented the splicing factors (snRNP U1, U2, ...) and showed how their association into complexes helps to finely tune splicing via a multi-recognition process. Transcription and splicing are coupled in time and space, and he discussed the possible regulation of splicing by external stimuli.

The organization subgroup: spatio-temporal organisation at different levels in biology

The course given by Hans Meinhardt showed how positional information could be generated by the functioning of the system. He applied a concept of fundamental importance—local activation and long-range inhibition—to explain how a diffusion-reaction schema can operate at many levels in biology to generate dynamic structures. He illustrated his talk by treating the problem of pattern formation in systems as different as flies, hydra, seashells and bacteria. Jan Traas and Pierre Barbier de Reuille based their talk on the aerial growth of the model plant, *Arabidopsis thaliana*. This growth depends on the formation of highly organized, stable groups of cells (shoot apical meristems) despite these cells dividing and differentiating rapidly. The explanation may lie in the network of interactions between the cells which involves a plant hormone and membrane-linked transporters. He presented a "virtual meristem" model which has parameters that should allow the action of genes to be identified and the behaviour of mutants to be predicted.

Many mathematical approaches to biological systems are based on differential equations or on partial derivatives although these approaches are

sometimes confounded by the discrete nature of certain biological phenomena. Alexander Bockmayr presented a new, hybrid, type of program language, hybrid concurrent constraint programming, which overcomes this problem. The language, "Hybrid CC", is a declarative one (i.e. it obeys familiar mathematical conventions) with a limited set of primitives that can be used to describe both continuous and discrete transitions. He showed how the language can be used to model splicing in HIV. Marie Aimar addressed the question of how to design a program that can be used in different simulations—and reused as the data change—and also be easy to validate and maintain. She explained why hydrid systems are valuable and focused on an object-oriented language SBML (a Structured Language for Biology) which she used in the context of the "virtual mitochondrion" project to discuss relevant problems and their solutions.

The consensus subgroup: physical chemistry and intracellular organization

To understand fully the controls over gene expression and progression through the cell cycle, it is essential to appreciate the factors responsible for determining the state of the chromosome. Conrad Woldringh explained that within the bacterial cell these factors include the behaviour of *Kuhn segments*. The chromosome can be considered as a chain of relatively stiff 158 nm Kuhn segments that act as springs trying to force the chromosome apart. This self-interaction force is opposed by a cross-interaction energy that acts to increase the volume available for the soluble proteins. The result in the crowded cytoplasm is a phase separation of the nucleoid. He went on to propose that chromosome segregation results from a self-enhancing combination of the Brownian motion of condensed DNA segments plus the attachment of expressed genes to the membrane via the coupled transcription, translation and insertion of proteins into membrane.

If the Holy Grail in the post-genomics era is to obtain a realistic simulation of a cell, that realism is going to have to hold at the molecular level and the interactions of water and macromolecules must be considered. Water is the most abundant molecule within the cell, most of which is within two layers of water molecules around biomolecules. Moreover, water's physico-chemical properties are central to virtually every enzymatic reaction. Pascale Mentré introduced some of the basic concepts needed to understand how the water in contact with the surfaces of proteins and other molecules is structured. She explained that hydrophilic substances can be surrounded by a hydration shell that prevents them from precipitating. The oriented dipolar molecules of water around ionized domains of biomolecules may be in a state of electrostriction in which their density and pressure could reach 1.2 and 34 kbars respectively. Polar domains of biomolecules may also make H-bonds with water; the bonds between π electrons and H_2O can keep hydrophobic residues on the surface of proteins whilst H-bonds between hydrogens borne by aliphatic carbons and H_2O ($CH\cdots OH_2$) might be important in the structure of both DNA and proteins. Moreover, proton conduction in the water surrounding enzymes may be critical for their activity. The rapid movement of water itself might be facilitated by hydrophobic domains within the cell via the mechanism of hydrophobic hydration. Finally, she stressed the importance of the fact that cells must contain different intracellular compartments characterized by different water properties (affecting for example the concentrations of ions). A physicist's view of intracellular water was provided by Marie-Claire Bellissent-Funel. She concentrated on interfacial water (as opposed to bulk water) which includes the water on the surfaces of proteins and lipid membranes. She described neutron scattering studies of translational and rotational diffusion as well as the vibrational density of states of confined water. Her examples included water confined in porous media, in the presence of organic solutes and on the surface of a deuterated C-phycocyanin protein. She showed how the vibrational density of states of interfacial water varies as a function of temperature and of the degree of hydration of this protein. She proposed a picture of interfacial water at room temperature in terms of an increase of the extension of the H-bond network of water as it occurs in supercooled water at a temperature some 25 K lower.

In the context of simulating cells, Eric Fourmentin gave a brief review of projects that include Cybercell, Alpha Project and Silicon Cell. He then described a project initiated by the *Fondation Fourmentin-Guilbert*, SIMEBAC, in which the ultimate object is to contribute to a realistic simulation of *E. coli* via a bottom-up, fine-grained simulation of bacterial metabolism. He focused on the transcription of genes by RNA polymerase and discussed the problems that would have to be overcome in simulating it.

The observability subgroup

Janine Guespin presented the concept of epigenesis, namely, how cells or organisms with the same genotype can have stably different phenotypes as

a result of differences in their history. She illustrated epigenesis by citing experiments in which brief exposure to an inducer of the *lac* operon converted a population of *Escherichia coli* from one in which *lac* expression was stably *off* in a particular medium to one in which it was stably *on* in the same medium. These epigenetic states are examples of positive feedback leading to multistationarity, and they exhibit hysteresis. She explained how a single positive feedback loop is needed for multistationarity in a system of non-linear interactions whilst a negative loop is needed for homeostasis (with and without oscillations). She then applied these concepts to the case of mucus production in *Pseudomonas aeruginosa*, which is of importance in cystic fibrosis, to show how the operation of feedback circuits may mean that this production is actually due to epigenesis (see below).

Jean-Pierre Vannier began by describing how the initial stages in haematopoesis are responsible for both the production of the different cells that will differentiate into particular types of cells and the maintenance of a population of stem cells that gives rise to these differentiating cells. Autocrine secretion is likely to be a part of positive feedback circuits responsible for epigenetic states in which stem cells are either quiescent or active in multiplying and differentiating. Vannier and David Campard then presented a model, based on the Boolean method of logical analysis formulated by René Thomas, in which an important rôle is played by the cells' microenvironment. Plants are able to store environmental stimuli and to respond to them much later. Janine Guespin, standing in for Michel Thellier, explained that an asymmetrical growth of cotyledons occurs when the apex of a *Bidens pilosa* L. seedling is decapitated; the asymmetric nature of this growth reflects asymmetric treatments inflicted on the seedling before the decapitation step. This system has been used to study how plants store and integrate signals before committing themselves to a growth strategy adapted to the environment. The logical analysis method can, she revealed, be used to explain the interplay between storage and recall functions and to predict stable states that can be tested experimentally.

The question of how to test the idea that an epigenetic state is responsible for mucus production in *P. aeruginosa* infections of the lung (see above) paved the way for the next speaker, Gilles Bernot. Many models in biology contain parameters that at best cannot be measured directly and at worst are uncertain or hidden. Gilles Bernot explained that this leads to the idea that only a class of models—rather than one specific model—can be validated. Development of a programme of formal logic to assist in such validation should take into account firstly the coherence of hypotheses and data and secondly, since such coherence does not necessarily mean a hypothesis is correct, the need to generate pertinent experimental tests. He then showed how algorithms used in testing computer programs could be combined with computational tree logic to suggest key experiments.

Daniel Claude gave a course on the essentials of control theory and introduced the notions of commandability, observability, and identifiability. Commandability means that there is always a command that allows the system to be driven from one state to another via a defined trajectory. Observability is about being able to distinguish between different initial states of the system by following the evolution of observable parameters. Identifiability is about identifying the parameters of the system by studying input/output relationships. He said that it is now possible to use a probabilistic algorithm firstly to obtain the set of observable parameters of a system and secondly to decide how many other, non-observable, parameters are required. He illustrated the use of this algorithm in the case of the toxicity of certain antibiotics—aminoglycosides—on the human kidney where it shows that the best time of day for administration of the antibiotic is 13:30.

Conclusion

The advantages of the long-term schedule of regular meetings at Genopole became clear at the Dieppe conference. The friendly, relaxed atmosphere at Dieppe was conducive to fruitful exchanges between specialists from different disciplines and between specialists and students. Indeed, the level of student participation was remarkably high and the (anonymous) evaluation by participants was positive. The ultimate take-home message of the conference may be that post-genomic biology is going to be dominated by multidisciplinary teams formed both from existing specialists and from a new generation of interdisciplinary students. Collaborative interactions between such teams are essential for progress towards an integrated picture of the cell and all the rewards that that will bring.

Acknowledgments

We thank Catherine Meignen for secretarial help. For funding, we thank Genopole, the CNRS, INRA, the Conseil Régional d'Ile-de-France, the Région de Haut Normandie, the Fondation Fourmentin-Guilbert, and the universities of Rouen, Grenoble, Toulouse, Montpellier and Evry.