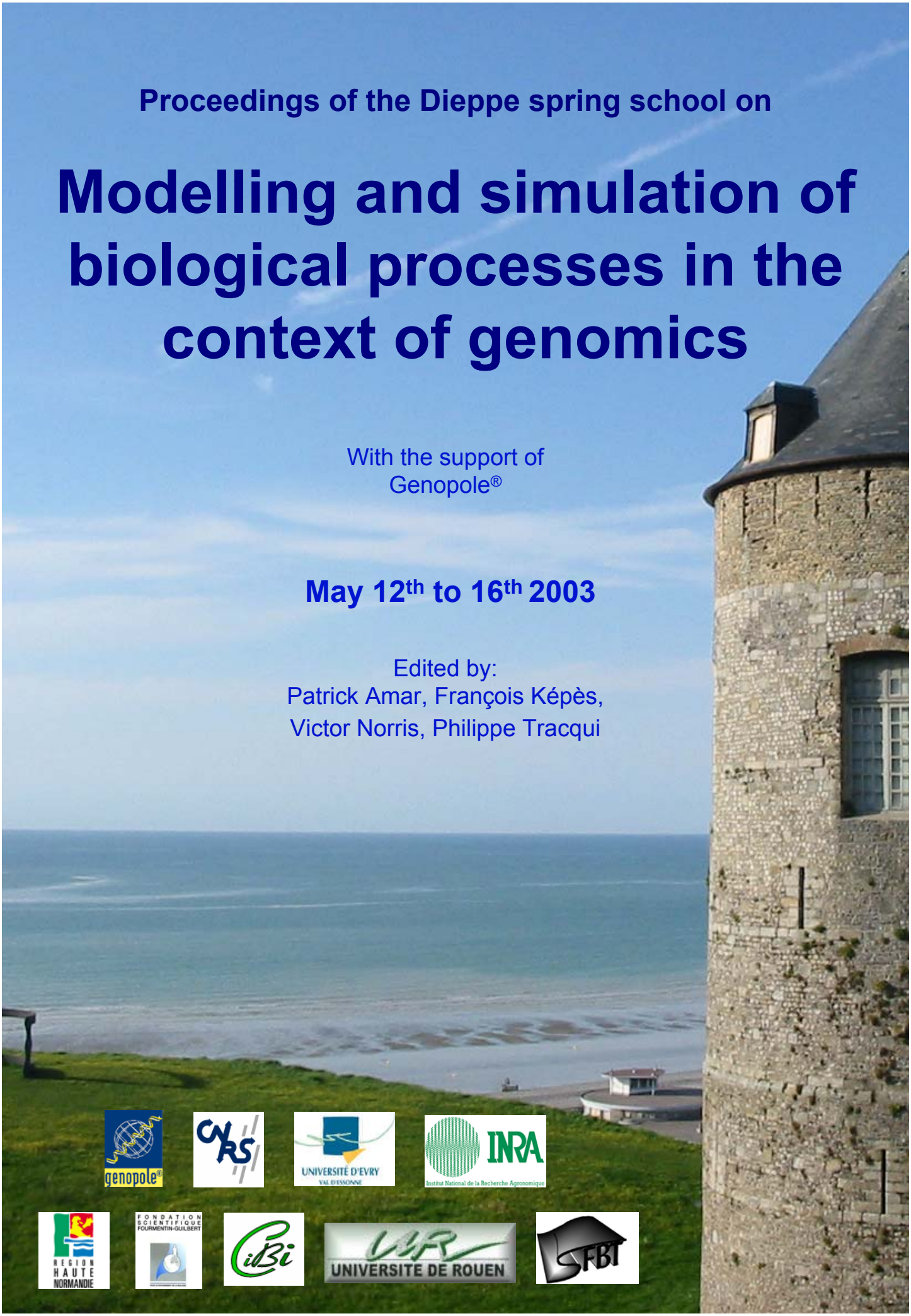
Proceedings of the Dieppe spring school on

Modelling and simulation of biological processes in the context of genomics



With the support of
Genopole®

May 12th to 16th 2003

Edited by:
Patrick Amar, François Képès,
Victor Norris, Philippe Tracqui

FOREWORD

The post-genomic era of biology is characterised 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-organising 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 also 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 multi-disciplinary 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, 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 *organisation* 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 organised 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. These two first editions of the spring school in Autrans and Dieppe will be followed by others each year.

Patrick Amar, Pascal Ballet, Gilles Bernot, Alessandra Carbone, Jean-Paul Comet, Franck Delaplace, Jean-Marc Delosme, Maurice Demarty, Marie Dutreix, Christine Froideveaux, Jean-Louis Giavitto, Christophe Godin, Misha Gromov, Janine Guespin, Roberto Incitti, Marcelline Kaufman, François Képès, Aurélien Mazurie, Olivier Michel, Jean-Pierre Nadal, Victor Norris, François Radvanyi, Alain Rambourg, Michel Thellier, Philippe Tracqui, Abdallah Zemirline.

ACKNOWLEDGMENTS

We would like to thank the seminar participants, who have contributed in a way or another to this book. It gathers overviews of the talks, discussions and roundtables, original articles contributed by speakers, abstracts from attendees, and courses proposed by the epigenesis group to review or illustrate matters related to the scientific topics of the seminar.

Of course, the organisation team would like to express gratitude to all the staff of the *Hotel La Présidence* for the very good conditions we have found during the seminar.

Special thanks go to Hélène Pollard (directrice de Genopole Recherche) and Catherine Meignen (chargée de l'organisation des manifestations scientifiques à Genopole Recherche) for their help during the preparation of the spring school and for their assistance in preparing this book for publication.

We would like to thank Paul Hossenlopp (responsable de la formation cellule de la politique de l'Emploi, CNRS) for his encouragement to this spring school and for his support to the organisation of the next editions of this school.

We would also like to express our thanks to the sponsors of this seminar for their financial support allowing the participants to share their enthusiasm and ideas in such a constructive way. They were:

- Genopole® Évry (http://www.genopole.org),
- Centre National de la Recherche Scientifique - CNRS (http://www.cnrs.fr),
- LaMI - Université d'Evry (http://www.lami.univ-evry.fr),
- Institut National de la Recherche Agronomique - INRA (http://www.inra.fr),
- Région Haute Normandie (http://www.region-haute-normandie.fr),
- Fondation Scientifique Fourmentin-Guilbert (http://www.fourmentinguilbert.org),
- Centre Interdisciplinaire de Biologie Intégrative and Université de Rouen (http://www.univ-rouen.fr),
- Société Francophone de Biologie Théorique - SFBT (http://sfbt.org).

The editors
Patrick Amar, François Képès,
Vic Norris and Philippe Tracqui

CONTENTS

PART 4 - LECTURES / ABSTRACTS

PART 5 - POSTERS

PART 6 - HIGH LEVEL COURSES

PART 7 - LIST OF ATTENDEES

Part 1

CONFERENCE SYNTHESIS

CONFERENCE SYNTHESIS

Vic Norris[1], Patrick Amar[2,3], 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, UPRESA CNRS 6037, Faculté des Sciences & Techniques, Université de Rouen, 76821, Mont-Saint-Aignan, France

[2]La.M.I. Université d'Évry Val d'Essonne & CNRS UMR 8042, Tour Évry 2, 523 Place des terrasses de l'agora, 91000 Évry, France

[3]Laboratoire de Recherches en Informatique, Université Paris Sud & CNRS UMR 8623, 15 avenue George Clémenceau, 91405 Orsay Cedex, 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, France

[7]Laboratoire des Techniques Imagerie Modélisation Cognition, CNRS UMR 5525, Faculté de Médecine, 38706 La Tronche, France

[8]ATelier de Génomique, Cognitive, CNRS UMR8071/genopole, 523 Terrasses de l'Agora 91000 Evry, France

The organisation 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 organised, 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: 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' 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 affects 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 or 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 hyper-structures

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 micro-environment 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 role 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-organising 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, but use a virtual grid to speed up the research of close neighbours). 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 polymerisation of one of the principal constituents 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-nuclear bodies (PNBs) which she presented 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 are pre-aggregates which form 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 PNBs formation and propagating concentration waves. The possible roles 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 organisation 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 consensus subgroup: physical chemistry and intracellular organisation

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 Kuhn segments which are relatively stiff segments of 158 nm 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 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 and most of it is within two layers of water molecules around biomolecules. Moreover, its 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 ionised domains of biomolecules may be in a state of electrostriction in which their density and pressure can 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 born by aliphatic carbons and H_2O ($CH\ldots OH_2$) are 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 stresssed the importance of the fact that cells must contain different intracellular compartments characterised 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 the 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 vary 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 H-bond network of water as it occurs in supercooled water at a temperature some 25ºK 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 exhibit hysteresis. She explained that 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 role 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 program 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 for administration of the antibiotic is 13.30.

Conclusion

The advantages of the long-term schedule of regular meetings in Genopole became clear at the Dieppe conference. The friendly, relaxed atmosphere 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 the participants was positive. The take-home message of the conference may be that the post-genomic biology is going to be dominated by multi-disciplinary teams formed both from existing specialists and from a new generation of interdisciplinary students. And collaborative interactions between such teams are essential for progress towards an integrated picture of the cell and all the rewards that that will bring.

Part 2

ARTICLES

BioDyn: a Software for Cell Physiology Simulation

Pascal Ballet, Abdallah Zemirline, Lionel Marcé

Département d'Informatique, EA2215 Université de Bretagne Occidentale BP 809, 29285 Brest Cedex
Pascal.ballet@univ-brest.fr, zemirlin@univ-brest.fr

Abstract

We describe a software for cell physiology simulation, which uses dynamic graphs and their composition and decomposition operations for modelling the multiagent systems and their behaviours.

1. Introduction

Nowadays the multiagent systems (MAS, for short) [1] are used as a tool to understand the structure, the behaviours and the functions of biological processes. Indeed, the MAS enable to simulate living organisms in order to perceive what characterises these self-organizing systems. They are also used to discern how they can produce achievements that exceed the sum of all the performances of their components. A MAS is a system of multiple active objects, called agents, which have the ability to perceive, to act, to interact and to move into their environment.
Each of the MAS here considered is constituted of the following components :
(i) an environment which is a three-dimensional space,
(ii) one or several categories of agents; two agents, which are not from the same category, differ in the simulator by their colour or/and their shape or/and their behaviour,
(iii) to each category is associated a small set of rules which governs the agent behaviours.

A cell can be divided into numerous autonomous entities in interaction. Molecules, ions, membranes and organelles can be seen as interacting reactive agents. A reactive agent is an abstraction or a simplification of a biological entity. It has sensors to get information from its local environment, a simple behaviour to change its internal state and actuators to modify the environment. Many data come from the big scale biology. These data mainly concern the genome but they also deal with the cell physiology, including membranes, hyper-structures and compartments. To understand the cell functionalities the genome is essential. Nevertheless, we think that the physiological environment of the DNA macromolecule plays a crucial role too. That is why we develop a software, named BioDyn, allowing the biologists to test their assumptions about the interactions between DNA and its local environment. The engine of the software is a *dynamical mass spring system* (DMSS) coupled with a MAS. The DMSS is designed to basically represent biological entities like ions, molecules, macromolecules, membranes and compartments. The MAS gives high-level behaviours to the DMMS to perform complex tasks like chemotactism, endocytosis, invagination and mitosis. The structure of any agent of this MAS is what is called a *dynamic graph*. Therefore, the agents of the MAS are able to interact together accordingly to the *composition and decomposition operations* defined for graphs.

This work comprises two parts. The first one introduces the dynamic graphs and describes the composition and decomposition operations for graphs. The second part deals with the software engine that is used to simulate biological processes.

2. Dynamic graphs

Graphs [2], [3] were defined for making visible to the eye a binary relationship between the objects in some domain. A *graph* $G = (V, E)$ consists of a *nodeset* V and of an *edgeset* (resp. *arcset*) E, where an *edge* (resp. *arc*) is an unordered (resp. ordered) pair of nodes. When E is an edgeset (resp. arcset) the graph $G = (V, E)$ allows to represent by a picture a *symmetric* (resp. an *asymmetric*) binary relation, with points for the nodes and lines (resp. arrows) for the edges (resp. arcs).

This representation of binary relations has allowed to model and then to solve a large number of complex problems that arise in engineering. The MAS approach which is presented in this work uses what is called dynamic graphs for visualizing and simulating biological processes.

Dynamic graphs appeared in the eighties with the networks subject to failures as the telecommunication networks and the computer networks, and then with the mobile networks as the mobile telephony and the satellite network. A *dynamic graph* is a graph which is modified in the course of time by adding and/or deleting nodes or edges (resp. arcs). Thus, a dynamic graph can be considered as a sequence of classical graphs each one representing one state of the dynamic graph at a given time. In this paper, the definition which seems appropriate for the dynamic graphs used for representing MAS is the following :

2.1. Definition

A *dynamic graph* (resp. *a directed dynamic graph*) is a graph $G = (V, E)$ evolving in a three dimensional space such that the elements of the nodeset V and those of the edgeset (resp. arcset) E are depending on time, more precisely :

- the sets $V = V(t)$ and $E = E(t)$ depend on time and do not have necessarily the same elements at different times;
- each $v \in V$ is a point (x, y, z) of $|R^3$ such that $x = x(t)$, $y = y(t)$ and $z = z(t)$ where t is a time variable.

Obviously, all the problems, properties and composition and decomposition operations known for graphs remain defined for dynamic graphs. In particular, this is true for characterizing some special class of graphs, recognizing members of a given family of graphs, finding solutions to some graph optimisation problems, looking for some special cutsets for a given family of graphs or for special compositions or decompositions for graphs which preserve some properties. Hereafter, composition and decomposition operations for graphs are described. Such compositions and decompositions can be used with dynamic graphs to simulate biological processes, with eventually some restrictions taking in account the specificity and the context of the simulated phenomena.

2.2. Examples

1) Graphs $G1(t)$ and $G2(t)$ of Figure 1 are isomorphic but are distinct dynamic graphs.

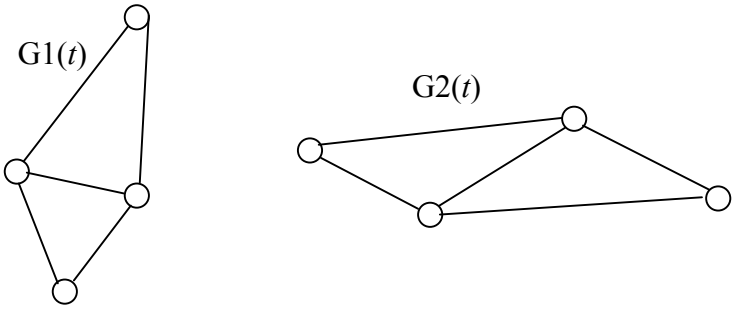


Figure 1

2) Internet provides several examples of dynamic graphs; we can mention the following :
- the web page graph, that is the dynamic graph having as nodes the web pages and as arcs the hypertext links between the web pages);
- the graph of the physical connections between machines having an IP number.

2.3. Compositions and decompositions for graphs and dynamic graphs

The composition and decomposition operations presented hereafter mainly arise from the study of perfect graphs [4], [5]. A graph $H = (W, F)$ is an induced subgraph of a graph $G = (V, E)$ if $W \subseteq V$ and $F = \{e \in E ; e = uv$ and $u, v \in W\}$. If $H = (W, F)$ is an induced subraph of G, H can be denoted $G(W)$. Besides, let $G = (V, E)$ be a graph and $U \subset V$; the subset U is a *clique* (resp. *stable set*) if for all $u, v \in U$ such that $u \neq v$ then $uv \in E$ (resp. $uv \notin E$). Let $\alpha(G) = \max \{|U|; U \subset V$ and U is a stable set in $G\}$, where $|U|$ denotes the size of U; and $\omega(G) = \max \{|U|; U \subset V$ and U is a clique in $G\}$. $\omega(G)$ (resp. $\alpha(G)$) is the size of the greatest clique (resp. stable set) of G. We define the *chromatic number* $\chi(G)$ of $G = (V, E)$ as being $\chi(G) = \min \{k$; there is a partition of V into k disjoint stable sets$\}$. A graph G is *perfect* if for all induced subgraph H of G, $\chi(H) = \omega(H)$. To recognize that a graph is perfect is not easy. Progress towards solving this difficult problem is made by designing a polynomial-time algorithm to recognize members of some restricted class of perfect graphs. This approach led searchers to define some composition and decomposition operations for perfect graphs [4], [5], [6], [7], [8] which preserve perfection.

In the following pairs of operations, each operation is the converse of its twin sister.

2.3.1. Addition and deletion of nodes or edges

Let $G = (V, E)$ be a (dynamic) graph; and let $U \subseteq V$ and $F \subseteq E$.

Deletion of nodes : The graph obtained from G by deleting all the nodes in U and all the edges having at least one end in U is the induced subgraph $G(V \setminus U) = (V \setminus U, E')$, where $E' = \{e \in E ; e$ has its two ends in $V \setminus U\}$.

Deletion of edges : The graph obtained from G by deleting all the edges in F is the graph $H = (V, E \setminus F)$.

The converse operation of a deletion consists in adding to G new nodes and/or new edges.

2.3.2. Contraction and substitution

Substitution : Let $G = (V, E)$ be a graph and $v \in V$; and let $H = (W, F)$ be another graph such that $W \not\subset V$.

The graph G', obtained from G by replacing the node v by the graph $H = (W, F)$ and each edge of E incident to v by $|W|$ edges, each one is incident to a distinct node in W, is called the graph obtained from G by substituting H to v. See Figure 2.

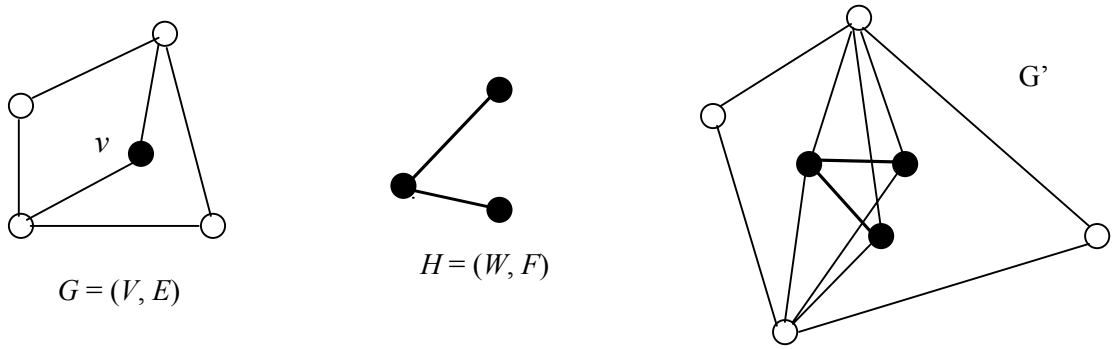


Figure 2

Contraction : Let $G' = (V', E')$ be a graph and $W \subseteq V'$.

The graph G, obtained from G' by deleting the edges of the induced subgraph $H = G'(W)$ and replacing the subset W by a new node v so that each edge of E' having one end x in W and the other one y in $V' \setminus W$ is replaced by an edge with v and y as ends, is called the graph obtained from G' by contraction of W. See Figure 2.

2.3.3. Cutset and set identification

Cutset : Let $G = (V, E)$ be a connected graph and $A \subset V$.

The subset A is called a *cutset* if the induced subgraph $G(V \setminus A)$ is disconnected. When A is a cutset, G can be decomposed into two smaller graphs $G1 = G(V1)$ and $G2 = G(V2)$, with $A \subset V1 \subset V$ and $A \subset V2 \subset V$. See Figure 3.

Set identification : Let $G1 = (V1, E1)$ and $G2 = (V2, E2)$ be two graphs.

If there exist $A1 \subset V1$ and $A2 \subset V2$ such that the induced subgraphs $G1(A1)$ and $G2(A2)$ are isomorphic, a composed graph $G = (V, E)$ is obtained from $G1$ and $G2$ by identifying $G1(A1)$ and $G2(A2)$, each node of $A1$ is identified with its image in $A2$ in the isomorphism between $G1(A1)$ and $G2(A2)$ getting in this way $G(A)$. See Figure 3.

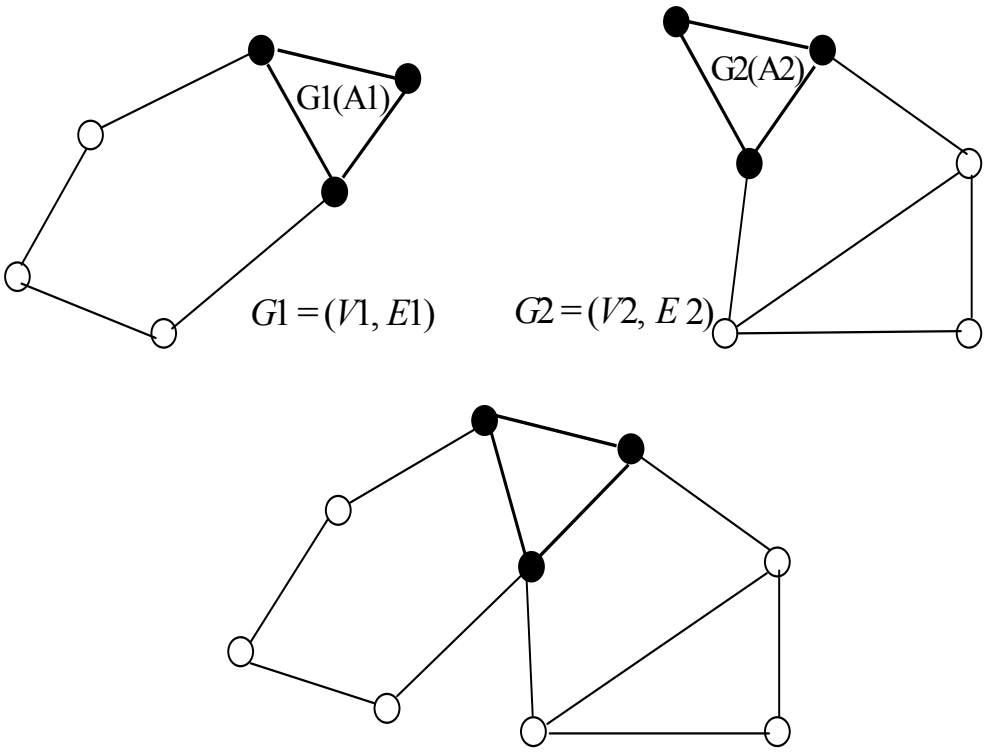


Figure 3

2.3.4. Cut and join

Cut : Let $G = (V, E)$ be a connected graph and $C \subset E$.
The edge set C is called a *cut* if the graph $H = (V, E \setminus C)$ is disconnected.
When a connected graph G has a cut, it can be decomposed into two connected subgraph $G1$ and $G2$ by deleting the edges of a minimal cut. See Figure 4.

Join : Let $G1 = (V1, E1)$ and $G2 = (V2, E2)$ be two graphs and $B1 \subseteq V1$ and $B2 \subseteq V2$ be two subsets. A composed graph $G = (V, E)$ can be obtained from $G1$ and $G2$ by joining some vertices of $B1$ to some vertices of $B2$ such that:
$V = V1 \cup V2$ and $E = E1 \cup E2 \cup \{$edges uv with u in $B1$ and v in $B2\}$. See Figure 4.

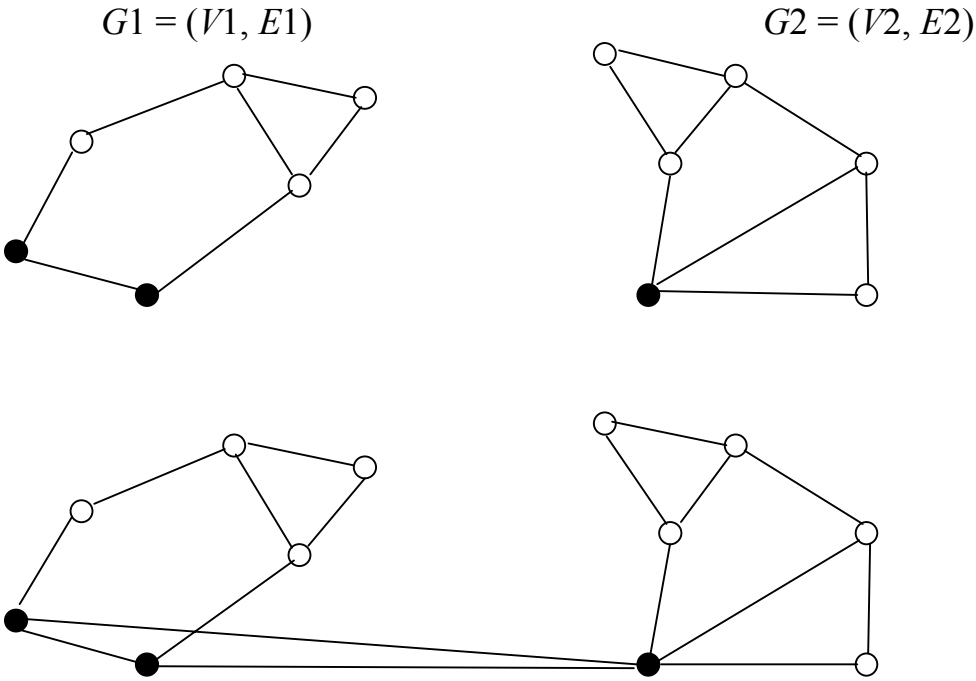


Figure 4

2.3.5. Amalgam

A composition operation that is applied on two graphs $G1 = (V1 \cup A1 \cup B1, E1)$ and $G2 = (V2 \cup A2 \cup B2, E2)$ (see graphs G1 and G2 of Figure 4) where

(1) for $i = 1, 2$, Vi, Ai and Bi are pairwise disjoint,

(2) $G1(A1)$ and $G2(A2)$ are isomorphic,

is defined in the following way :

The composed graph G is obtained from $G1$ and $G2$ by

. identifying the subsets $A1$ and $A2$ (each node of $A1$ is identified with its image in $A2$ in the isomorphism between $G1(A1)$ and $G2(A2)$) ; and

. joining some vertices of $B1$ to some vertices of $B2$.

The graph $G = (V, E)$ is called the *amalgam* of $G1$ and $G2$. See Figure 5.

Conversely, if for a graph G there exist two graphs $G1$ and $G2$ satisfying conditions (1) and (2) such that G is the amalgam of $G1$ and $G2$, then the graph G is said to be decomposed by this amalgam into $G1$ and $G2$.

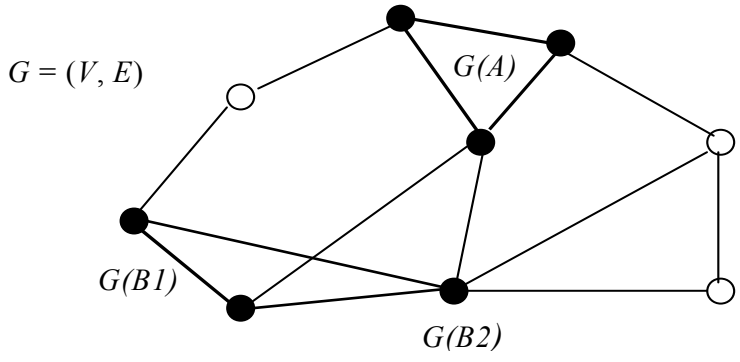


Figure 5

3. The BioDyn Engine

3.1. The BioDyn Engine is composed with a DMMS, which allows physical constraint for the simulated biological entities and a MAS designed to give high-level behaviours for set of entities. We begin with a short description of the DMMS and then, we will see how the MAS is built.

The DMMS is made of mass entities (nodes) and spring entities (edges). A basic mass entity has a position and can be attached to spring entities according to receptors (see figure 6). A spring entity has a length and a stiffness. The mass entities and the spring entities, attached with receptors, represent a more complex entity, which is called "composite agent".

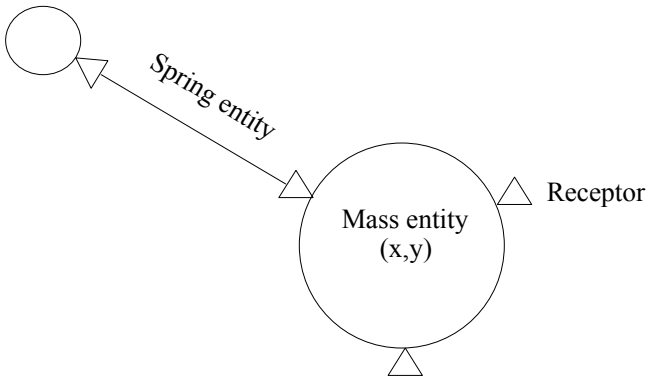


Figure 6 : Basic entities of the DMSS

This complex entity can be deformed to represent, for example, a membrane deformation (see figure 7).

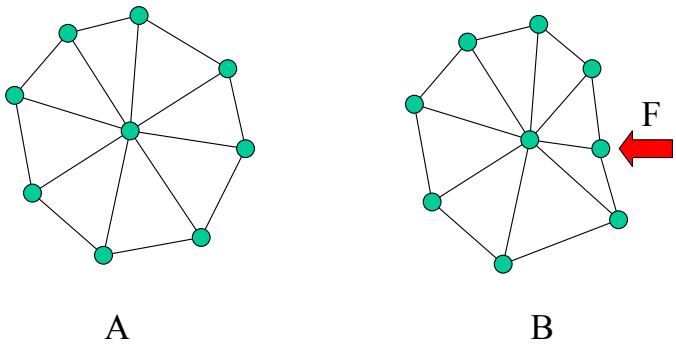


Figure 7 : Deformation of a composite agent.

The system represents entities from one angstrom to one micro-meter when it works at the molecular level. We suppose that the environment is the intra-cell compartments. I this case, the forces are from one pico-newton (low energy bounds) to nano-newton (covalent bounds). The time step is 10^{-10} second and the temperature is 311 kelvins. The dynamical viscosity is set to 100 $kg.m^{-1}.s^{-1}$.

When the system is used to represent entities at the cellular level, we suppose that the environment is something like lymphatic ganglion. The dimensions are from 10^{-7} meter to one milli-meter. Covalent bounds are unbreakable and low energy bounds are about one pico-newton. The broonian movment, coming from the temperature are 10^{4} time less important than at

the molecular level, but the fluid movments (like blood in a veine) is crucial. The time step is 10^{-6} second and the dynamical viscosity is again set to 100 $kg.m^{-1}.s^{-1}$.
This DMMS is a bottom-up system: the structures build themselves thanks to their movment and their interactions. Sometime, it could be usefull to add high-level behaviors to perform more complex simulations. That is why, we introduce a MAS, which has a top-down architecture.
The MAS is made of composite agents. A composite agent has one or several scenarii, which can be applied to perform high-level actions like mitosis, growth, chemotactism or pseudopodes. To make a high-level action, the composite agents of the MAS receive information (called events) from the DMMS entities (bottom-up approach). The events coming from the DMMS are, for a node: *on link attachment*, *on unlink*, *on delete*, *on clone*, *on timer* and *on age*. For a link, the possible events are: *on delete*, *on timer* and *on age*.
The actions for a composite agent (high level actions) consist in modifying the DMMS (top-down approach). The composite can change the stiffness and the size of its links and it can remove or create links and nodes. It can also apply forces to its nodes.
To realise a high level behavior, the composite looks at its received messages. If the messages validate a specific test (for example *link_attachment(ident) and age > 4 s or node_delete*), it performs a set of actions (creation of nodes or links for instance).
To illustrate our words, we propose you an example at the cellular level into the next chapter.

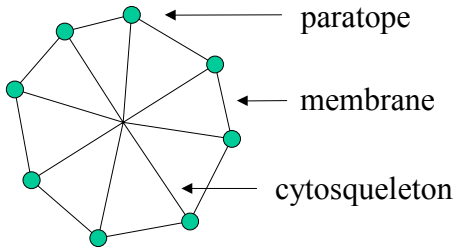
3.2. Example at the cellular level

This is a simple example concerning a humoral immune response. We take into account a B-Cell with its B receptors, antibodies and bivalent antigens. The very simplified mechanisms we simulate is the following:

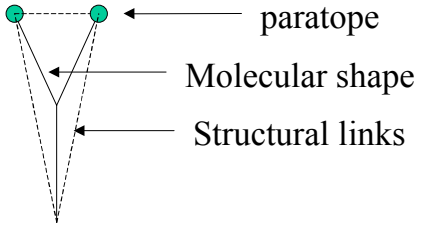
1- when an antigen is bound with a B-Cell receptor, the B-Cell endocytose the antigen.
2- Then, the B-Cell creates antibodies that diffuse into the lymphatic system.
3- The antibodies fix the bivalent antigen.

We begin by the DMSS.

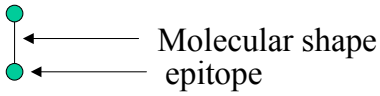
First, we create a B-cell with epitopes:



Second, we create a symbolic antibody. A symbolic antibody is an entity that represents a set of antibodies, because of the different scale levels.

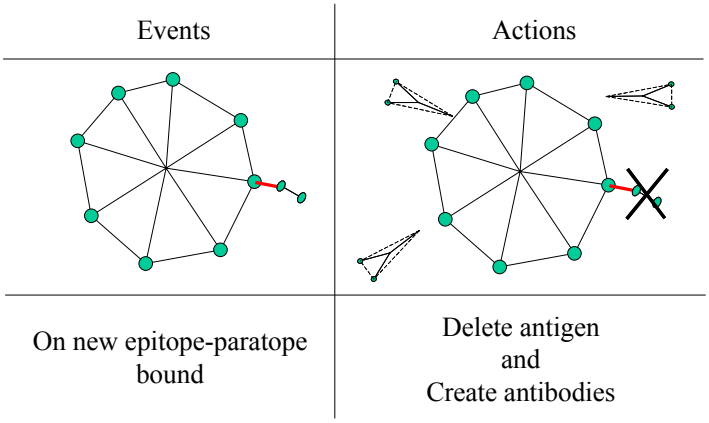


Third, we build the bivalent antigen :

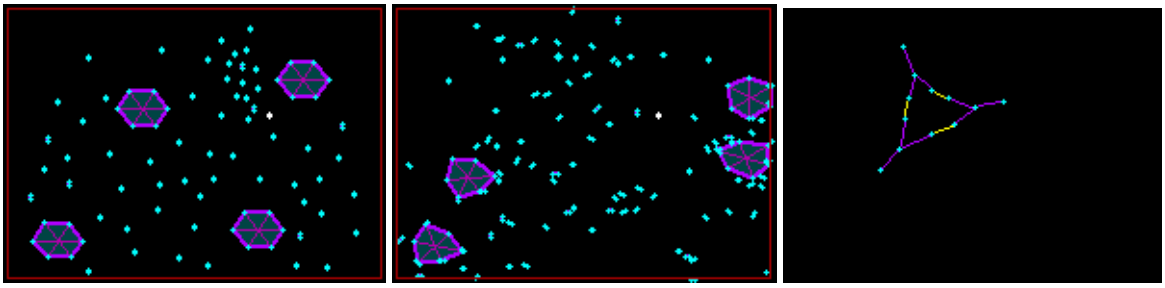


Then, we add a high level behavior thanks to the MAS.

For the B-Cell, we represent the endocytosis :



Finally, we put four B-Cell agent and one hundred antigen agents into an environment of 50 micrometers by 50 micrometers (Step 1) and we observe the evolution of the system (Step 2 and 3):



Step 1 Step 2 Step 3

At the Step two, we observe that B-Cells are creating antibodies and at the step 3 we zoom in to observe the formation of an antibody-antigen structure (see plate 1).

3.3. Conclusion

All the entities (from the MAS or from the DMSS) are “living” in a continuous environment, which has a gravity acceleration, a viscosity and a time step. These parameters should be useful for modelling biological mechanism into a dynamic environment. We think this software could be interesting for molecular, macromolecular and physiological applications.

4. References

[1] J. Ferber, *Les Systèmes Multiagents*, InterEdition, Paris 1995.

[2] C. Berge, *Graphs and Hypergraphs*, North-Holland, Amsterdam 1973.

[3] R. Distel, *Graph Theory*, Springer-Verlag, New York 1997.

[4] C. Berge, V. Chvátal, *Topics on Perfect Graphs*, Ann. Discrete Math., 21, North-Holland, Amsterdam, 1984.

[5] A. Brandstädt, V.B. Le, J.P. Spinrad, *Graph Classes a survey*, SIAM Monographs on Discrete Mathematics and Applications, Philadelphia 1999.

[6] W.-L. Hsu, *Decomposition of perfect graphs*, J. Combin. Theory B, 43, 1987, 70-94.

[7] M. Burlet, J. Fonlupt, *Polynomial algorithm to recognize a Meyniel graph*, Ann. Discrete Math., 21, 1984, 225-252.

[8] A. Zemirline, *Quasi-star-cutsets and some consequences*, European Journal of Operational Research 137, 2002, 326-335.

Chromatin dynamics are a hyperstructure of nuclear organization

Arndt Benecke

Institut des Hautes Études Scientifiques - 35, route de Chartres, 91440 Bures-sur-Yvette, France &
INSERM U417 - 184, rue du Faubourg St.-Antoine, 75012 Paris, France
tel: + 33 (0)1 60 92 66 65 fax: + 33 (0)1 60 92 66 09 e-mail: arndt@ihes.fr

Abstract

Chromatin structure and dynamics have received over the last years increasing attention as being fundamentally involved in all DNA-based nuclear processes. In a necessary trade-off between spatio-temporal packaging of the genome and molecular processes such as gene transcription, genome replication, and DNA repair, chromatin was initially regarded as a 'barrier' to latter processes. Only recently it was understood that on the contrary chromatin also plays an active and equally essential role in assuring the correct assembly and functioning of the required molecular machines. Particularly interesting is the fact that chromatin has, both negative and positive, regulatory effects on DNA-based nuclear processes. Given recent biochemical work on transcription coregulators and their crosstalk with other genome maintenance and regulation pathways a model has been developed which suggests a secondary layer of coding by the chromatin structure of the genome. This code, interpretated by coregulators, might not only be regulatory information for the transcription process but also serve to coherently orchestrate all DNA-based nuclear processes and moreover provide the basis for functional nuclear organization. The dynamic structure of the chromatin-code and its mechanical and functional characteristics are highly reminiscent of cellular hyperstructures. We evidence here that nuclear organization is achieved through formation of a functional chromatin-hyperstructure which induces local symmetry-breaking to provide for non-equilibrium regulatory and mechanical processes.

Keywords: hyperstructure, non-equilibrium thermodynamics, symmetry breaking, chromatin-sequence, transcription coregulator, nuclear organization.

1. Chromatin structure

DNA in eukaryotes is tightly associated with a group of basic proteins called histones. The four different core histones, H2A, H2B, H3, and H4 form a hetero-quatromer structure that dimerizes to form an eight polypeptide (two of each core histone) protein domain, around which roughly 170bp of DNA organize in two and a half helical turns to form a nucleosome. The histone-protein core thereby strongly interacts with the poly-anionic DNA. The nucleosome is the basic repeat unit of eukaryotic genomes, and except for some rare

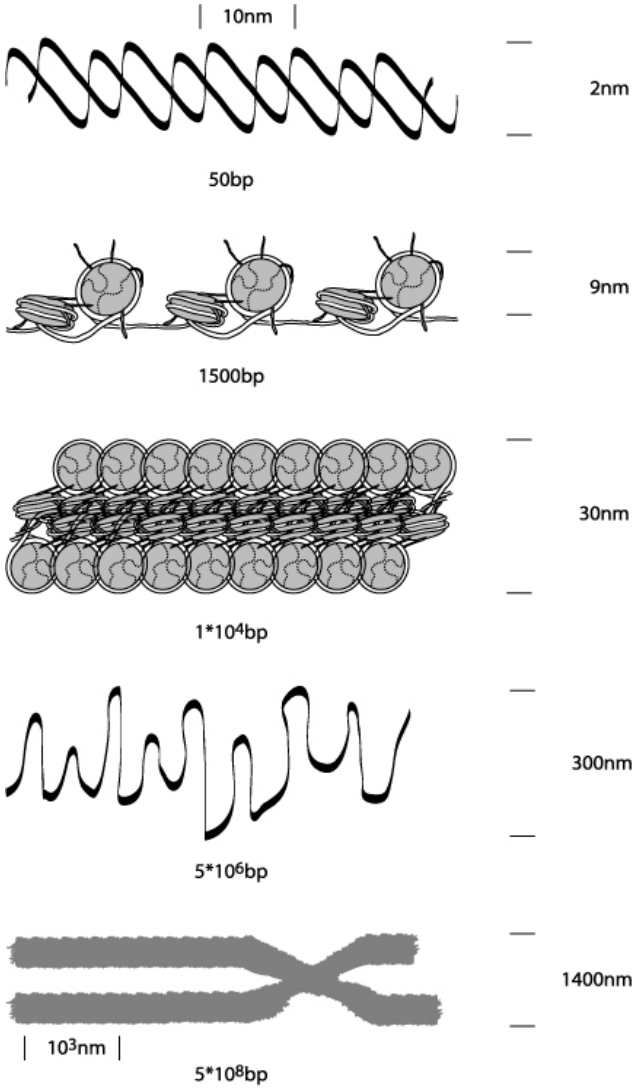


Figure 1: Macroscopic chromatin structure.

The different organizational units of chromatin are schematized. Individual nucleosomes (9nm fiber) associate to higher-order nucleosomal arrays (30nm fiber), often accompanied by linker Histone (H1, and H1-like) and/or HMG protein incorporation. The 30nm fiber can be further condensed into chromatin loops (300nm structure) that eventually form condensed chromosomes during mitosis. In a resting cell greater 90% of the DNA is compacted into highly condensed chromatin (heterochromatin; corresponding essentially to the 300nm loops), and only the reminder is active with respect to DNA-based nuclear processes (euchromatin). Obviously, this ratio is dynamic and the parts of DNA accessible to the nuclear machinery are thought to vary significantly.

circumstances the entire genomic DNA is present in the form of nucleosomes. Two nucleosomes are separated by a short stretch of non-nucleosomal DNA, the linker DNA. Approximately, every two hundred base-pairs of DNA are thus associated with one nucleosome. Given an estimated length of the human genome of 3x10E9 base-pairs, every nucleus has thus roughly 1.5x10E7 nucleosomes. The crystal-structure of an individual nucleosome has been determined, and illustrates very well the 'bead on string' concept developed previously based on electron micrographs [1, 2].

Genomic, nucleosomal DNA can be classified into different structurally distinct forms of organization. This classification is a result of electron and light microscopic studies including probing of the structure with molecular probes such as short DNA-tags, and correlates more or less with the activity of the underlying DNA-sequence. Figure 1 illustrates the different major classes of nucleosomal DNA or chromatin. From these studies it became evident that nucleosomes are at the basis of the spatial-temporal packaging of DNA. The total genomic DNA of a human cell has a length of two meters along its major axis, and a radius of about two nanometers. Any DNA-based nuclear process, such as replication and sister-chromosome segregation during mitosis needs to account for this architectural particularity of having a difference of nine orders of magnitude between the two axes. Packaging DNA into nucleosomes and having higher-order nucleosomal structures such as the 30nm and 300nm fibers evidently facilitates such processes and achieves compaction of up to five orders of magnitude, as impressively illustrated by the formation of condensed chromosomes during mitosis [3].

The different chromatin states are a function of incorporation of non-core histones such as histone H1, the (a) linker histone, and other basic proteins such a high-mobility group (HMG) proteins. Linker histones are thought to tie together two neighboring nucleosomes into dimers. Furthermore, other non-histone, nucleosome-interacting proteins such as Sir factors or polycomb group proteins are able to induce higher-order nucleosomal arrays (i.e at telo- or centromeres). Finally, so-called matrix attachment regions (MAR) and nuclear bodies (NR) provide centers of interaction with the nuclear cytoskeleton, and organizational centers to promote certain higher-order forms such as chromatin-loops. These different structures are dynamic, being regulated by cellular activity, however, thought to be close to equilibrium states of chromatin [4].

2. Chromatin Dynamics

Chromatin structure and its dynamics have received over the last years increasing attention as being fundamentally involved in all DNA-based nuclear processes. In a necessary trade-off between spatio-temporal packaging of the genome and molecular processes such as gene transcription, genome replication, and DNA repair, chromatin was initially regarded as a 'barrier' to latter processes. Only recently it was understood that on the contrary chromatin also plays an active and equally essential role in assuring the correct assembly and functioning of the required molecular machines. Particularly interesting is the fact that chromatin has, both negative and positive, regulatory effects on DNA-based nuclear processes [5].

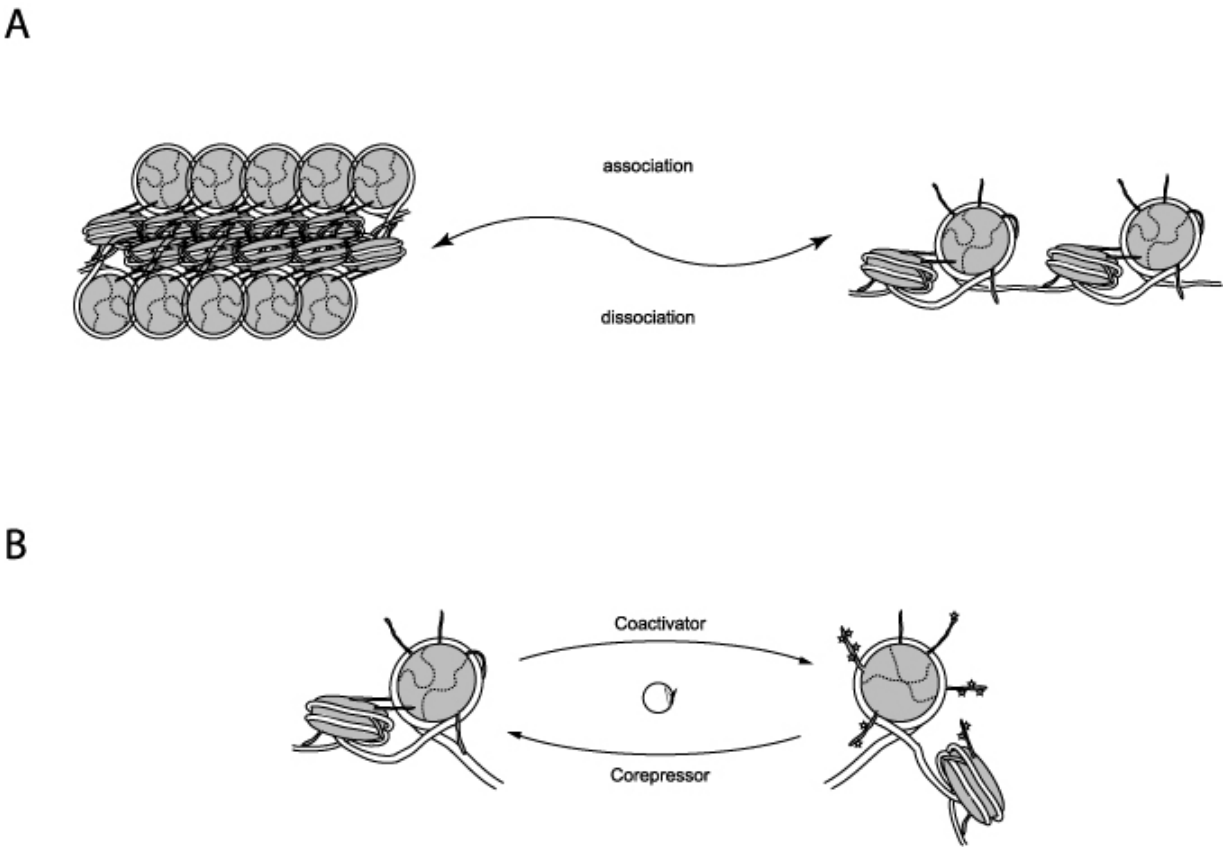


Figure 2: Two distinct types of chromatin dynamics.

Recent progress in understanding the interactions between nuclear machineries operating on the DNA code and the chromatin environment of these DNA segments has unrevealed two distinct types of chromatin dynamics. Panel A illustrates the 'classic' example of different association and dissociation reaction that occur in chromatin. Note, that these processes can be under enzymatic control and often require the presence of additional partner molecules such as linker histones. Fundamentally different are metabolic hypercycles introduced by the enzymatic action of coactivators and corepressors. The unmodified nucleosomes to the left in panel B are substrates for coactivator enzymes that modify their histone-tails enzymatically. The products of these reactions, depicted to the right, are themselves substrates for a second, complementary class of enzymes, the corepressors, which reverse the action of the coactivators to reestablish the initial state. These hypercycles are accorded particular attention since they might function as molecular motors to keep certain euchromatic regions actively 'breathing' and therefore allow random access of the nuclear proteome to DNA. Furthermore, they are thought to play a crucial role in the structuring of chromatin (see below).

Only very recently it has been realized that two distinct types of chromatin dynamics exist. While above mentioned association and dissociation reactions are believed to happen, despite their regulation via the availability of non-core nucleosomal proteins and enzymes, close to equilibrium (see Figure 2A), a second class of dynamics can be put forward based on the action of enzymatic activities encoded in coregulator poly-peptides. These coregulators are accessory factors to transcription, repair, replication, and recombination activities and act by specifically modifying the chromatin environment at and around specific sites of DNA activity. The enzymatic modifications carried out by coregulators mostly concern the N- and C-termini of the histone proteins, and consist in acetylation, methylation, phosphorylation, and likely glycosylation and sumorylation reactions. As illustrated in Figure 2B, these coregulator enzymes always occur in pairs, the one reversing the effect of the other. Such pairs thus establish, through acting on two different nucleosomal states metabolic hypercycles, a concept first put forward by Eigen and Eigen and Schuster for cellular metabolic pathways [6, 7]. They are of highly nonlinear nature, and can be best thought of molecular motors that drive nucleosomes far from equilibrium states to facilitate DNA-based nuclear processes.

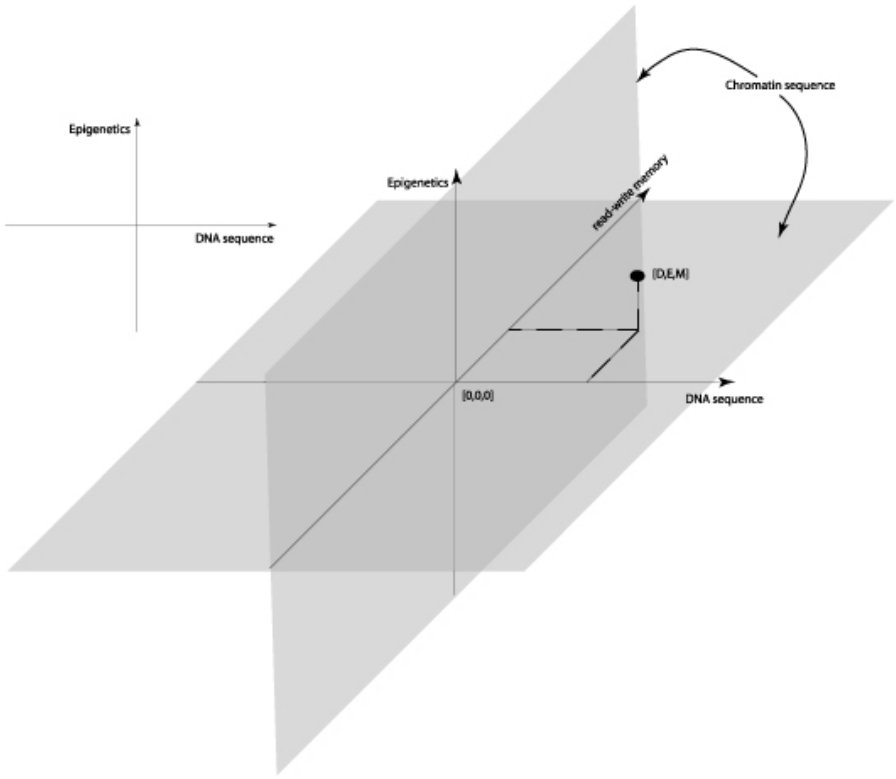


Figure 3: Metabolic hypercycles introduce a third time-scale to nuclear coding.
Through above discussed hypercycles of coactivator and corepressor action a secondary code is being introduced. The enzymatic modifications conveyed upon the individual nucleosome change its state in a manner that is *a priori* independent from the underlying DNA-sequence and epigenetic signals. This change of state can be accounted for by the cell (as illustrated through the action of corepressors, and proteins that specifically interact with modified nucleosomes) and thus constitutes a code. In analogy to computer science this code can be regarded as a read/write memory of an individual nucleus/cell. DNA sequence, the primary code, is being shared between all cells of a multicellular organism, and evolves at time scales which are a multiple of the generation time of the organism. The secondary chromatin code, however, changes on the order of minutes and hours, thus on a similar scale as switching between individual genetic programs in a cell. It is thus anticipated that the chromatin code might play a role in defining the activity of such cellular programs.

Through energy consumption they introduce stochastic chromatin 'breathing' [8].
Interestingly, the occurrence of such hypercycles in chromatin dynamics strongly suggests that chromatin dynamics are a hyperstructure of nuclear organization. Hyperstructures are a very powerful concept in formally describing dynamic cellular processes and was initially developed by Norris and Norris and colleagues [9, 10]. Indeed, we are currently investigating this most interesting aspect [11, 12].

3. Three time-scales of nuclear coding

Coregulator-induced hypercycles between distinct nucleosomal states is a very powerful concept for the understanding of genomic organization and expression in the nucleus. As a matter of fact the different modifications instituted by coregulators - which define distinct nucleosomal states - establish a chromatin code. This code at the time reflects underlying DNA activity, epigenetic regulation, and coregulator activity. Given the reversible nature it can be viewed as a read-write memory function of the genome. Interestingly this concept provides a new time scale in nuclear coding (illustrated in Figure 3). While changing DNA-sequence code happens roughly at specification time scales, and changing epigenetic signaling at generation time scales of an organism, the chromatin sequence code is being modified within minutes to hours, thus on the same time scale as i.e. switching between different cellular genetic programs. It is highly anticipated that the chromatin sequence code indeed is taking part in such a decision making process on the cellular level [12].

4. Chromatin models based on reaction kinetic graphs

In light of above concepts we have started to analyze chromatin dynamics by using reaction-kinetic graphs. Essentially, we have assigned different nucleosomal states (as taking from the published literature and our own experimentation, i.e. [13]) nodes in a graph that were subsequently linked by the enzymatic reaction transforming one nucleosomal state into another (the edges). An example of such a reaction-kinetic based chromatin model is given in Figure 4. These graphs grasp already on a visual level the different concepts of chromatin dynamics (see i.e. the different metabolic hypercycles), and based on their structure help to understand how chromatin dynamics are implemented on a molecular rather than phenomenological level. Due to space constraints the experimental evidence and scientific reasoning behind the graph displayed in Figure 4 can not be detailed here in an adequate manner. A complete description will be found at [14, 15]. Put forward by the nomenclature of the vertices (nodes) of the graph it will be realized for instance that the different nucleosomal states labeled '[0CH----]' and '[1C-----]' correspond to the 30nm and 9nm fibers of chromatin (see Figure 1). The [1C-----] state can be considered the ground-state of euchromatin and is being further modified through a metabolic hypercycle to result in the pre-active state [2C-----]. This state should be considered as the pre-requisite for any DNA-based nuclear process (represented here by '[3C3P--]').

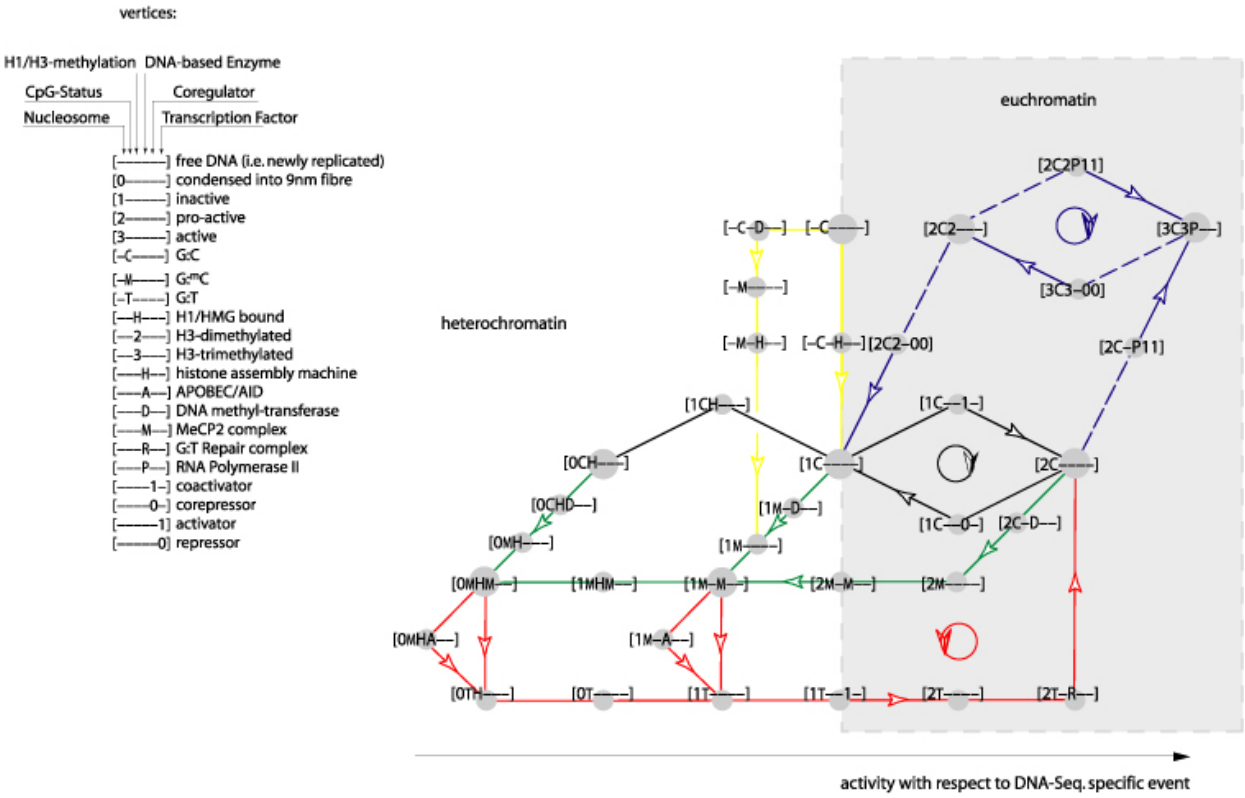


Figure 4: A mathematical chromatin model based on reaction kinetics.

The different known states of chromatin, as well as their modifications through DNA and nucleosomal binding proteins can be represented as large, fully connected reaction kinetic graphs. An example of a primitive chromatin-graph is given. Each node corresponds to a nucleosomal state. Large nodes indicate major stable nucleosomal states, whereas the smaller nodes correspond to transition states. The edges of the graph represent molecular reactions linking the different nucleosomal states. Both bi- and unidirectional edges are observed. A nomenclature for the different nodes is depicted to the right. Interestingly, this simple chromatin-graph contains three different metabolic hypercycles that essentially define the euchromatic compartment. Details on the different experimental observations and underlying assumptions, as well as a description of this graph can be found elsewhere (see text for reference).

5. *arc - Advanced Reaction-kinetic modeling of Chromatin*

While these reaction-kinetic graphs provide clues to the dynamics of chromatin, the underlying first-order ordinary differential equation system can not be solved analytically. We therefore devised a computer algorithm that is specifically tailored to the numerical simulation of such large reaction-kinetic graphs (note that some of the chromatin graphs we are studying contain well over one-hundred nodes; Figure 4 representing just an example of a core-structure). The Figure 5 illustrates and the legend explains the predominant concept underlying the development and functioning of *arc*. Please note the output that is particular suited to analyze graphs with multiple non-trivial equilibrium states. A full description of *arc* is in the process of being published [16], and the algorithm will be made publicly available for non-commercial use pending acceptance of the publication at http://www.ihes.fr/~arndt.
Figure 6 illustrates the type of analysis that we undertake using the *arc* algorithm. A reaction-kinetic sub-graph displaying periodic behavior of several of its components was analyzed in a bifurcation diagram. The order-parameter k is a single reaction rate along one of the edges of the graph. The Feigenbaum diagram displayed illustrates that for all k with 100<k<kc the reaction attains a single stable equilibrium state in finite time. For all kc<k<10E5kc the reaction runs in finite time into a limit cycle (illustrated by the left inlet). The singularity at kc

thus defines two distinct functional states of the graph under investigation. Such analysis obviously helps to comprehend and further refine chromatin graphs such as the one shown in Figure 4. A more detailed description of this bifurcation analysis will be found in [14].

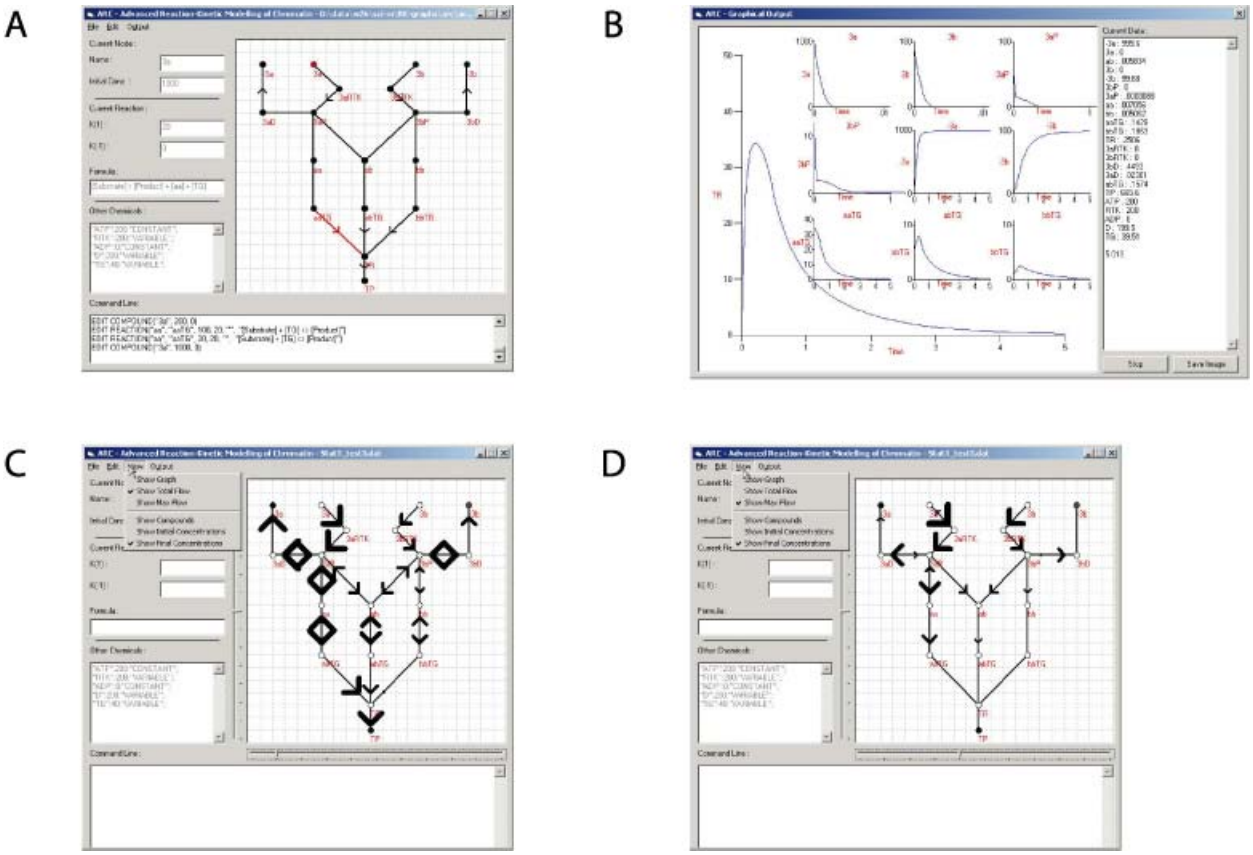


Figure 5: *arc* - an algorithm to study chromatin reaction kinetic graphs.

An algorithm, *arc* for Advanced Reaction-kinetic modeling of Chromatin, has been developed. *arc* posses several particularities compared to conventional reaction chemistry tools. The algorithm allows to associate 'hidden' reactions (such as coregulator - cofactor association reactions that are not part of the nucleosomal graph but need nevertheless to be considered for the numerical enumeration) with the graphs designed in the INPUT window (panel A). This reduces the displayed graphs to their inherent underlying backbone, and thus puts visual emphasis on the core structure. Since our brain is much more powerful in detecting structural modules in such a type of graph than computer programs (for which the graph is by order of magnitudes to small to provide for sufficient information) this visual aid should not be underestimated. Furthermore, *arc* displays output over the entire numerical simulation both in numerical (not shown) and graphical form (panels B-D). Especially, in cases were low complexity steady-states are not attained such outputs aids in the comprehension of the dynamics of the nucleosomal graphs (i.e. regions of high turn-over, panel D). A detailed description of the algorithm will be found elsewhere (c.f. main text for reference).

6. *Perspectives*

The reaction-kinetic graphs of chromatin will be used as basis for building one- and later three-dimensional cellular automata approaches. The chromatin graphs thereby will through the use of renormalization, support vector, and SET analysis techniques provide for the transition probabilities between the different nucleosomal states depending only on a limited set of parameters. This reduction of complexity is paramount to simulation of some 10E7 individual objects/agents (the nucleosomes) in different spatial dimensions. Through the introduction of spatial dimensions a cellular automata approach will allow to incorporate the

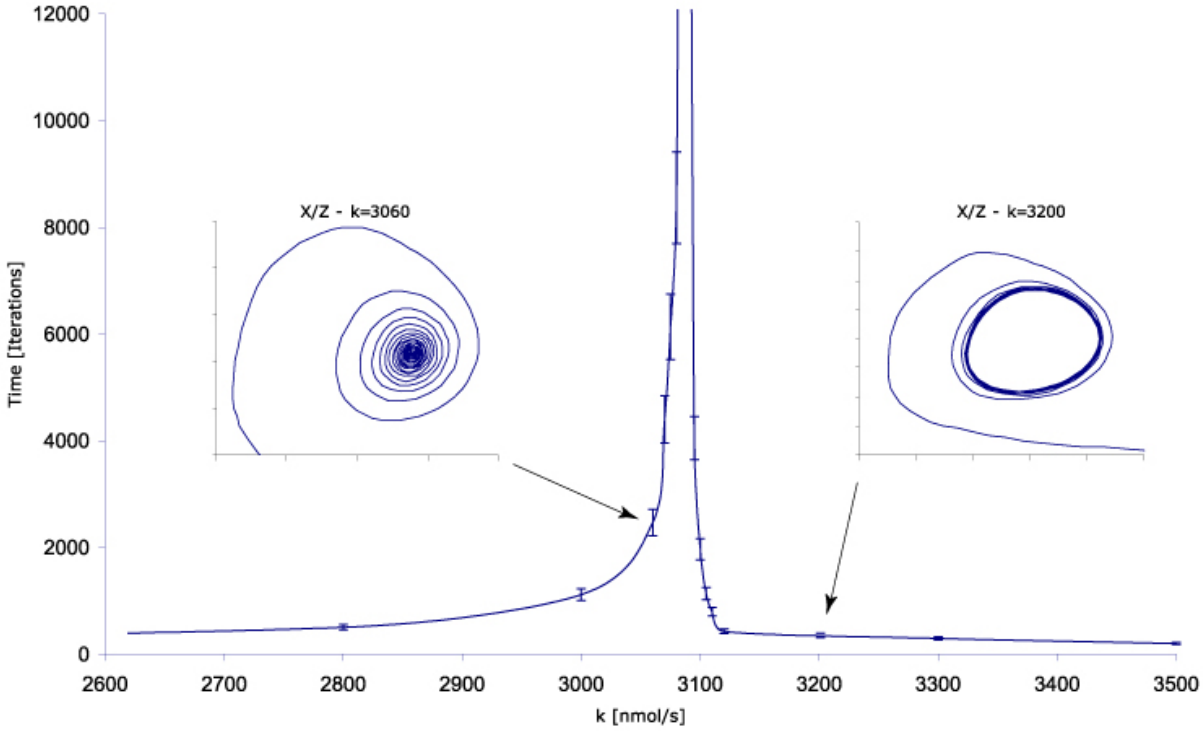


Figure 6: Feigenbaum diagram of a reaction kinetic graph displaying two types of periodic behavior.

An example of reaction-kinetic graph analysis with help of *arc* is given. A simple reaction-kinetic graph displays two forms of periodic behavior depending on the value of a single reaction kinetic constant (k, X-axis). For k<kc the reaction intermediates X and Z show declining periodic fluctuations and their ratio reaches in finite time a single steady state (inlet on the left hand side). If k>kc the reaction remains for infinite time periodic, thus the stable states form a limit-cycle (inlet on the right). The singularity gives us kc, the phase-shift or bifurcation value. Furthermore, the characteristics of the limit cycle are amendable to analysis. In this particular case the excluded area of the limit cycle is unchanged in form over five orders of magnitude.

non-linear thermodynamics of biased diffusion in the nuclear compartment, as well as neighborhood and long-range nucleosomal interaction effects.

We hope to achieve a mathematical model of chromatin amendable to experimental scrutiny and usable to predict effects of chromatin dynamics on nuclear organization and DNA-based nuclear processes.

7. Acknowledgements

The author is indebt to the organizing commitee and especially Victor Norris and Patrick Amar for putting together the exciting workshop "Modélisation et simulation de processus biologiques dans le context de la génomique (May 12-16 in Dieppe, France), and their invitation to present the work summarized here. Furthermore, Victor Norris and François Képès were and are instrumental to the development of the chromatin-hyperstructure concept. Graham Smith is the driving force behind the programming of *arc*, and his ingenuity and outstanding competence is warrant for success of this ambitious project.
Our work is supported by the I.H.E.S., the I.N.S.E.R.M., the Deutsche Forschungsgemeinschaft, and the Fondation de France.

8. References

[1] Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Nature 389:251-60.

[2] Bellard M, Oudet P, Germond JE, Chambon P (1976) Eur J Biochem 70:543-73.

[3] Dietzel S, Nelmont AS (2001) Nat Cell Biol 3:767-70.

[4] Umov FD, Wolffe AP (2001) Oncogene 20:2991-3006.

[5] McKenna NJ, O'Malley BW (2002) Cell 108:465-74.

[6] Eigen M (1971) Naturwissenschaften 58:465-523.

[7] Eigen M, Schuster P (1977) Naturwissenschaften 64:541-65.

[8] Benecke A (2003) Complexus 1:65-76.

[9] Norris V, Fralick J, Danchin A (2000) Mol Microbiology 37:696-702.

[10] Norris V, Demarty M, Raine D, Cabin-Flaman A, Le Sceller L (2001) Biochimie 84:341-7.

[11] Amar P, et al. (2002) Acta Biotheor 50:357-73.

[12] Norris V, Kepes F, Benecke A (2003) in preparation.

[13] Tini M, Benecke A, Um SJ, Torchia J, Evans RM, Chambon P (2002) Mol Cell 9:265-77.

[14] Smith G, Benecke A (2003) Proc SNA submitted.

[15] Benecke A (2003) in preparation.

[16] Smith G, Benecke A (2003b) Bioinformatics submitted.

Modelling and Simulation of Large Assemblies of Proteins

Patrick Amar[1,2], Gilles Bernot[2], Victor Norris[3]

[1]Laboratoire de Recherches en Informatique, Université Paris Sud & CNRS UMR 8623, 15 avenue George Clémenceau, F-91405 Orsay Cedex
[2]La.M.I. Université d'Évry Val d'Essonne & CNRS UMR 8042, Tour Évry 2, 523 Place des terrasses de l'agora, F-91000 Évry
[3]Laboratoire des Processus Intégratifs Cellulaires, CNRS UMR 6037, Faculté des Sciences et Techniques, Université de Rouen, F-76821, Mont-Saint-Aignan Cedex

Author for correspondence: Patrick Amar, email: pa@lri.fr

1. Introduction

To study the assembly, the movements and the dissociation of large numbers of molecules in a virtual cell, a simulation program has been developed. The simulator is driven using a description of the model system written in a language we have also developed that does not limit the simulation program to a particular model. We present here the simulation of two completely different models: the growth of actin-like filaments in a prokaryotic cell, and the association and dissociation of proteins into large assemblies.

The program simulates a virtual cell as a three dimensional space bounded by a spherical membrane. This cell is initially filled with a population of molecules of various types. When the simulation begins, these molecules diffuse and interact according to the reaction rules described in the model. Periodically, the simulator shows the content of the cell with a 3D *OpenGL* user interface. During the simulation, the user can rotate the cell and focus on and zoom a particular area. A histogram showing the distribution of the lengths of the assemblies is also displayed. The program can also display curves showing the number of molecules of each species during the simulation.

2. Simulator description

The simulator is a stochastic automaton driven by reaction rules between molecules. Each molecule is represented by a record that includes its type, its position, a list of links to certain other molecules and other internal data. The simulator keeps track of each assembly in real time from the computer point of view.

A step of simulation (called a *generation*) is done by applying the following process:

- Choose the source molecule *S* (randomly, in order to avoid artefacts).
- Check if close enough to *S*, in a location randomly chosen *L*, there is another molecule *T*, the target.
- If so, and if a reaction rule is given between a molecule of the type of *S* and a molecule of the type of *T*, this rule is applied, according to a probability representing the reaction kinetics.
- Else, molecule *S* may move to the empty location *L*, according to a probability representing the diffusion speed.

When all the molecules involved in the cell have been processed, the current generation is completed and a new one can begin. The generation simulated time slice is set to 100 micro-seconds, which corresponds to the average time for a protein to move a distance of 10 nanometres (approximately its diameter).

To keep memory usage low, the simulated space is managed with a technique called a *hash table*. With this technique the computer memory used is proportional to the number of molecules in the simulation (and not to the size of the simulated space). The extra computer time used by the hash table is less than twice the time used to access a standard array. The proportion of this access time to the total time being less than one quarter, this extra cost is acceptable compared with the possibility of having a potentially infinite simulated space.

Another way to reduce the computer time used in this programme is to maintain two lists of molecules. The first one is the list of the *active* molecules, which can be *sources* of interaction. The second one is the list of *inactive* molecules, which can only be target of interaction. The simulator processes only the list of active molecules which can be much smaller than the total number of molecules. During the processing of a couple of molecules, the programme automatically changes the status of these molecules. In the actin example shown hereafter, the main constituent of an actin strand is the F-actin molecule, which is inactive (unless near the ends of the filament).

2.1. The rules

The simulator implements four kinds of interaction rules between two molecules: the source *S* and the target *T*:

- **Reaction**: *S* reacts with *T* to produce two other types of molecules *S′* and *T′*.
- **Association**: *S* binds to *T* to produce the complex *S*-*T*.
- **Dissociation**: the complex *S*-*T* can break and leave individual molecules *S* and *T*.
- **Catalysis**: the complex *S*-*T* can be transformed to *S′*-*T′*.

As in the reaction and catalysis rules, the association and dissociation rules can change the type of the molecules. Each rule is given a probability of execution that, on the long run, corresponds to a reaction kinetics. For the association rule, a maximum number of links can be specified.

2.2. Configuration

The simulator uses a configuration file to describe the model the user wants to simulate. This file contains four sections. The first section describes the molecules involved in the model. The second specifies the diffusion rate of each molecule. The third section describes all the reaction rules that will be applied during the simulation. The last section describes the initial population and location for each kind of molecules. Here is an example of the types of molecules involved in the simulation of the growth of actin filaments:

```
molecule
      P (200, 0, 0),                    // filamentous actin 'plus' end.
      M (0, 200, 0),                    // filamentous actin 'minus' end.
      AF (200, 200, 0, inactive),       // inactive F-actin.
      AG (200, 100, 0);                 // phosphorylated globular actin.
```

With these definitions, the *plus* end (**P**) will be displayed in red, the *minus* end (**M**) in green, the filament itself (**AF**) in yellow and the free globular actin (**AG**) in orange.

In this example only the free G-actin molecules can diffuse. The filaments themselves are frozen:

```
speed (AF) = 0.0;       // diffusion speed is zero
speed (P) = 0.0;        // for the filaments.
speed (M) = 0.0;
speed (AG) = 1.0;       // high diffusion speed.
```

The following rules show the formation of the polarised dimers from two free phosphorylated free G-actin molecules:

```
AG  +  AG  ->   M(1) * P(1)      [0.05];
M   *  P   ->   AG   + AG        [0.5];
```

The right part of the first rule states that a minus end **M** can be bound to only one plus end **P**, and conversely, a plus end **P** can be bound to only one minus end **M**. The second rule is the reverse reaction, the depolymerisation of the complex, giving two free G-actin molecules.
The next rules show the growth of a filament from the plus end:

```
AG  +  P  ->   P(1)  =  AF(1)    [0.8];
P   *  AF ->   AG    +  P        [0.001];
```

The first rule shows how a free G-actin molecule can be bound to the plus end of an already existing filament. The equals sign in the right part of the rule means that the link must be aligned with the filament. The second rule shows the reverse reaction, the depolymerisation from the plus end giving one free G-actin molecule.

The next two rules show the growth of a filament from the minus end. One can notice that the polymerisation kinetics is lower than for the plus end which may lead to a linear movement of the filament towards the plus end equivalent to tread-milling.

```
AG  +  M   ->   M(1)  =  AF(1)   [0.3];
M   *  AF  ->   AG    +  M       [0.005];
```

2.3. Initialisation

The simulation is initialised by the statement:

```
cube (0, 0, 0, 12, AG);
```

A cube of length *12* located at the centre of the cell *(0, 0, 0)* is filled with $12^3 = 1728$ molecules of free globular actin. This compact cube leads to a very high local concentration of G-actin. To avoid this generating artefacts, the simulator is first started in *diffusion only* mode for a few thousand generations. Then it is switched back to *reaction* mode so the molecules diffuse and react according to the rules. First, polarised dimers assemble and then the filaments grow until one end touches the membrane or an equilibrium state is reached (see Fig. 1).

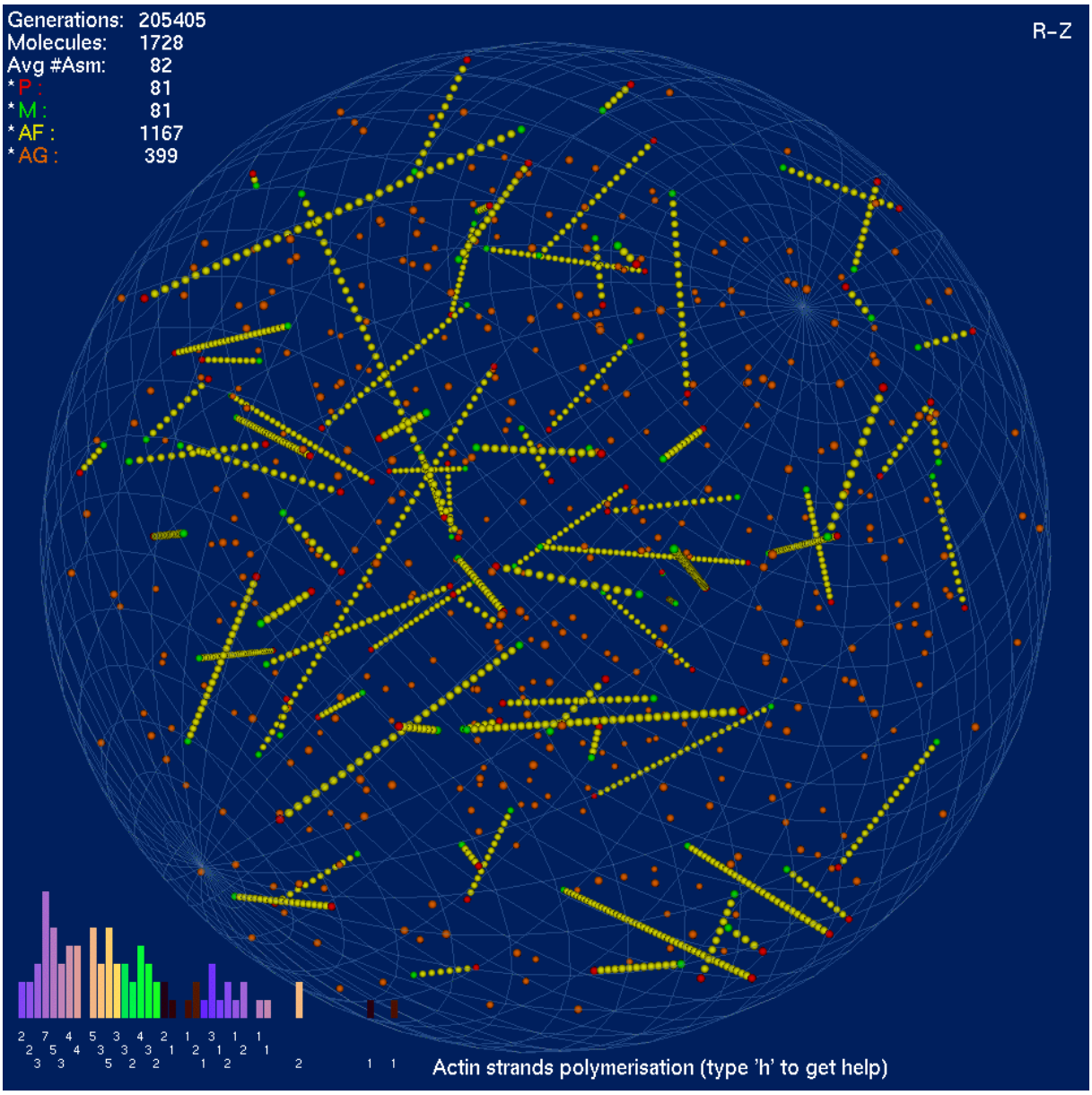


Figure 1: A view of the virtual cell filled with dynamic actin filaments. On the bottom left corner of the screenshot a histogram of the lengths of the filaments is displayed (see plate 2).

3. Simulation of hyperstructures

In addition to the rules mentioned above, the simulation language has some specific features that allow the user to study a large number of different model systems. Hyperstructures are large assemblies of molecules such as enzymes within cells. In this section we show how to model a hyperstructure in the form of a metabolic pathway in which the product of one enzyme is the substrate of the next one in the pathway. The simulation shows how hyperstructures can, on demand, assemble, work and disassemble when the simulation only specifies an increase of affinity between two enzymes in the presence of their substrate. Finally, we show how to confine an object to the cell membrane so allowing the creation of membrane receptors which can only diffuse in two dimensions in the membrane (see Fig. 2).

As in the previous example, the initial state of the simulation is obtained by switching the simulator to a *diffusion only* mode and then running it for a few thousand generations.

This disperses the cytoplasmic enzymes throughout the cytoplasm. As the diffusion speed of the membrane receptors (Enzyme 1) is very slow, they stay in roughly the same place in the membrane during this first phase of simulation. The main simulation phase begins when the simulator is switched to the *reaction* mode.

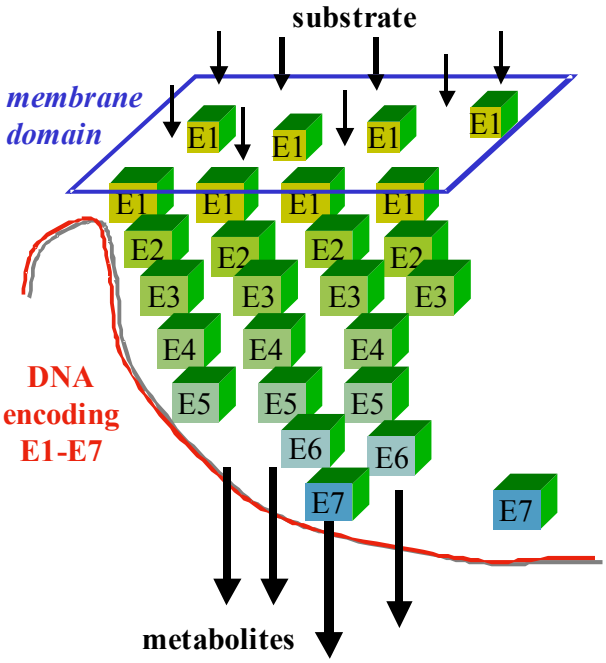


Figure 2: Formation of a non-equilibrium hyperstructure due to changes in the affinity of its constituent enzymes for one another. Enzymes E1 can only diffuse in the plane of the membrane whilst the other enzymes, E2 to E7 diffuse in the cytoplasm. The binding of a substrate, such as a sugar, to the E1 enzymes leads to an increase their affinity for one another and their assembly into an E1 domain. On binding its substrate, each enzyme in the pathway acquires an increased affinity for the following enzyme. This results in the assembly of metabolons E1 to E7 and the assembly of the hyperstructure (here, a group of metabolons). Note that transcription of the genes encoding E1 to E7 and the simultaneous translation of the mRNA may help the assembly of the hyperstructure.

With the current implementation of the programme it is not yet possible to have molecules outside the cell membrane. To simulate the fact that the initial substrate is translocated by the membrane receptor, we put this substrate directly into the cytoplasm. To keep this example as simple as possible we only use one kind of membrane receptor and a pathway of four enzymes.

3.1. Configuration

In the first section of the configuration file we declare all the molecules involved in this simulation: the membrane receptor $\mathbf{E_1}$, the enzymes $\mathbf{E_2}$ to $\mathbf{E_4}$ and the substrates $\mathbf{S_1}$ to $\mathbf{S_5}$. The *membrane* keyword declares molecules confined to the cell membrane:

```
membrane   E1 (0, 0, 250);
```

The other enzymes and the substrates are declared using the *molecule* keyword as in the previous example. The reaction rules are divided into four almost identical groups, one for each enzyme in the pathway.

The group of rules for enzyme E_n indicates that:

- enzyme E_n can fix its substrate S_n
- when E_n has bound its substrate, E_n has its affinity for enzyme E_{n+1} increased so it can be bound by it
- conversely, when enzyme E_n does not have its substrate S_n bound to it, E_n loses its affinity for E_{n+1}
- enzyme E_n transforms its substrate S_n to its product S_{n+1} , which is the substrate for enzyme E_{n+1}
- when enzymes E_n and E_{n+1} are bound together, the product S_{n+1} is transferred to E_{n+1} (and this new product is then freed from E_n and linked to E_{n+1})

Here are the rules for the beginning of the pathway, the membrane receptor E_1:

```
S1 + E1          -> S1(1) * E1(1)      [0.6];
```

This rule means that when the substrate S_1 is close enough to the membrane receptor E_1, it is captured with probability 0.6. The number between the parentheses in the right part of the rule means that the enzyme can bind only one copy of its substrate.

```
{S1}E1 + E2      -> E1 (1) * E2 (1)    [0.9];
```

This second rule shows how the *environment* of a molecule can be important in the application of a rule. The left part of the rule means that if enzyme E_2 is close enough to a membrane receptor E_1 that is already bound to its substrate S_1 (and only if E_1 is bound), then enzyme E_2 will bind to E_1 to form a complex. This is done with a high probability to model the high affinity for the two enzymes in presence of the substrate.

```
{~S1}E1 * E2     -> E1 + E2            [0.001];
```

Finally the third rule shows how enzyme E_1 loses its affinity for enzyme E_2 when E_1 is not bound to its substrate. The left part of the rule means that if E_1 is bound to E_2 but E_1 is not bound to its substrate S_1, the link between the two enzymes is broken. One can notice that the probability is very low; this is because the two molecules being linked are very close and the program gives a greater chance for this event to occur.

The two last rules of the group are used for transform S_1 to S_2 and to transfer it from enzyme E_1 to enzyme E_2.

```
{E1}S1 + {E1}E2 -> S2(1) * E2(1)      [1.0];
```

here S_1 which is bound to E_1 also binds to E_2 and at the same time is transformed to the product S_2. Each time the initial conditions are set (the left part of the rule) the rule will be statistically applied because the probability is set to one.

```
{E2}E1 * {E2}S2 -> E1 + S2            [1.0];
```

this rule is used to break the link between the product S_2 and the enzyme E_1 to complete the transfer.

With this set of five rules repeated three times (for each of the different enzymes E_2, E_3 and E_4) in the configuration file, plus a last rule to release the final product S_5 we have finished with the rule section. The initialisation section fills the cell with 36 copies of the membrane receptor, 64 copies of enzymes E_2, E_3 and E_4, and 729 copies of substrate S_1.

```
surface (E1);
cube (0, 6, 8, 4, E2);
cube (6, 0, 8, 4, E3);
cube (0, 0, 0, 4, E4);
cube (0, 0, -6, 9, S1);
```

3.2. Simulation results

After the first phase of diffusion to get a homogenous distribution of all the molecules in the cytoplasm, the membrane receptors bind their substrate. Then after a short period of time, we can see the first assemblies appear and quickly transform the intermediate substrates to the final product like an assembly line in a factory (see Fig. 3).

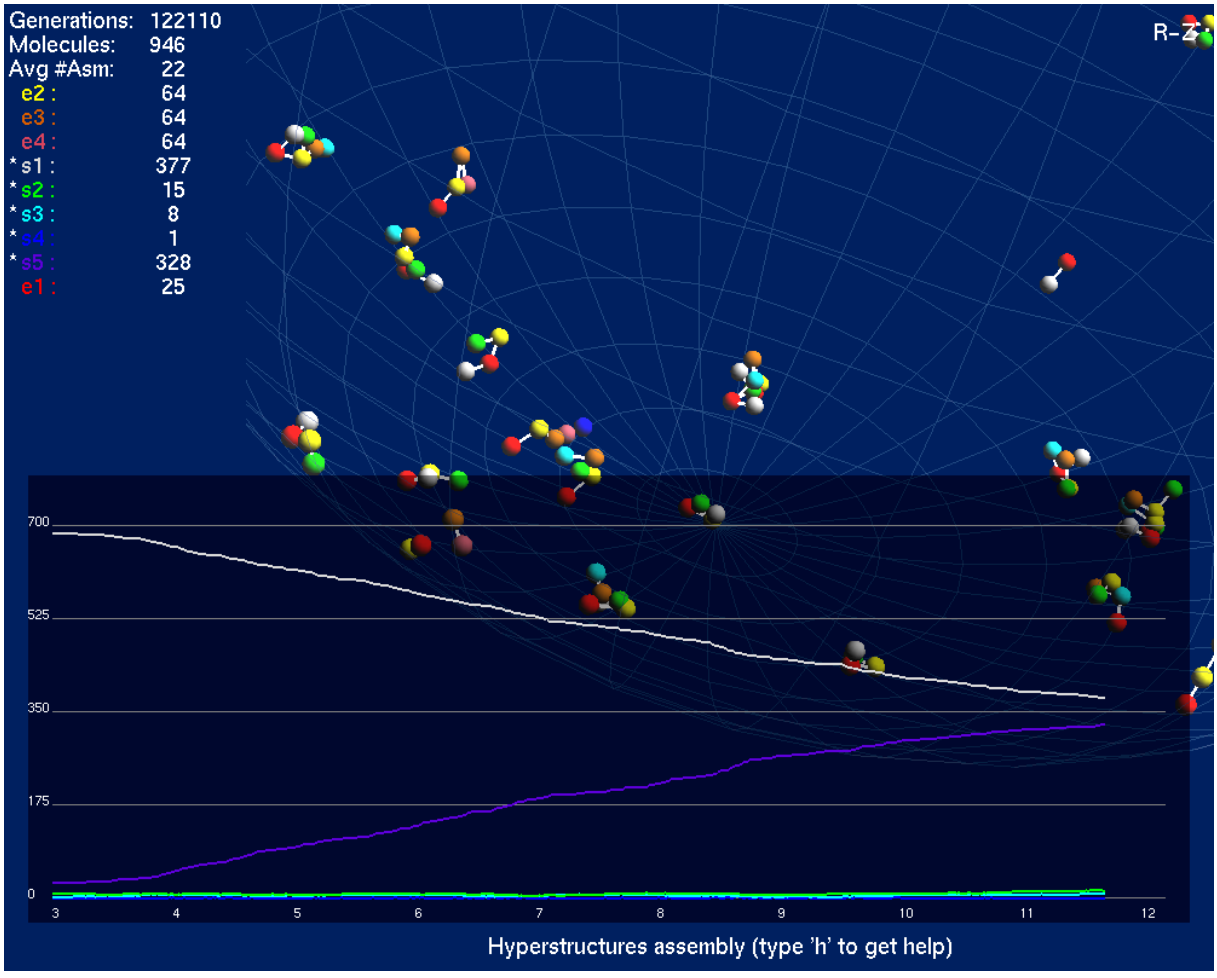


Figure 3: The virtual cell with the hyperstructures linked to the membrane receptors (in red). The curves in the bottom part show the decrease in the concentration of substrate S_1 along with the increase in concentration of product S_5. The horizontal axis is graduated in seconds of simulated time (the real time is approximately 3 times slower on a standard PC). The vertical axis shows the number of copies of each kind of molecules (see plate 3).

Before all the copies of substrate S_1 have been transformed to the final product S_5, the assemblies begin to break up and finally disappear. Since these assemblies are attached to the membrane even if the total concentration of enzymes is low, the local concentration is high enough to produce S_5 at high rate.

4. Conclusion

With these two examples one can see that this simulation programme is very versatile. The efficiency of the implementation in terms of computer time is high enough to include in a future release some real time controls. These controls may include the ability to modify the reaction kinetics (the probability part of the rules) or the number of copies of each kind of molecules, etc.

Another improvement, which is in progress, is to replace the spherical membrane of the cell with a simulated membrane made with lipid molecules. The number of molecules used to make the membrane is very high, but they can be *inactive*, and so they do not use computer time. The next step is to allow deformations of this membrane, because of the pressure of the actin filaments for example.

References

Amar, P., P. Ballet, G. Barlovatz-Meimon, A. Benecke, G. Bernot, Y. Bouligand, P. Bourgine, F. Delaplace, J.-M. Delosme, M. Demarty, I. Fishov, J. Fourmentin-Guilbert, J. Fralick, J.-L. Giavitto, B. Gleyse, C. Godin, R. Incitti, F. Képès, C. Lange, L. Le Sceller, C. Loutellier, O. Michel, F. Molina, C. Monnier, R. Natowicz, V. Norris, N. Orange, H. Pollard, D. Raine, C. Ripoll, J. Rouviere-Yaniv, M. Saier jnr., P. Soler, P. Tambourin, M. Thellier, P. Tracqui, D. Ussery, J.-P. Vannier, J.-C. Vincent, P. Wiggins and A. Zemirline.
Hyperstructures, genome analysis and I-cell. Acta Biotheoretica **50**: 357-373, 2002

Borisy, G.G. and T.M. Svitkina.
Actin machinery:push-pull the envelope, Curr. Opin. Cell Biol. **12**:104-112, 2000.

Geigant E., Ladizhansky K. and A. Mogilner.
An integrodifferential model for orientational distribution of F-actin in cells. SIAM J. Appl. Math. **59**:787-809, 1998.

Kier LB, Cheng CK, Seybold PG.
Cellular automata models of chemical systems. SAR QSAR Environ Res. 2000;11(2):79-102.

Kier LB, Cheng CK, Testa B, Carrupt PA.
A cellular automata model of enzyme kinetics. J Mol Graph. 1996 Aug;14(4):227-31, 226.

Le Sceller L., C. Ripoll, M. Demarty, A. Cabin-Flaman, T. Nyström, M. Saier Jnr. and V. Norris (2000).
Modelling bacterial hyperstructures with cellular automata. Interjournal Paper 366: http://www.interjournal.org.

Wurthner JU, Mukhopadhyay AK, Peimann CJ.
A cellular automaton model of cellular signal transduction.Comput Biol Med. 2000 Jan;30(1):1-21.

Modelling, observability and experiment: a case study
- Positive feedback loop in a genetic regulatory network -

Gilles Bernot[1], Janine Guespin-Michel[2], Jean-Paul Comet[1], Patrick Amar[1,3], Abdallah Zemirline[4], Frank Delaplace[1], Pascal Ballet[4], Adrien Richard[1]

[1]Laboratoire de Méthodes Informatiques, Université d'Évry Val d'Essonne & CNRS UMR 8042, Tour Évry 2, 523 Place des terrasses de l'agora, F-91000 Évry
[2]Laboratoire des Processus Intégratifs Cellulaires, Université de Rouen & UPRESA CNRS 6037, F-76821 Mont-Saint-Aignan Cedex
[3]Laboratoire de Recherche en Informatique, Université Paris Sud & CNRS UMR 8623, 15 avenue George Clémenceau, F-91405 Orsay Cedex
[4]Departement Informatique, EA 2215, Université de Bretagne Occidentale, BP 809, F-29285 Brest Cedex

Abstract

We propose an new methodology for modelling of biological regulatory networks inspired by the design and validation of large computing systems. We take into account the capability to validate a model by a set of biological experiments. So defining a model goes with experimental methods and conditions to validate or invalidate it. As in the design of large sized softwares, we will distinguish two activities: first to build an accurate model specifying the assumed behaviour, second to design plans of experiments to verify *a posteriori* the model predictions. We wish to experiment, through the case of the modelling of the mucus production by the bacterium *Pseudomonas aeruginosa*, the application of this working methodology.

1 Introduction

Biologists put a large number of meanings in the term "model". Even when precised as "mathematical model" we are far from the unicity of meaning. One of the difficulties comes from the interdisciplinarity already contained in the expression "a mathematical model in biology". Who makes the model, who is using it, and first, what is it utility? Although some biological scientific communities already use modelling routinely, other are still very reluctant to that usage, and molecular biologists most often think modelling useless for their field because it does not contribute to experimental knowledge. This may involve a large range of reactions, from violent reject to polite interest This leads often model makers to use data from the literature, which may lead to very interesting models, but without enough feedback from and toward experimentation.

Interdisciplinarity needs not only learning how to work together, but also the common design of usable tools. It demands moreover that each contributor finds a scientific interest working together, in other words, that the collaboration will benefit to both disciplines. This is where analogy between computing systems specification and modelling in biology is involved. In computer science, the design of systems requires to:

- specify, i.e. build a rigorous model of the desired behaviour of the future computing system;
- verify *in fine* if a system corresponds to its specification, i.e. to the desired behaviour as described by the theoretical model previously built.

This last activity is mainly based on sophisticated software testing methods *via* test generation from model theories. The goal is then to propose a set of experimentations on the delivered software which is sufficient to establish, by extrapolations, that the software under test will have a behaviour compatible with its model.

Within this framework, the notions of operability and observability constitute a major issue:

- the *operability* is the capability to make a program run some chosen pieces of its internal code (in order to test them), sometime activated in rare, complex or very specific configurations. It is also the capability to make a program modify the value of variables hidden in the very large set of data managed by the program. These actions have to be done by only using the limited user interface of the program.
- the *observability* is the capability to make the effects produced by the previous manipulations visible, in order to verify their correctness according to the desired model of behaviour.

Some models or softwares are not testable, either because of a lack of operability or a lack of observability. A necessary step to design a software is to know if a model can be validated by a reasonnable sized set of tests (experiments). The reader can easily transcribe this argumentation to the case of biological modelling:

- *operability:* what would be the utility of a too much detailed model of some biological entity if no biological experiment of those details can be done?
- *observability:* what would be the utility of an experiment which cannot let us observe a revealing behaviour?

Some mathematical models for biology are not very helpful because of a lack of operability or observability. A necessary first step to propose a model for biology is to know if it can be validated by a set of biological experiments at a reasonnable cost.

Hence a good model sould be delivered with a set of experimental methods/conditions able to validated or invalidate it. By using the same kind of theories developed in computer science for the validation and the verification of softwares, we wish to experiment, through the case studied here, a new interdisciplinary working method for the modelling in biology.

In the case we will study here a first modelling step (already published), based on the multistationnarity theory, makes an innovating hypothesis plausible (section 2). A second mathematical step, based on formal logics in computer science, allowed us to determine the biological experiments sufficient to validate or invalidate the hypothesis (section 3).

Far further this biological example we have created a new software environment which integrate sophisticated algorithms from computer science (`SMbioNet`). This software environment allows in the one hand to better commit biologists in the modelling process by giving them a model validation tool, and in the other hand, to increase the scope of existent tools in computer science.

Of course such an approach needs a tight collaboration between biologists and model makers. The models designed this way, are not only an attempt for *a posteriori* explanation of results from biology, but a guide for biological experimentation, which will be *in fine* the determining criterion.

2 The chosen case

The biological system chosen is the production of mucus (alginate) by the bacterium *Pseudomonas aeruginosa*. Bacteria of this specie do not generally produce this mucus if they have not experienced a sojourn inside the lungs of patients suffering from cystis fibrosis (production which is the main cause of lethality in this disease). Not only do these bacteria produce alginate in the patients' lungs, but they continue doing so, more or less stably, once extracted from these lungs and cultivated in the laboratory. It is generally admitted that mutations arising inside these lungs cause the ability of these bacteria to produce this mucus in other conditions [1].

But another hypothesis has been put forward, according to which the ability to produce or not alginate are two stable states that arise from each other by an epigenetic modification, prior to the selection of mutants [2].

A very simplified model of the regulatory network has been constructed [2] as depicted in Figure 1. The 3 variables are x for the AlgU protein, y for the AlgU inhibitors, and z for the alginate production. The 4 arcs[1] represent: the self-regulation of variable x (arc $x \to x$), the transcription of the genes encoding the anti-AlgU (arc $y \to x$), the transcription of the genes involved in alginate production (arc $x \to z$)), and finally the inhibition of AlgU by the anti-AlgU (arc $x \to y$). Two feedback circuits control AlgU, a positive feedback loop at the transcriptional level, and a negative feedback circuit involving the activity of the AlgU protein. It has been observed that in the mutant bacteria y is removed from the corresponding graph. The negative action of y on x disappears.

The extreme simplification of this model is directly related to the theory that supports it, which stipulates that feedback circuits are the only elements that are determinants for the emergence of epigenesis [3]. It is proved in this framework that the other known regulatory interactions are of minor importance with regard to the question of the existence of an epigenetic modification.

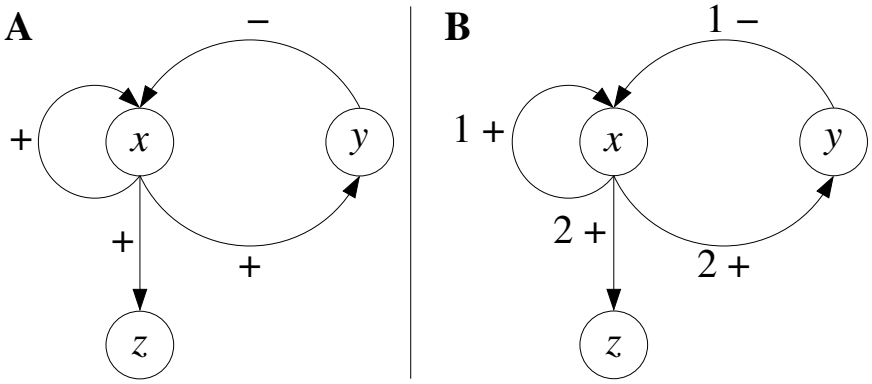


Figure 1: Genetic network regulating the production of mucus (alginate) by the bacterium *Pseudomonas aeruginosa* (simplified model). **A** shows the activation ($+$) or the inhibition ($-$) between the genes (resp. proteins) of the network. In **B** each arc is labelled with the minimal value of the threshold for which this action (activation or inhibition) is triggered. The variables means: x AlgU protein, y anti-AlgU, z mucus synthesis.

This model can be studied by a system of differential equations or by generalised logical analysis [4]. The generalised logical approach has been proved to be a correct approximation of the differential approach. To summarise, when variable x interacts with variable y, the curve that represents the level of y as a function of the level of x is a sigmoid. This sigmoid defines a threshold $S_{(x,y)}$ (Figure 2-a). Similarly the influence of x on another variable z defines another threshold $S_{(x,z)}$ (Figure 2-b). The two thresholds are generally different and lead to three different possible behaviours of variable x depending whether it is below both thresholds, between them or above them (Figure 2-c). Thus it is possible to ascribe discrete values to the different levels of variable x. Then the thresholds correspond to interaction values between the variables. In order to describe that AlgU must be present above threshold 2 to trigger the expression of the alginate genes, it will be noted in the graph $x \xrightarrow{2+} z$ (Figure 1-B). In order to describe the dynamic evolution of the system, we have to specify the level reached by a variable v as a function of the levels of the other variables that influence it (Figure 1). These values are represented by function K.

[1] the arrows in Figure 1

In our model, two feedback circuits co-exist. The functionnality of the positive feedback loop $x \rightarrow x$ is a necessary condition for the existence of two stable states: if x is high, it is self-maintained, else, if it is below the first threshold, it remains so. The negative feedback circuit may attract the system toward one or the other of these stable states depending on its strength (i.e depending on function K).

A mathematical study of this model [2] has shown that the epigenetic hypothesis (the possibility that two stable states may exist depending on the previous history of the system) is coherent and that biologically consistent values of Ks can lead to properties that are precisely those of the system. But there is more to it. For instance it can be predicted that, if the hypothesis is true, a pulse of AlgU will suffice to switch the bacteria to a mucoid state. The model is thus predictive as well as explanatory.

An interesting question is: if such an experiment succeeds, if a pulse of AlgU is able to induce mucus production, or at least induce the expression of the first genes involved in mucus production, will this be sufficient to prove the underlying hypothesis of epigenetic modification? Conversely, if this experiment fails, will it prove the unreliability of the epigenetic hypothesis in this case?

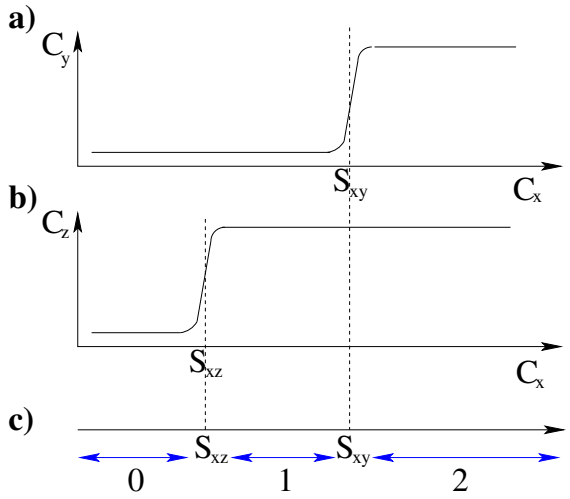


Figure 2: Influence of the variable x on the variables y and z

3 Formal logic to propose experiments

In this section, we outline the general methodology through the example of the production of mucus in the bacterium *Pseudomonas aeruginosa* described in section 2. Although the mathematical proofs are informally presented, they can all be formally performed on a computer. Indeed a model is used to establish properties on a system, to express and handle these properties in order to extract some non trivial other ones. It is thus necessary to formalise the properties in such a manner that they are easy to handle by a computer. The objective laid down here relates to the generation of scenarii of experiments, in which time plays a central part. This leads naturally to temporal logics (see [5], [6] for a general description of temporal logics).

For a given graph, there is a large number of models depending on the values of the tresholds and on all the possible functions K. Each one of these models defines a specific temporal behaviour. A temporal logic formula expresses a property which can be used to split the set of models into two parts: the models which satisfy the property and those which do not. Formalising biological knowledge into temporal formulae will allow us to consider only sensible models.

Our software environment `SMbioNet` allows the user to draw the regulatory network as in Figure 1, assign the thresholds when they are known, and then generate all the compatible models. `SMbioNet` allows the user to enter temporal logic formulae to keep only the models which satisfy them.

In our case study the epigenetic hypothesis means to prove that *in the presence of* y in the graph it is possible to have a recurrent state in which mucus is produced. By construction of the graph for the wild bacterium, y is present and the topology of the graph has been biologically validated as well as the signs of interactions. We have entered this graph in `SMbioNet` which has generated more than 600 models each one with its own behaviour corresponding to different thresholds and function K.

The epigenetic hypothesis is *consistent* if and only if there exist at least one model such that the behaviour can reach a state where z (mucus production) is expressed in a recurrent way. The only work we have to do is to write a temporal logic formula which expresses the recurrent expression of z. By giving this formula to `SMbioNet` we will automatically know how many models satisfy the formula.

The language of temporal logics offers the traditional connectors such as for example, the "or", noted $\vee$, the "and", noted $\wedge$, the implication noted $\Longrightarrow$. It also offers modalities particular to this type of logic related to time. We can for example create the modality $\mathcal{F}_s$, which means that the formula which follows the modality is true in the "strict future". We call here "strict future" the future starting after a certain amount of time (i.e. excluding the current state). This amount of time must be chosen according to biological considerations on the studied system.

We want to verify if there is a model such that, if at a given time the bacterium is in a mucous state, then later (in a strict future) it will be again in a mucous state. This can be formalised as:

$$(z = 1) \Longrightarrow \mathcal{F}_s(z = 1)$$

Indeed we know enough about the graph structure to deduce that $z = 1$ is equivalent to $x = 2$. We know experimentally that the threshold associated with the interaction $x \to z$ is the maximal value of x (equal to 2), and we know by construction that the threshold associated to $y \to x$ is 1 (y has influence only on x, therefore there is only one threshold for y). On the other hand we do not know the thresholds of the arcs $x \to x$ and $x \to y$. In other words, we do not know the relative quantities of the variable x necessary to obtain a self-induction effect, an effect on y, or a combined effect. So, the epigenetic hypothesis means that the relative forces of these two circuits are such that it is possible to make recurrent the state $(x = 2)$. Consequently the epigenetic hypothesis can be written as:

$$(x = 2) \Longrightarrow \mathcal{F}_s(z = 1)$$

Note that this formula expresses that $z = 1$ in a recurrent way because when $z = 1$ at a given time, $x = 2$ at the same time, which in turn implies that $z = 1$ will be true again in the future, and so on.

`SMbioNet` has shown that several models satisfy this formula, which means that the epigenetic hypothesis is consistent.

$\Rightarrow$	0	1
0	1	1
1	0	1

Table 1: Truth table of $p \Rightarrow q$

The consistency of the epigenetic hypothesis being established, it remains to find experiments which prove it *in vivo*. As we can see on table 1, when the left part of the formula $(x = 2)$ is false,

the whole formula is true regardless the value of the right part. So the only revealing experiments always start by assigning (artificially if necessary) x to 2. The scenario of experiments is thus the following:

1. Start by imposing ($x = 2$).
2. Wait a lapse of time (of course the length of it remains empirical) then stop imposing (x =2) and test the phenotype for as many subsequent generations as possible.
    - If the bacterium has not changed it's phenotype, then the experiment *a priori* fails.
    - If the bacterium has become mucoid ($z = 1$) and remains so for several generations after the external signal has been removed, then, epigenesis is proved.

Operability and observability

The next question is then, is this prediction amenable to experimentation, is it both operable and observable? In other words, is it possible to raise x up to 2, then quit the conditions that have allowed this, and observe the production of mucus in the "strict future"?

Indeed, there are several ways to increase x without introducing the bacteria inside the lungs of a cystic fibrotic patient. The most rigorous one would be to introduce into wild type cells of *Pseudomonas aeruginosa*, a plasmid carrying gene *algU* under the control of an artificially inducible promoter. A short pulse of expression of this gene would lead to an artificial increase of the amount of protein AlgU inside the cell.

To observe the results of this experiment, again several experimental devices are currently available, either by measuring the mucus produced, or, more easily by measuring expression of the first gene of the alginate biosynthesis chain (gene *algD*).

Limits of the approach

The graph on which the models are based (Figure 1) is actually only a subgraph of a more general graph showing all the variables of the organism. So it would be necessary to consider all the interactions with the neglected part of the general graph. Having neglected the outgoing arcs of the graph does not have any consequences since we are only interested in the subsystem involving the production of mucus. On the other hand, having neglected the arcs entering this subsystem can have an important impact. By construction of the graph, some situations can be eliminated.

1. By definition of z, the only arc controlling z is the one we take into account: $x \to z$, and thus it does not exist any other entering arc on z.
2. All the arcs which were not considered in the model but which control x or y are not involved in a circuit. The number of steady states does not change [7].
3. If there are arcs entering on x and y whose influence does not vary, the only consequence of having extracted a subgraph is to shift the various thresholds associated with the variables x and y. The system will have other values for the thresholds and possibly for the function K, but the variables will always be discretised the same way. Thus, the satisfiability of the formula remains the same.

Only one case remains awkward: when regulators external to this subgraph (on x and y) have an influence which varies in time. The study presented here makes the assumption that these influences are negligible. This work remains valid under the assumption that a merge of the subgraph into the global graph have constant influence on the variables x and y.
Lastly, let us recall that the amount of time mentioned above between step 1 and step 2 of the experiment remains empirical.

4 Conclusion

The interdisciplinary work undertaken by our *Observability* working group in Genopole® gives a methodological framework to define models including a tool kit for experimental validation or refutation. This way, our work resolutely enforces the modelling activity. It increases its credibility in biology. Indeed, following a Popperian approach, this methodology offers the opportunity to strongly and properly link the modelling activity and the experimental activity, which is central in biology.

Establishing such an approach requires a theory which fixes the rules allowing to reason from a model. The theory introduced here is *temporal logic*, usually employed for the logical analysis of the discrete dynamic systems in computer science. According to this theory, our case study proves that a discrete qualitative model of gene expression based on the work of René Thomas fulfills the methodological requirement mentioned. It makes it possible to determine, in a computer aided manner, a protocol of experimentation to prove or refute the epigenetic assumption described by this model (section 3).

Because our approach is inspired by the software engineering testing methods, this suggests that we can automate it. SMbioNet is a software assistant for the design of biological models which fully handles temporal logic. More than this, this approach can be used to generate and optimise biological experiment scenarii. We have shown the feasibility of the approach on the *Pseudomonas aeruginosa* example. Extending SMbioNet to also suggest experiment scenarii requires to continue our investigations on the application of formal methods from computer science to life science.

Acknowledgements

We thank Christelle Koundibia, Catherine Meignen and Hélène Pollard for organisational help and Genopole Recherche for supporting our interdisciplinary work groups.

References

[1] J. Govan and V. Deretic, "Microbial pathogenesis in cystic fibrosis: Mucoid pseudomonas aeruginosa and burkholderia cepacia," *Microbiol Rev*, vol. 60, pp. 539–574, 1996.

[2] J. Guespin-Michel and M. Kaufman, "Positive feedback circuits and adaptive regulations in bacteria," *Acta biotheoretica*, vol. 49, pp. 207–218, 2001.

[3] R. Thomas, "On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations," *Springer Series in Synergies*, vol. 9, pp. 180–193, 1980.

[4] R. Thomas and M. Kaufman, "Multistationarity, the basis of cell differentiation and memory. ii. logical analysis of regulatory networks in terms of feedback circuits," *Chaos*, vol. 11, pp. 3375–3382, 2001.

[5] M. R. Huth and M. D. Ryan, *Logic in Computer Science: Modelling and Reasoning about Systems.* Cambridge University Press, 2000.

[6] R. Lalement, *Logique, réduction, résolution*. Masson, 1990.

[7] E. Snoussi, "Qualitative dynamics of a piecewise-linear differential equations : a discrete mapping approach," *Dynamical Stab. System*, vol. 4, pp. 189–207, 1989.

Models of Biological Pattern Formation

Hans Meinhardt

Max-Planck-Institut für Entwicklungsbiologie, Spemannstr. 35, D-72076 Tübingen, Germany,
Hans.Meinhardt@tuebingen.mpg.de

Abstract

The development of a higher organism is controlled by a complex network of biochemical reactions that are under genetic control. In the following a short overview is given for some of the models we have proposed to describe essential steps in this process. Many of these models have found meanwhile direct support by molecular-genetic experiments. By computer simulations it has been shown that the models describe many of the observed phenomena. Animated simulations are available at http://www.eb.tuebingen.mpg.de/meinhardt

1. Basic mechanisms of biological pattern formation

For the generation of patterns in originally more ore less homogeneous tissues we have proposed that local self-enhancement and long-range inhibition is the driving force [1, 4]. Basic types of patterns can be generated in this way: organizing regions, gradients, periodic structures and stripe-like patterns. The simulations in Fig. 1 provide some examples.

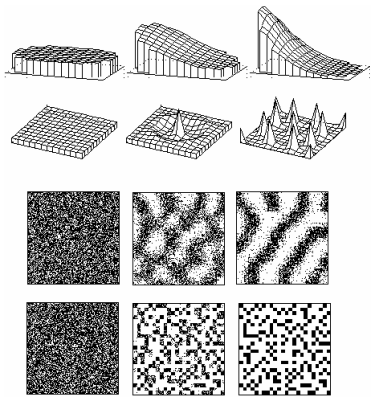


Figure 1: Elementary patterns generated by local self-enhancement and long-ranging inhibition. Left is the initial situation, at right the final stable pattern. An intermediate pattern is at the centre. If the range of the activator is compatible with the field size, graded distributions emerge. Such a pattern is able to generate polarity in an initially homogeneous field (top). If the size is larger than the range of the inhibitor, periodic patterns are formed. If the autoregulation saturates at high activator concentrations, stripe-like patterns are formed [15]. In the absence of activator diffusion, activated and non-activated cells are distributed in a salt-and-pepper fashion.

The following set of equations describes a prototype of an interaction between an autocatalytic activator *a* and the inhibitor *b* that allows pattern formation.

$$\frac{\partial a}{\partial t} = s\,\frac{a^2 + b_a}{b} - r_a\, a + D_a \frac{\partial^2 a}{\partial x^2} \qquad (1a)$$

$$\frac{\partial b}{\partial t} = s\,a^2 - r_b\, b + D_b \frac{\partial^2 b}{\partial x^2} + b_b \qquad (1b)$$

The inhibitor has to diffuse much faster in order that pattern formation can occur *($D_a << D_b$)*. In order that the pattern is stable in time the decay rate of the inhibitor must be higher than that of the activator *($r_b > r_a$)*, otherwise oscillations will occur. b_a describes a small activator-independent production rate of the activator necessary to initiate the autocatalytic activator production at from low levels, e.g. during regeneration. The factor *s* describes the ability of the cells to perform the autocatalytic reaction. In these simulations *s* is assumed to be uniformly distributed except some small random fluctuations that initiate the patterns.

Biological pattern formation shows in many situations a high degree of pattern regulation. This is a feature of the reaction described above. For instance, after the removal of the activated region, the inhibitor decays until the autocatalysis triggers again, restoring the original pattern.

The antagonistic reaction can also be based on the depletion of a substrate or co-factor that is required for activator autocatalysis. The following equation provides an example:

$$\frac{\partial a}{\partial t} = s\,b\,a^2 - r_a\,a + D_a\frac{\partial^2 a}{\partial x^2} \qquad \text{(2a)}$$

$$\frac{\partial b}{\partial t} = b_b - s\,b\,a^2 - r_b\,b + D_b\frac{\partial^2 b}{\partial x^2} \qquad \text{(2b)}$$

In this interaction, the substrate or co-factor *b* is produced by all cells of the field in which the reaction takes place with the constant rate b_b and removed during the autocatalytic activator production. As shown in the simulation below, in growing fields such an interaction leads to maxima that have the tendency to split. A maximum has the tendency to shift towards regions where higher substrate concentration is available. This will play a role in the formation of net-like structures (see Fig. 2)

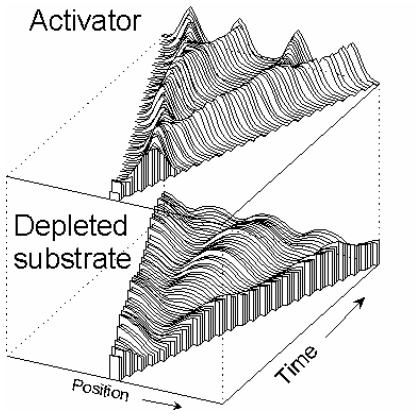


Figure 2: Pattern formation in which a production of an autocatalytic activator (top) proceeds on the expense of a factor (bottom) that becomes depleted during activator synthesis (Equation 2). Again, the co-factor must be long ranging. Shown is the activator and co-factor (depleted substrate) distribution as function of time in a growing linear field of cells. Whenever the distance between maxima exceeds a certain level, the maxima will split. In this way, the distance between maxima remains essentially constant.

2. Generation of complex structures

The complexity of an organism requires many coupled reaction. Using the freshwater polyp Hydra as an example, it has been shown of how it can be achieved that two organizing regions (head and foot) appear at the opposite terminal positions [7]. Hydra has only a single axis. It can be regarded as a living fossil that tells us about evolutionary inventions necessary to achieve bi-

lateral symmetry. The crucial step was the formation of a midline of the body. As shown in Fig. 1, stripe like patterns result by a saturation of autocatalysis. However, this leads to many stripes. The formation of a single stripe requires a cooperation of a patch-forming and a stripe-forming system. The patch-forming system makes sure that only a single stripe is formed while the stripe forming system is responsible for the formation a high concentration in a continuous line that is stable against decay into individual patches. We have shown that insects and vertebrates use different strategies. In vertebrates, a patch-like system, the node, elongates a stripe like system, e.g. the notochord. In insects, a dorsal organizer repels the midline to the ventral side [9].

3. Gene activation: a pattern formation among alternative genes

The temporary nature of signalling systems based on diffusion allows pattern formation only at a small scale. The formation of a large organism requires a permanent memory within the cells to which signals they have been exposed at earlier stages. Gene activation also requires self-enhancement and competition to allow the activation of only a particular gene among several alternative genes. Thus, cell differentiation can be regarded as a pattern formation in the gene space [4].

The equation below shows a very simple case to illustrate that a positive feedback can lead to threshold behaviour. A gene is assumed whose gene product *g* has a non-linear feedback on the activation of its own gene [3].

$$\frac{\partial g}{\partial t} = \frac{r_g g^2}{1+\kappa g^2} + r_g g + m \qquad (3)$$

At low *g* concentrations, the negative term dominates and the g-level will decline further. From a certain threshold level onwards, the autoregulatory term dominates and the concentration increases further until the saturation level is reached. A graded external signal m leads to stable the activation of the gene whenever *m* is above a certain threshold. This gene activation would remain unchanged in this sharply confined region even after switching off of the signal. Fig. 3 shows a simulation using this equation.

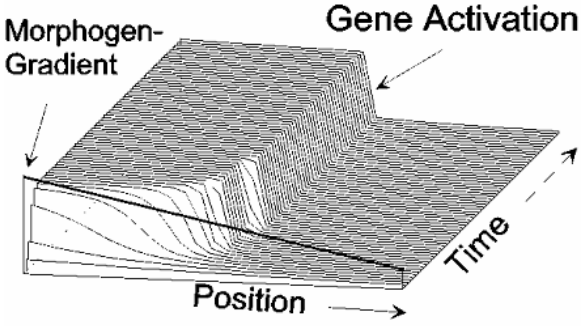


Figure 3: A stable switch-like activation of a gene based on an autocatalytic feedback loop that saturates at high concentration (Equation 3). Only those cells that are exposed to a certain threshold concentration in the morphogen concentration *m* switch from a low into a high concentration.

4. Generation of net-like structures

A feedback of a position-dependent gene activation on the pattern that has caused its activation can lead to very complex patterns. As an example the formation of filament-like branching structures will be discussed. This pattern is very common in almost all higher organisms. The vena-

tion of leaves, the tracheae of insects, the blood or lymph vessels as well as neurons are examples. How can such complex patterns emerge?

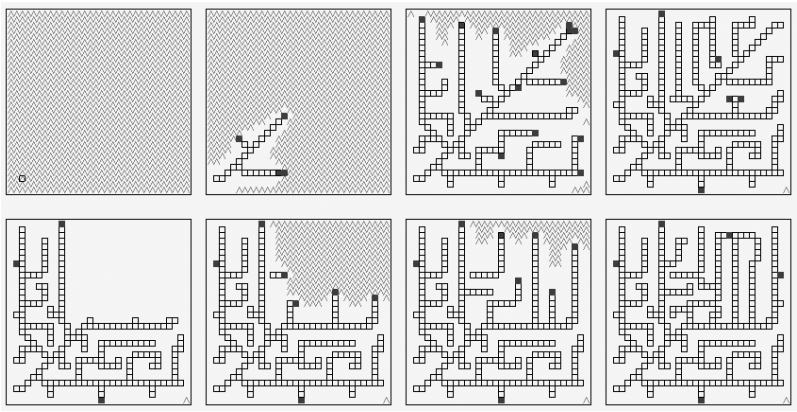


Figure 4: Formation of a net-like structure. The interaction of four substances is sufficient to generate a structure with branching filaments (equation 4a-d). A signal for the local elongation of the filament is generated by an activator *a* (black dots) / inhibitor system. In this simulation the signal is used to differentiate cells (squares). Differentiated cells remove a substrate *c* (wavy lines). Since the activator/inhibitor system depends on the substrate, the activator maximum is shifted to that neighbouring cell which has the largest distance from other differentiated cells. This is usually the cell in front of tip of the filament. The patterning process comes to rest if a certain density of the filaments is reached [3].

As shown above, the reactions given above allow the generation of patterns (Fig. 1) and this signal can be used to trigger a stable gene activation when a threshold is exceeded (Fig. 3). The exposed cells differentiate and become, for instance, a part of a vascular system. It is assumed further that differentiated cells repel the signal. The signal will be shifted into a neighbouring cell that will differentiate too. In this way, it becomes a part of the vascular system. A repetition of this process - differentiation, shift of the signal, differentiation - leads to a strand of differentiated cells behind a wandering activator maximum (Fig. 4).

For the simulation, the following set of equations has been used [3]; it is a combination of the equation given above).

$$\frac{\partial a}{\partial t} = \frac{sca^2}{b} - r_a a + D_a \Delta_a + b_a d \quad (4a); \qquad \frac{\partial b}{\partial t} = sca^2 - r_b b + D_b \Delta b + b_b d \quad (4b)$$

$$\frac{\partial c}{\partial t} = b_c - r_c c - c_c c\, d + D_c \Delta c \quad (4c); \qquad \frac{\partial d}{\partial t} = \frac{r_d\, d^2}{1 + s_d d} - r_d d + b_d a \quad (4d)$$

The concentration of *d* within a cell is an indicator of whether a cell is differentiated or not. As mentioned above (Fig. 3) only two stable states are possible. The concentration of *d* is either high or low. At a certain concentration of the activator sufficient *d* is produced to turn on the autocatalysis of *d,* the cell switches from low to high *d* production and remains at the high *d* level even after a decrease of the inducing signal. To achieve a shift of the activator maximum away from the differentiated cells, it is assumed that all differentiated cells remove a substance *c* that is produced by all cells. The maximum rate of activator autocatalysis is assumed to depend on *c*.

After differentiation of a cell, the *c* concentration drops and the maximum *a* concentration shifts to neighbouring cells containing the highest substrate concentration (see also Fig. 2). Usually, this is the cell in front of the tip of the incipient filament. The filament becomes elongated in this way. A growing filament may change the direction of its elongation due to the inhibitory influence nearby filaments or reflection of the inhibitor from the margins of the field.

Branches are formed whenever activator maxima become sufficiently remote from each other during elongation of filaments. Then, the inhibitor concentration can become locally so low that a new activator maxima is triggered along an existing vein due to the term $b_a d$. After removal of some filaments, the system is able to regenerate the missing veins (or whatever it is) since in these regions, *c* is no longer removed and the rising c concentrations attract activator maxima from the non-injured region.

This simple example shows that by superposition of several reactions pattern-forming systems can emerge that have a far richer repertoire then the components. This allows a tailoring of systems such that requirements for specific developmental situations can be met.

5. Initiation of substructures such as legs and wings

How it is achieved that wings and legs are always formed at particular positions of an organism, with a particular handedness and a particular orientation relative to the main body axes of the embryo? The complex structure of a higher organism requires the reproducible generation of new coordinate systems for sub-patterns. Such a new coordinates system can be generated if differently determined cells cooperate with each other to produce a new set of signalling substances [5, 6]. New signalling regions emerge that are centred over differentiation borders. The formation of a new coordinate system for a leg or a wing requires the intersection of two differentiation borders, one with A/P, and the other with a DV orientation. This model has found meanwhile much support in insects and vertebrates [14, 2].

6. Pigment patterns on shells of tropical sea shells and centre finding in bacteria: Highly dynamic pattern emerge by destabilization of a once established pattern.

Since a mollusc can enlarge it shell only by accretion of new material at the margin, the patterns on the shells are a time record of a one-dimensional process. They provide a natural picture book to study dynamic systems. Travelling waves play a crucial role therein that have unusual properties. For instance, they can penetrate each other in a soliton-like fashion (crossing lines on the shells) or waves can split, forming spontaneously backwards-running waves. The complex patterns result from several patterning superimposed patterning systems. Thus, the shells preserve a time record of the pattern along the growing edge; one can decipher the logic of the underlying mechanism.

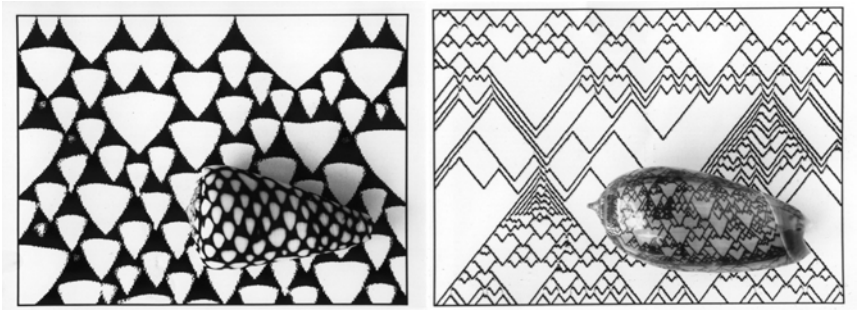


Figure 5: Pigment pattern on tropical sea shell. The pattern on *Conus marmoreus* (left) results from a permanent change between a widening and breakdown of a steady state pigment reaction The pattern on *Oliva porphyria* result from travelling waves that generate backwards-running waves whenever the number of waves drop below a critical level [12, 10].

Central in shell patterning is that a once generated pattern become destabilized shortly after their generation. Other biological systems use the same trick. The initiation of new leaves in Phyllotaxis [11, 15], the orientation of cells or growth cones by minute external signals [8] or the determination of the division plane in an E.coli bacterium [13] are examples. These mechanisms allow a dynamic adaptation to changing conditions although non-linear reactions have normally a substantial hysteresis.

References

[1] Gierer, A. and Meinhardt, H. (1972). A theory of biological pattern formation. Kybernetik 12, 30-39 (available for download at http://www.eb.tuebingen.mpg.de/meinhardt)

[2] Martin, G.R. (1995). Why thumbs are up. Nature 374, 410-411.

[3] Meinhardt, H. (1976). Morphogenesis of lines and nets. Differentiation 6, 117-123.

[4] Meinhardt, H. (1982). Models of biological pattern formation. Academic Press, London (available for download at http://www.eb.tuebingen.mpg.de/meinhardt)

[5] Meinhardt, H. (1983a). Cell determination boundaries as organizing regions for secondary embryonic fields. Dev. Biol 96, 375-385.

[6] Meinhardt, H. (1983b). A boundary model for pattern formation in vertebrate limbs. J. Embryol. exp. Morphol. 76, 115-137.

[7] Meinhardt, H. (1993). A model for pattern-formation of hypostome, tentacles, and foot in hydra: how to form structures close to each other, how to form them at a distance. Dev. Biol. 157, 321-333.

[8] Meinhardt, H. (1999). Orientation of chemotactic cells and growth cones: Models and mechanisms. J. Cell Sci. 112, 2867-2874.

[9] Meinhardt, H. (2002). The radial-symmetric hydra and the evolution of the bilateral body plan: an old body became a young brain. Bioessays 24, 185-191.

[10] Meinhardt, H. (2003). The Algorithmic Beauty of Sea Shells. (3rd edition) Springer, Heidelberg, New York S,.

[11] Meinhardt, H. (2003). Complex patterning by a self-destabilization of an established pattern: Chemotactic orientation of cell polarity and phyllotaxis as examples. Comptes Rendus Biologies 326, 223-237

[12] Meinhardt, H. and Klingler, M. (1987). A model for pattern formation on the shells of molluscs. J. theor. Biol 126, 63-69.

[13] Meinhardt, H. and de Boer, P.A.J. (2001). Pattern formation in E.coli: a model for the pole-to-pole oscillations of Min proteins and the localization of the division site. PNAS 98, 14202-14207.

[14] Vincent, J.P. and Lawrence, P.A. (1994). It takes three to distalize. Nature 372, 132-133.

[15] Koch, A. J. and Meinhardt, H.(1994). Biological pattern-formation - from basic mechanisms to complex structures. Rev. Mod. Physics 66, 1481-1507

Storage and recall of morphogenetic signals in plants

Michel Thellier

Laboratoire des Processus Intégratifs Cellulaires, UPRESA CNRS 6037, faculté des sciences et techniques, université de Rouen, 76821 Mont-Saint-Aignan Cedex, France

Introduction

Plants are sensitive to various environmental stimuli, such as wind, rain, mechanical contact, pricking by insects, wounds inflicted by herbivorous animals, infection by pathogens (viruses, bacteria or fungi) or stresses (e.g. water or thermal stresses). They respond to such stimuli, in a few cases by specific movements (e.g. *Mimosa pudica*, or some carnivorous plants), but most generally by modifications of their morphogenesis. The initial stimulus and the morphogenetic response may be disconnected in space, i.e. the response may occur not only in the stimulated area of the plant but also in tissues at a distance from the stimulated region: this implies that some sort of signal has to be transported from the stimulated to the reactive sites. The initial stimulus and the morphogenetic response may also be disconnected in time, i.e. the signal initiated by the initial stimulus may be "stored" in the plant without taking effect for a significant lapse of time, until another sort of stimulus finally causes this signal to be "recalled" and to take effect in the form of a modification of the plant morphogenesis.

In the study of such processes, the authors have often assumed implicitly that there is a linear chain of events between the reception of the initial stimulus and the final morphogenetic response. We shall see that this is not necessarily the case; and there may be several intertwined circuits of functions that exhibit positive or negative feedback loops. The "logical" (i.e. Boolean) language developed by René Thomas and co-workers is well adapted to formulate the behaviour of such systems, especially in predicting the possible occurrence of multistationarities and discussing their stability.

The experimental system of the Bidens *seedlings*

The most comprehensive set of experimental data has been obtained from the study of the "breaking of the symmetry of bud growth" in 3-week old seedlings of *Bidens pilosa* L. At this age the only expanded leaves are the two cotyledons; the terminal bud, also termed the "apex", develops very actively and the buds at the axil of the cotyledons (i.e. the "cotyledonary" buds) are quiescent as a result of apical dominance. The *Bidens* seedlings thus are bilaterally symmetrical. Each seedling may be given an arbitrary orientation by defining one of its cotyledons as being *a*, and the other *b*. When apical dominance is released by removal of the apex (seedling "decapitation"), the cotyledonary buds start to grow. Under optimal conditions of photosynthesis and mineral nutrition, both cotyledonary buds of each seedling begin to elongate at the same rate, with the consequence that these seedlings remain bilaterally symmetrical. However, under limiting conditions of light and mineral supply, one of the cotyledonary buds begins to elongate before the other, causing each seedling to lose its bilateral symmetry. For a set of seedlings, this symmetry-breaking effect may be characterised quantitatively using a non-dimensional "asymmetry index", **g**, defined by

$$g = (n_B - n_A)/n$$

where n is the total number of seedlings and n_A and n_B are the numbers of seedlings where it is the bud at the axil of cotyledon *A*, or *B*, which is the first to start to grow. A set of seedlings with g close to zero is said to be "symmetrical", whereas $0 < g \leq 1$ or $0 > g \geq -1$ correspond to

asymmetrical phenotypes in favour of side *B* or *A*. When the seedlings are kept under homogeneous conditions and/or subjected to symmetrical stimuli, a symmetrical response ($g \neq 0$) is always observed. But when an asymmetrical treatment is applied, a significant asymmetry in favour of bud *B* (g up to 0.3-0.6) may be observed. Note that in most experiments the asymmetrical stimulus consisted of pricking the cotyledon *A* of each seedling four times with a blunt glass needle, but identical results are obtained by applying non-traumatic asymmetrical stimuli to cotyledon *A* (e.g. by deposition of a drop of an appropriate solution or by gently rubbing the cotyledon).

When intact (non-decapitated) seedlings are subjected to the usual asymmetrical stimulus (4 pricks on cotyledon *A*) no evident external change results. However, if apical dominance is released by removing the terminal bud at a time Δt (up to 14 days) after the pricking treatment, again the bud at the axil of cotyledon *B* may start to grow before that at the axil of cotyledon *A*, with g-values again in the range of 0.3 to 0.6. The signal initiated by the asymmetrical pricking treatment thus can be stored (*STO* function), without significant loss of information, until it can finally take effect as a consequence of the release of apical dominance.

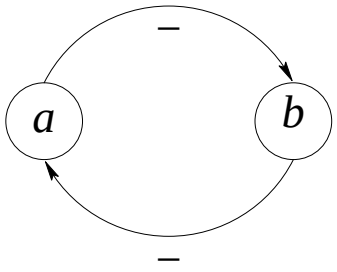
Obtaining the asymmetrical response as a result of applying the asymmetrical stimulus however is not compulsory. For instance, when seedling decapitation is carried out at the onset of daylight, indeed an unequivocal asymmetry in favour of the growth of the bud at the axil of cotyledon *B* is observed (g-value significantly above zero), but when the seedlings are decapitated in the middle of the day (other experimental conditions unchanged), this is no longer the case (g-value not significantly different from zero)! This does not mean that the stored symmetry-breaking information was erased as a consequence of decapitating the seedlings in the middle of the day, because giving a thermal (symmetrical) treatment to the plants immediately after their decapitation was enough to restore significant asymmetry (g-value appreciably different from zero). This means that, depending on the experimental conditions, the seedlings switch from a state where they can "recall" the stored symmetry-breaking signal to a state where they cannot, or conversely (*RCL* function).

From a great number of experiments it was concluded that i) the *STO* function is an "all-or-nothing" and irreversible process, while ii) the *RCL* function can be easily reversed, "*on*" and "*off*", depending on a great number of factors such as the mineral status of the seedlings, the application of symmetrical or asymmetrical pricking treatments, the conditions of seedling decapitation, the application of a thermal treatment or of a water stress, etc. Moreover, applying the symmetric stimuli (active on the *RCL* function) either before or after the symmetry-breaking stimulus (active on the *STO* function) had no appreciable effect on the intensity of the final response (the g-value). This suggests that the sequences of events corresponding to the *STO* and *RCL* functions are relatively independent from each other.

Simulation of the STO/RCL *plant behaviour using René Thomas's logical formalism*

In the formalism used here, the symbol "+" refers to the logical sum (inclusive "or"), while the symbol "·" or simply a blank refers to the logical product. A superscript corresponds to "not", i.e. $\bar{x}$ = not x; as an alternative, x and $\bar{x}$ can also be written x = 1 and x = 0, respectively. We have to distinguish between "internal" variables, which are characteristic of the studied system itself, and external, "input" variables, whose values depend on the experimenter. In the case of the internal variables, a small letter symbolises the variables itself, while the corresponding capital letter symbolises its "*image*", which corresponds to a conditional projection into the future. For example, if the cotyledonary buds are labelled *a* and

b, variable *a* refers to the presence or absence of bud *a* and its image *A* refers to its development. Thus, $a = 0$, $A = 1$ means that bud *a* is still absent but that the conditions for its development are fulfilled. For the input variables one may use indifferently capital or small letters. In this language, writing $A = \bar{b}$ and $B = \bar{a}$ means that bud *a* will develop if and only if bud *b* is absent, and that bud *b* will develop if and only if bud *a* is absent. This is equivalent to drawing the feedback circuit



and, drawing the state Table, it appears that the image of state 01 is 01, and the image of state 10 is 10. This means that states 01 and 10 are both stable logical states, and we therefore label them **01** and **10** (bold). According to whether a feedback circuit contains an even or an odd number of negative interactions, it is termed a positive or a negative circuit, respectively. Thus circuit (1) is a particular case of a positive feedback circuit. It is a general rule that multiple steady states require a positive feedback circuit.

Since all the experiments considered here were carried out under deficient conditions, it is not necessary to use a variable to represent the state (deficient or not) of the seedlings. However, in order to take into account explicitly that growth of the cotyledonary buds can take place only after apical dominance has been released, we rather write $A = T \cdot \bar{b}$ and $B = T \cdot \bar{a}$, in which *T* refers to the ablation of the terminal bud. A new state Table then has to be drawn, in which the image of the state vector depends not only on the internal variables, *a* and b, but also on the input variable, *T*, corresponding to the fact that the experimenter can decide at any time to remove the terminal bud. Hence there are two columns in which *T* has the value 0 or 1. It must be stressed that a description of the system using the single input variable, *T*, is valid only under conditions when no asymmetric stimulus has been applied to the seedlings.

When an asymmetrical stimulus (i.e. the pricking treatment *P*) is applied, the storage function, which is symbolised *S* (*STO* in the publications cited above) while s is its associated variable, is "*on*" and remains irrevocably "*on*". In the case when various treatments reversibly allow or forbid the irreversibly stored asymmetric signal to take effect, this leads us to define a "recall" function, *R* (*RCL* in the publications cited above), required for the actual expression of the asymmetry at the level of the population of seedlings. The variable is labelled *r*, and its image, *R*, in agreement with the conventions used in our formalism. The introduction of these other variables increases the number of columns in the state Table. By following the logical pathways in the state Table, it becomes possible to understand how the interplay of the storage and recall functions, *s* and *r*, describes the relative growth of the cotyledonary buds and which are the stable states under the various experimental conditions. The reversibility of *r* is clearly apparent since changing $\bar{r}$ to *r* and inversely will cause the seedling state to change from 11 to 01 and inversely as long as the seedlings have not reached either of the possible stable states.

Other examples of STO/RCL behaviour in plants

Here, the logical symbolism has been applied to the breaking of the symmetry of bud growth in *Bidens* seedlings. Potentially, this approach may be extended to a lot of other experimental situations where other *STO*/*RCL* effects occur in plants. For instance, stimulating a *Bidens* seedling (whether asymmetrically or not) induces an inhibitory signal for hypocotyl growth; but hypocotyl growth inhibition will occur only under appropriate ionic conditions in the nutrient medium of the plants. The pricking of one cotyledon of seedlings of flax and of two cultivars of tomato (*Groseille rouge* and *Marmande*) broke the symmetry of the growth of the cotyledonary buds in a way similar to that observed with the *Bidens* seedlings but with smaller values of the asymmetry-index; in similar manner, cotyledon pricking induced inhibition of hypocotyl growth only under appropriate ionic conditions. In all these cases, retarding the application of the appropriate ionic conditions retarded the morphogenetic response. Applying various sorts of stimuli to flax seedlings results in the induction of meristem in the hypocotyl epiderm, on the condition that a depletion in calcium nutrition has been applied; again, retarding calcium depletion retarded of an equal number of days the development of the meristems.

Discussion and conclusion

One may wonder about the role of such *STO*/*RCL* effects in the control of plant morphogenesis, as a consequence of the reception of environmental stimuli. An attractive hypothesis is that plants should not react to each individual stimulus received from the environment, but should rather make an integrated response to the various stimuli received at a given time. Clearly, the fact that signals induced by certain stimuli may be stored while being recalled only if some other stimuli have also been received, would be an appropriate mechanism for that.

The logical language developed by Thomas thus may be of rather general interest for the simulation of such integrated responses of plants to environmental stimuli.

Using an attribute estimation technique for the analysis of microarray data

Jeremy Mary[1], Géraldine Mercier[2], Jean-Paul Comet[3], Antoine Cornuéjols[1], Christine Froidevaux[1] & Marie Dutreix[2]

[1] Laboratoire de Recherche en Informatique, Université Paris Sud & CNRS UMR 8623, 15 avenue George Clémenceau, F-91405 Orsay Cedex
[2] Institut Curie, Université Paris Sud & CNRS UMR 2027, 15 avenue George Clémenceau, F-91405 Orsay Cedex
[3]Laboratoire de Méthodes Informtiques, Université d'Evry Val d'Essonne & CNRS UMR 8042, Tour Evry 2, 523 Place des terrasses de l'agora, F-91000 Evry

Abstract

We present an original method for the analysis of microarray data based on an attribute estimation technique. We show, in the context of radioactive environmental stress, that this method can be used to detect a small number of genes that, highly probably, are involved in the response to this stress in yeast. We also studied the different processes in which these informative genes are involved. Moreover, we ranked the genes according to their ability to distinguish between two experimental conditions (irradiated versus non irradiated). Using this method, we highlighted the transcriptional response of yeast cells to low doses of radiation.

1 Introduction

Yeast microarrays provide information about the transcription of all genes in a cell population. Genome-wide monitoring of transcript changes in yeast could provide information about previously unrecognized cellular responses to environmental stress and reveal specific genes that allow yeasts to survive in such conditions. As microarrays generate massive amounts of data, data analysis methods are required to determine whether changes in gene expression are indeed significant.

In our approach, we first looked for a method to analyze the transcriptome of these yeasts with the aim of identifying genes involved in the response to environmental changes. We then tried to identify the functional groups to which these newly highlighted genes belong.

In the first stage, we were particularly interested in answering several questions:

(1) the number of genes involved in the transcriptional response;
(2) the identity of these genes (to determine whether genes induced by low-level irradiation are also induced by acute irradiation);
(3) the possibility to use transcriptional data to predict whether a given sample belongs to one of the considered (environmental) classes (i.e. had the yeast been exposed to radiation or not).

This standard problem is named "supervised classification" [1]: some examples from both classes (treated yeasts and non-treated yeasts) are available for training purposes and the aim is to distinguish them by means of the expression levels of a subset of pertinent genes to be determined.

This task is however difficult for several reasons.

(1) Available data are noisy due to:
   - imprecise measurements: classical noise assumed to be Gaussian
   - the fact that some data are aberrant due to spotting or hybridization problems.

(2) The number of attributes is huge: expression level of 6135 genes.

(3) The number of training examples is usually very low: 18 in our study.

(4) Classes are not well balanced (they do not contain the same number of slides): 12 non-treated cultures and 6 treated cultures in our study.

(5) The expression of genes is correlated (common regulatory pathways), which goes against the assumption of probabilistic conditional independence.

In an ideal learning process, the measurement of gene expression levels alone should make it possible to identify all the genes involved in a given environmental response. Unfortunately, with the learning methods available it is difficult to identify such sets of genes given the large number of attributes (the expression levels of genes) and the low number of training examples (the available microarray slides). Thus, we had to choose between two alternatives:

- The detection of almost all the genes possibly involved in the transcriptional response (small number of false negatives) even if there is a risk that genes not involved in the response will be considered to be pertinent (many false positives).
- The detection of only genes that are almost certainly involved (small number of true positives), with the possibility that a lot of involved genes will be overlooked (numerous false negatives).

As small subsets of genes make it easier to identify involved biological functions and as we wanted to limit the number of false positives, we chose the second approach. Using the RELIEF attribute estimation method (see section 3), we showed that it is possible to detect a small number of yeast genes that are almost certainly involved in the response to radioactive stress.

2. Biological design and data collection

Our environment is altered by human and industrial activities that release different kinds of chemicals. These products, which are potentially highly toxic for public health, are difficult to detect because of their very low concentrations and because they are mixed with other compounds. We started this study from the assumption that exposure to genotoxic agents alters the transcription of some genes and that these changes can be estimated by comparing the messenger RNA (mRNA) levels of two populations growing in different environments. We used DNA microarrays to measure all the mRNAs of the *Saccharomyces cerevisiae* yeast in different growth conditions. The aim of this study was to identify the specific response induced by a given chemical. We limited our study to the quantification of the biological effects of low doses of ionizing radiation, particularly when they are delivered at low dose rates, as is the case in most environmental conditions.

To analyze the effect of continuous exposure to low doses of radiation, we quantified the expression level of most of the yeast genes in cells grown with (I) and without (NI) low doses of irradiation. For this purpose, cells were exponentially grown in rich medium for 20 hours (corresponding to 12 division cycles) in the presence or absence of β radiation (1.71 Mev). The number and the distribution of cells at each stage of the cell cycle were then determined by microscopy: the frequencies of single cells (G1), budding cells (S) and doublets (G2/M) in the NI and I populations should be similar if growth is not affected by the irradiation. Doses were considered to be "low" when no differences in cell growth or mutation (or recombination) frequencies were observed between NI and I populations [7]. In the first set of experiments, the dose rate was between 10 and 30 mGy/h. Although we observed no physiological (growth) or

genetic (recombination and mutagenesis) effects in the irradiated populations, the expression levels of numerous genes were significantly changed. The relative expression level of each gene was estimated by use of glass slide microarrays (produced by Corning) spotted with 6135 denatured DNA sequences corresponding to all of the open reading frames (ORFs) of *S. cerevisiae.* We labeled the control cDNAs with Cy3-dye and the NI or I cDNAs of interest with Cy5-dye. The same control cDNA was used for all the experiments. It was prepared from a pool of independent cultures grown without treatment.

Hybridized microarrays were scanned using a Genepix 4000B machine (Axon Instruments). Separate images were acquired for each type of fluorescence, at a resolution of 10 □m per pixel. Images were analyzed with Genepix pro 3.0 (Axon), after manual rectification of the outline of each spot. The median values for both types of fluorescence were used for each spot. A quality control standard (QCS) was estimated, calculating the difference between the median pixel values within the spot (spot) and in the background (bkgrd), corrected with the square root of the sum of the standard deviations (stdev).

$$\frac{\text{median}(spot) - \text{median}(bkgrd)}{\sqrt{\left(\frac{\text{stdev}(spot)\cdot\sqrt{\pi}}{\sqrt{2}\cdot\text{nbpixels}(spot)} + \frac{\text{stdev}(bkgrd)\cdot\sqrt{\pi}}{\sqrt{2}\cdot\text{nbpixels}(bkgrd)}\right)}} > \text{Quantile normal}\left(\frac{0.5}{2\cdot number\ of\ spots}\right)$$

Data with "non significant" QCS values for both types of fluorescence were considered to be missing.

The data generated by cDNA microarrays are affected by many experimental parameters other than differential expression and this systematically leads to variations in the measured intensities. These variations cannot easily be quantified by standard methods. Measured intensities must be normalized before we can compare measurements from different microarrays experiments. The main assumption underlying normalization is that there is a functional coherence between a true biological difference and the corresponding measured values. Looking for crude ratios of signal intensity in intensity-dependent dye normalization methods seems preferable to relying on global methods such as mean or median normalization. We used the location and scale normalization procedures originally developed by Yang [15]. These methods correct for intensity and spatial dye biases, by use of a robust local regression. They use the Splus LOWESS function (Insightful) to perform robust local regression and were applied to obtain a scaled within print group normalization to account for spatial dependence in dye biases, with scale adjustment between the blocks. To make this method more robust, we did not consider saturating points (with saturating fluorescence intensities) when estimating LOWESS fits. However, as these points contain relevant information, all measured intensities were normalized with the estimated LOWESS curves. The normalized data for each spot were defined as the estimated relative expression levels for experiments with irradiated populations (I values) and non irradiated populations (NI values).

3. Attribute estimation with the RELIEF algorithm

Given the characteristics of the available data (large number of attributes versus few examples), it is not worth trying to detect complex correlations, as most of the numerous correlations found will be spurious. Instead, we used approaches that examine the possibility of a direct correlation between each gene and class (I or NI). This is an attribute estimation problem that can be addressed by numerous techniques (see for instance [5, 8, 10, 12, 14]). For reasons that will

become clear shortly, we used an adaptation of the RELIEF method [5, 6], which is an attribute estimation method that looks for the attributes that seem to be the most significantly correlated to the class to be predicted. The overall principle is to calculate a normalized weight for each gene. These weights are comprised between -1 and +1, with positive weights indicating that there is a positive correlation between the relative expression level of a gene in a given example and its class. The weight of a gene is a function of the variation of its relative expression level within each class compared to the variation between classes. Indeed, the correlation between class and relative gene expression seems to be stronger if the intra-class variation is small compared to the extra-class variation.

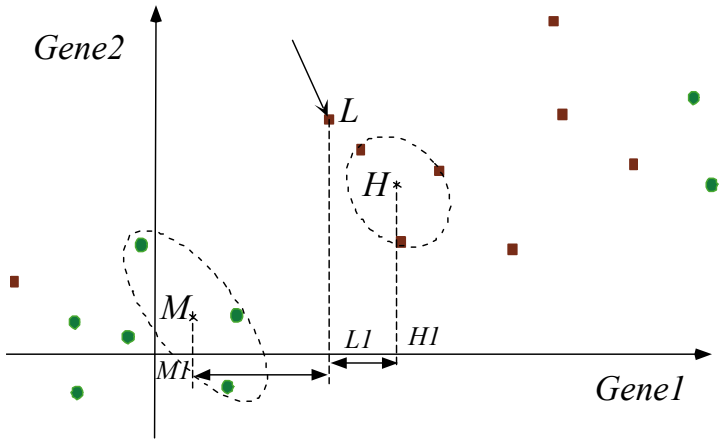


Figure 1 – This figure illustrates the RELIEF principle when only two genes are considered. Examples are in this case represented by points in the plan determined by the relative expression levels of the two genes. Here, there are 17 examples, 9 of which belong to the "square" class, and 8 to the "circle" one. In this case, the example *L* positively contributes to the overall estimated relevance of Gene1.

The weight associated with a gene (an attribute) is calculated as follows. One example *L* can be considered as a point in the attribute space (a 6135-dimensional space). For each of the *m* examples, labeled I or NI, we determine its *k* nearest neighbors in the same class and calculate the corresponding barycenter *H* (for nearest Hit), and then determine its *k* nearest neighbors in the other class and calculate the corresponding barycenter *M* (for nearest Miss). For each gene *g* in turn, we calculate the projection of the points *L*, *H* and *M* on the associated axis. For instance, the projection *L1* of point *L* on the gene *g* corresponds to the expression level of the gene *g* for the example *L*, whereas the projection of point *H* (respectively *M*) corresponds to the mean expression level of gene *g* for the *k* nearest neighbors of the same class (respectively of the other class). We then calculate, on one hand, the distance between the projection of *L* and the projection of *M*, and, on the other hand, the distance between the projection of *L* and the projection of *H* (see figure 1). The difference distance(*L1, M1*) - distance(*L1, H1)* between these two values provides the contribution of this example to the estimated relevance of gene *g*. This is repeated for all examples in turn. The sum of all the contributions thus obtained, divided by the number of examples, defines the weight associated with each gene.
This can be summarized by the following formula:

$$weight(gene) = \frac{1}{m}\sum_{L=1}^{m}\{[\exp r_{gene}(L) - \exp r_{gene}(M)] - [\exp r_{gene}(L) - \exp r_{gene}(H)]\}$$

where $expr_{gene}(x)$ is the projection of point x on the axis associated with the gene, and m is the total number of examples.
The weight that is calculated in this way for each gene is an approximation of the difference of two probabilities (see [6]) as follows:

Weight(gene) = P(gene has a different value / k nearest neighbors in a different class)
- P (gene has a different value / k nearest neighbors in the same class).

In our work, we chose to use the Manhattan distance to determine the neighbors in the 6135-dimensional space instead of the more classical Euclidian one, because the latter tends to overvalue the genes (attributes) that exhibit very noisy measures or aberrant values.
As well as its low computational cost (after we had optimized the code), one important advantage of RELIEF is that, unlike many other statistically oriented techniques, it does not assume that genes are independent and does not need information on the data distribution (e.g. Gaussian). Moreover, the parameter k makes it possible to control the tradeoff between sensitivity and robustness to noise. In our case, k=3 was determined empirically as the best choice. It avoids an unreasonable sensitivity to aberrant values while being more precise than methods that take into account only mean values.

4. Determining informative genes

It is difficult to decide how many genes are required to determine whether a sample has been irradiated or not by merely examining the weights obtained. Indeed, it is difficult to choose a minimal weight (threshold) beyond which a gene can be considered to be involved in the response to radiation. Moreover, the small number of samples, together with the intrinsic noise in the data, makes the task even more complex, meaning that some genes may appear to be strongly correlated by pure chance. To estimate the probability that irradiation does not lead to any variations in relative gene expression (null hypothesis), we compared the correlation between the relative expression level of each gene and the class in the experiment with the same correlation for random permutations of class labels. We labeled samples such that one class contained 12 and the other one 6 experiments, to respect our original distribution, as RELIEF is sensitive to proportions. For each permutation, we then re-applied RELIEF to obtain new weights for each gene. The process was repeated until the curve of average number of genes having a given weight stopped varying (see figure 2). The process was repeated 2000 times and the curve started stabilizing after 500 runs.
If there is indeed a correlation between the relative gene expression levels and the class, more genes should appear to be correlated in the experimental condition than in the null hypothesis condition for a given weight. Figure 2 shows the curves of the number of apparently correlated genes for a given weight corresponding to the two conditions (experimental and randomized). A difference is clearly visible, thus confirming that exposure to low radiation doses affects the expression of some genes.
More precisely, this figure allows us to select a set of genes that are highly likely to be involved in the adaptive response to low doses of radiation. One possibility would be to consider the point where no genes appear to be correlated with the class in the null hypothesis condition (beyond the 0.58 weight). At this threshold, 13 genes were correlated with class in the experiments and the probability was very low that any of them could be attributed to chance. It must be emphasized that other genes are certainly involved. To obtain a larger set of genes for the purpose of biological interpretation, we are willing to accept that 10% of the selected genes are false positives. This ratio is obtained for a threshold of 0.3: if we consider the genes with a weight greater than 0.3, 171 genes are selected in the experimental condition compared to about 17 in the

randomized condition (see figure 2). The 171 genes with a weight greater than 0.3 are considered as "informative".

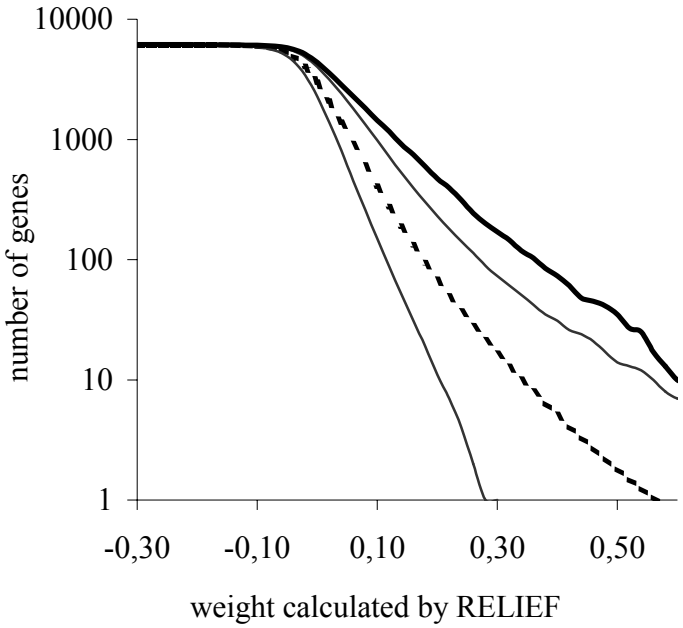


Figure 2 - Experimental (continuous bold) and randomized (dashed) curves showing the number of genes, the weight of which is greater than or equal to the threshold given on the x axis. The thin lines indicate the 95% confidence interval of the random distribution.

5. Interpreting biological results

In addition to its ability to separate the samples depending on whether they were exposed to radiation or not, the method allowed us to identify 171 informative genes responding to low doses of radiation. The analysis of the cellular functions involving the proteins coded by these genes could provide new information about the cellular target of the radiation and the functions induced in the response.

We analyzed the first 171 genes ranked by RELIEF. These genes were grouped according to the cellular process in which they participate. The processes implicated in radiation response were identified by comparing the proportion of genes involved in a given process within the first 171 ones with the proportion of genes involved in the same process within the 6135 measured genes. More precisely, the second column of the table indicates the number of ORFs induced (top part of the table) or repressed (bottom part) in response to radiation that take part in the cellular process indicated in column 1. The third column gives the proportion of the genes responding to each cellular process among the responding genes (91 induced genes in the top part and 80 genes in the bottom part). The fourth column indicates the proportion of the number of genes responding to the cellular process, represented on the microarray. The last column indicates the level of over-representation (ratio) of each cellular process among the 171 responding genes.

Only the processes involving more than four genes are considered. Three processes (highlighted in the table) are clearly induced by exposure to low levels of radiation. They all participate in the oxidative phosphorylation cascade. The genes participating in these processes are 14 to 30 times more frequent in the first 171 genes ranked by RELIEF than in the whole of the 6135 genes on the microarray (Table 1).

Without explaining the biological interpretation of these functional pathways in detail (ongoing work), our confidence in the method used for gene selection is strengthened by the fact that the functions revealed by the selection of the informative genes are known to be involved in the elimination of some cellular products caused by ionizing radiation (e.g. free radicals).

function of 91 induced genes/171	number of genes	%in this list	%total genes (6135)	ratio
unknown	38	41,8	50,4	0,8
oxidative stress response	4	4,4	0,3	14,3
oxidative phosphorylation	9	9,9	0,3	30,5
transport	4	4,4	2,2	2,0
gluconeogenesis	1	1,1	0,1	16,9
protein processing & synthesis	3	3,3	2,0	1,6
ATP synthesis	7	7,7	0,4	20,6
glucose repression	1	1,1	0,2	4,8
respiration	2	2,2	0,1	22,0

function of 80 repressed genes/171	number of genes	%in this list	%total genes (6135)	ratio
unknown	45	56,3	50,4	1,1
stress response (putative)	1	1,3	0,2	7,0
glycerol metabolism	2	2,5	0,1	30,8
protein processing & synthesis	3	3,8	2,0	1,9
secretion	2	2,5	2,0	1,3
transport	4	5,0	2,2	2,3
glycolysis	2	2,5	1,0	2,5

Table 1

6. Comparison with other methods

The use of RELIEF to select up- and down-regulated genes is original. We compared it to the very common ANOVA (ANalysis Of VAriance) method [2, 10], which is a generalization of the multi-class case of the Student's t-test, a statistical method that is at the root of the SAM method (Significance Analysis of Microarrays [12]). Analysis of variance has already been used for gene expression data [4] and led to the proposal that microarray data should be normalized, which in turn made it possible to analyze some aspects of the data (including the search for informative genes). In our study, we used analysis of variance to look for informative genes after applying a LOWESS normalization (see section 2).

ANOVA tests the equality of several means of a variable (here relative gene expressions) according to independent variables (here the class of microarray slides: I or NI). It is assumed that the data samples are chosen randomly and are independent, and that the populations have the same variance (this is a strong hypothesis not always true of microarray data). The principle of the method is to estimate the variance, first taking into account the classes and secondly without taking into account them. If the class is not correlated to the measured variable, these two estimated variances should be similar. We applied ANOVA to each of the 6135 genes, thus giving a number measuring the statistical correlation between relative gene expressions and

classes for each gene. To see if the Fisher statistics given by ANOVA agree with the null hypothesis that there is no correlation between relative gene expressions and classes, we used the same approach as for RELIEF: shuffling classes and calculating a large number of new Fisher statistics (see section 3). The informative genes are those for which the Fisher statistic (for the real classes I and NI) is very high as compared to the statistics obtained for the random cases.
We also compared the ANOVA results with those obtained with the SAM software [12], which identifies the most significant genes by ordering them according to a statistic based on the variations in gene expression levels rescaled with the standard deviation. Thus, the selected genes are those for which the rank is far away from the mean rank calculated with the shuffling process.
We compared the three methods by analyzing the intersections of the 500 top-ranked genes. For instance, the ranking obtained by SAM is relatively similar to that obtained by ANOVA, with 82% of the top 500 genes being identical. We then compared the results obtained by the RELIEF and ANOVA methods; 257/500 genes were found to be common to both rankings. The probability of obtaining an overlap of this size between two groups of 500 genes randomly selected among 6135 is very low (about 10^{-160}) according to the hypergeometric law. This probability is in the same range for different measured overlaps: if we take the first q ranked genes obtained by both methods (for q between 100 and 2000), the overlap size is greater than the half of the chosen number of genes (q/2). Thus, we can conclude that although ANOVA and RELIEF are based on different principles, they give partially the same information from the same data. Given the absence of independence between gene expressions in RELIEF and the fact that it can tolerate noise, the biological interpretation that we gave concerns the genes obtained by RELIEF (see Table 1).

7. Conclusion

The use of microarrays to distinguish between organisms living in various environmental conditions and to reveal the biological processes involved in survival in certain conditions is a great challenge. One of the main reasons for this stems from the fact that very few measurements are usually available compared to the large number of genes. Another reason is that the usual statistical assumptions regarding both the distribution of gene expression levels and the independence of genes do not appear to hold true (groups of genes are known to be correlated). We therefore propose an original attribute estimation method that remedies these drawbacks and is robust to noise. The analysis carried out with RELIEF made it possible to bring to the fore the transcriptional response induced by low doses of radiation. We ranked the genes according to their ability to distinguish between samples exposed to radioactivity or not. This revealed that some of the induced genes are involved in specific processes: three of these processes are clearly induced by exposure to low-level radiation and participate in the same oxidative phosphorylation cascade. This suggests that it could be possible to use dedicated microarray data (based on selected genes) for diagnostic purposes. We also plan to exploit the same approach for other biological studies such as, for instance, the classification of tumors in humans [13].

Acknowledgements

The authors gratefully acknowledge the "Consensus" research group for stimulating discussions, and the Genopole and the DRIDF for supporting group meetings financially. This work was supported by the Institut National de Recherche et de Sécurité (convention n°5011888), the CNRS, the Curie Institute, the Association pour la Recherche sur le Cancer (#5659) and Biogen n°74 (BioIngénierie 2001).

References

[1] Cornuéjols A. and Miclet L. : Apprentissage artificiel. Concepts et algorithmes. Eyrolles, 2002.

[2] Glantz S.A. and Slinker B.K., Primer of Applied Regression & Analysis of Variance, McGraw-Hill/Appleton & Lange, 2nd edition, 2000.

[3] Grant G., Manduchi E. and Stoeckert C., Using non-parametric methods in the context of multiple testing to determine differentially expressed genes, in Methods of Microarray Data Analysis: Papers from CAMDA'00, eds Lin SM. and Johnson KF., Kluwer Academics, pp. 37-55, 2000.

[4] Kerr M.K., Martin M. and Churchill G.A., Analysis of variance for gene expression in microarray data, Journal of Computational Biology, 2000, 7(6), 818-837.

[5] Kira K. and Rendell, L., A practical approach to feature selection. In Proc. International Conf. on Machine Learning, Aberdeen, D. Sleeman and P. Edwards (Eds.), Morgan Kaufmann, 1992, pp. 249-256.

[6] Kononenko I., Estimating Attributes : Analysis and Extensions of RELIEF, Proc. of the European Conference on Machine Learning, ECML-94, 171-182, 1994.

[7] Mercier G., Denis Y., Marc P., Picard L., Dutreix M., Transcriptional induction of repair genes during slowing of replication in irradiated *Saccharomyces cerevisiae*, Mutation Research 487, 2001, 157-172.

[8] Ng A. and Jordan M., Convergence rates of the voting Gibbs classifier, with application to Bayesian feature selection, citeseer.nj.nec.com/445984.html.

[9] Park P., Pagano M., Bonetti M., A non parametric scoring algorithm for identifying informative genes from microarray data, Pacific Symposium on Biology: 52-63,2001.

[10] Pavlidis P. and Noble W., Analysis of strain and regional variation in gene expression in mouse brain, Genome Biology, 2001, 2(10).

[11] Troyanskaya O., Garber M., Brown P., Botstein D. and Altman R., Nonparametric methods for identifying differentially expressed genes in microarray data, Bioinformatics, Vol.18, no.11, pp. 1454-1461, 2002

[12] Tusher V., Tibshirami and Chu GG., Significance analysis of microarrays applied to the ionizing radiation response, PNAS, April, 2001, Vol 98, n°9, 5116-5121.

[13] Van't Veer L., Dai H., van de Vijver M., He Y., Hart A., Mao M., Peterse H., van der Kooy K., Marton M., Witteveen A., Schreiber G., Kerkhoven R., Roberts C., Linsley P., Bernards R. and Friend S., Gene expression profiling predicts clinical outcome of breast cancer, *Nature*, 415, January, 2002.

[14] Xing E., Jordan M. and Karp R., Feature selection for high-dimensional genomic microarray, Proc. of the Int. Conf. on Machine Learning, ICML-2001, 601-608, 2001.

[15] Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 2002 Feb 15; 30(4):27-28.

Physico-chemical aspects of the bacterial genome

Sónia Cunha[1, 2], Theo Odijk[2] , Norbert O. E. Vischer[1], S. Elmore[1] & Conrad L. Woldringh[1]

[1]Swammerdam Institute for Life Sciences, BioCentrum Amsterdam, University of Amsterdam, Kruislaan 316, 1098 SM Amsterdam, The Netherlands
[2]Section Theory of Complex Fluids, Kluyver Institute for Biotechnology, Delft University of Technology, P.O. Box 5057, 2600 GB Delft, The Netherlands

Abstract

As the genome in bacteria is not enveloped by a membrane like in eukaryotic cells, bacterial DNA must be compacted by other forces. We propose a semi-quantitative physico-chemical model for the bacterial nucleoid as developed by one of us (Odijk, 1998). The model predicts the occurrence of a phase separation between DNA and cytoplasm based on excluded volume or depletion interactions. Observations on the compaction of isolated nucleoids by addition of polyethylene glycol and on the rate of expansion of the nucleoid after liberation by osmotic shock from spheroplasts confirm the predicted behavior of the nucleoid. To explain the directional motion of DNA within the cell during the process of segregation, we propose that the so-called transertion mechanism (i.e. co-transcriptional translation and translocation of membrane proteins; Norris, 1995; Woldringh *et al.*, 1995) plays a role in the bi-directional movement of DNA daughter strands.

1 Theory of phase separation by depletion forces

In *E. coli*, the circular chromosome with a contour length of about 1580 μm is assumed to occur in the form of a single branched, plectonemic supercoil with a contour length of 630 μm. This long worm-like chain is assumed to be composed of about 4000 so-called Kuhn segments: relatively stiff supercoiled regions with a length of two times the persistence length, i.e. 158 nm. When packed within the small volume of an *E. coli* cell (0.37 μm^3) and in the absence of soluble proteins, the Kuhn segments, colliding with each other through thermal fluctuations, try to expand like a spring. Due to this self-interaction an energy F_n can be assigned to the nucleoid that is proportional to the number and volume of interacting Kuhn segments and inversely proportional to the packing volume (Figure 1A).

The living cell, however, is crowded with numerous soluble and negatively charged protein particles. As DNA and proteins cannot occupy the same space, a cross-interaction energy F_c can be considered that is proportional to the cross-excluded volume between one protein and the DNA double helix times the number of interacting proteins per cell volume (Figure 1B). The crowded cell as a whole minimalizes its free energy by maximizing the volume that is available to the soluble proteins. This results in a volume exclusion of the DNA into a smaller nucleoid (Figure 1C). At equilibrium the phase separation results in a nucleoid with a volume of 0.068 μm^3 containing 43,000 proteins and a cytoplasmic phase containing about 10^6 proteins. As a result of this a phase separation F_n increases (because of a smaller volume for the supercoil segments) while F_c decreases (because fewer proteins cross-interact with the DNA double helix).

To test the theory of phase separation the free energy of the supercoiled nucleoid was experimentally estimated and found to be similar to the theoretical value (Cunha *et al.*, 2001). As a further test of the theory the rate of expansion of the DNA upon osmotic lysis of spheroplasts was measured and found to be comparable to the rate theoretically predicted by Odijk (2000).

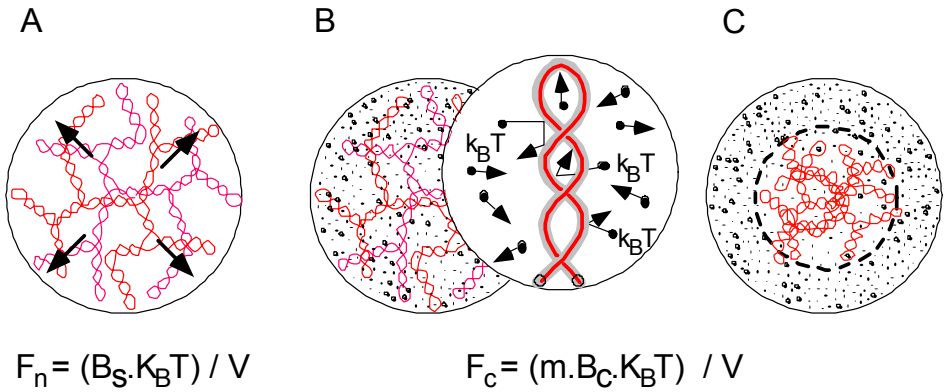


Figure 1: Schematic representation of self-interactions of DNA supercoil segments and of their cross-interactions with proteins. B_s, total excluded volume of Kuhn segments; B_c, cross-excluded volume between one protein and the DNA double helix; m, number of soluble proteins; k_BT, amount of energy per thermal fluctuation [Odijk, 1998 and Woldringh, 2002] (see plate 4).

2 The transertion model

While the equilibrium phase separation described above does not require energy for its existence, a growing cell represents a non-equilibrium steady-state in which DNA, RNA, protein and membrane surface are synthesized. As a result, the transcribed nucleoid is expanded and segregates into two compartments before cell division. Our working hypothesis is that Brownian diffusion rather than motor proteins in a mitosis-like mechanism, provide the driving force for segregation. In addition, the process of co-transcriptional translation and translocation of membrane proteins (called transertion, Norris, 1995) creates transertion areas (see Figure 2) that direct the bipolar movement of replicated daughter strands according to the following principles: (i) genes expressing membrane proteins become transiently constrained at the membrane by the transertion process; (ii) after initiation of DNA replication genes of both daughter strands compete for membrane surface to form transertion areas; (iii) cooperativity between neighboring genes on the same daughter strand causes a symmetry breaking in the random occupation of membrane surface resulting in an accumulation of transertion areas formed by the same daughter strand in one region of the cell; (iv) this spatial bias generates a self-enhancing movement of the two daughter strands in opposite directions; (v) during the transient and cooperative constrainment of genes in one region of the cell, Brownian diffusion of non-transcribed DNA supercoil segments represents the driving force for gradual movement of the bulk of the DNA.

Time-lapse microscopy of *E. coli* cells (strain SG102/pSG20) stained with DAPI and containing GFP-tagged *oriC*-spots suggest a diffusional motion of this DNA region within the nucleoid compartment. Upon division or “splitting” of *oriC*-spots no rapid separation of the spots was observed as both spots continued to move in a random way. These observations confirm the results obtained by Roos *et al.* (1999) on the localization of *oriC* by fluorescent *in situ* hybridization.

Acknowledgments

We thank Andrew Wright for the generous gift of the *E. coli* strain SG102/pSG20.
This work is supported by the Stichting Aard- en Levenswetenschappen (ALW project number 805.47-032-P), subsidiary of the Dutch organization for Scientific Research NWO.

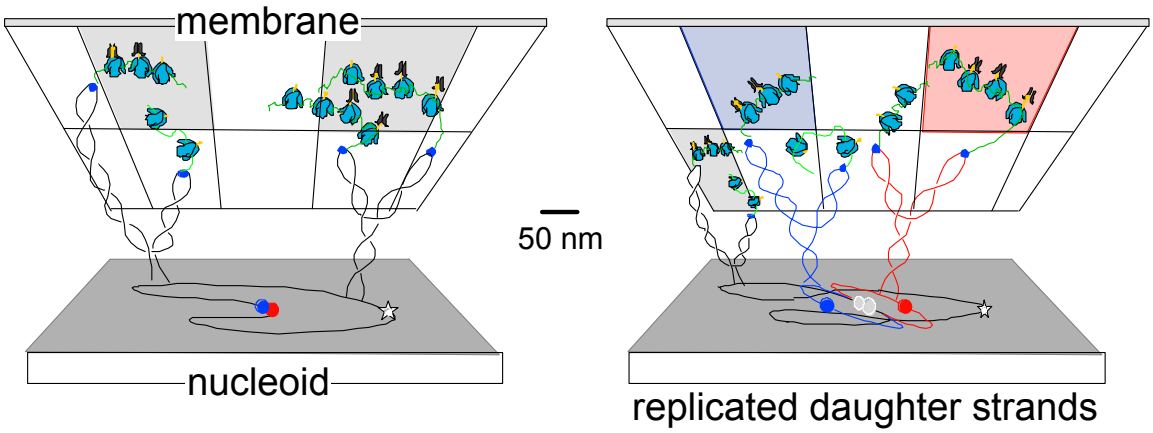


Figure 2: Schematic representation of transertion-mediated chromosome segregation. Left panel: genes on supercoiled DNA loops coding for membrane proteins are each transcribed by two RNA polymerases. Brownian movement of the DNA supercoils is constrained because the DNA is indirectly attached to transertion areas on the plasma membrane. Right panel: After initiation of DNA replication, bidirectionality is generated by the following principles: (i) transient DNA constrainment at transertion areas; (ii) competition for membrane space; (iii) cooperativity between neighbouring genes; (iv) spatial bias in placement of transertion areas formed by one daughter strand; (v) movement of bulk of non-transcribed DNA by Brownian diffusion. [for symbols see Woldringh, 2002] (see plate 5).

References

Cunha, S., C.L. Woldringh, Th. Odijk (2001) Polymer-mediated compaction and internal dynamics of isolated *Escherichia coli* nucleoids. J. Struct. Biol. **136:** 53-66.

Norris, V. (1995) Hypothesis: chromosome separation in *Escherichia coli* involves autocatalytic gene expression, translation and membrane-domain formation Mol. Microbiol. **16:** 1051-1057.

Odijk, T. (1998) Osmotic compaction of supercoiled DNA into a bacterial nucleoid. Biophys. Chem. **73:** 23-30.

Odijk, T. (2000) Dynamics of the expanding DNA nucleoid released from a bacterial cell. Physica A **277:** 62-70.

Roos, M., A.B.M. van Geel, M.E.G. Aarsman, J.T.M. Veuskens, C.L. Woldringh, N. Nanninga (1999) Cellular localization of *oriC* during the cell cycle of *Escherichia coli* as analyzed by fluorescent in situ hybridization. Biochimie **81:** 797-802.

Woldringh, C. L., P. R. Jensen, and H. V. Westerhoff. (1995) Structure and partitioning of bacterial DNA: determined by a balance of compaction and expansion forces? FEMS Micribiol. Lett. **131:** 235-242.

Woldringh, C.L. (2002) The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. Mol. Microbiol. **45:** 17-29.

Interfacial water: a modulator of the biological activity

Pascale Mentré

UMR 8646 MNHN-CNRS, Muséum national d'histoire naturelle, 53 rue Buffon, 75005 Paris, France
pmentre@mnhn.fr

Abstract

In contact with an extraneous surface, water adopts a particular structure. This water is called *interfacial water*, and although its organization depends greatly on the nature of the surface, its different forms have in common a certain number of properties: cooperativity, proton conductivity, behavior under pressure, density different from°1, selective solving power, etc. *Interfacial water is an integral partner of biomolecule and, the modulator of their activity.* Moreover, the quasi-totality of cell water is interfacial so simulation of the role of water in biological processes must not ignore the numerous and complex properties of interfacial water: *The obvious question is whether such studies are currently feasible*.

Before the seventies, only a very few biologists envisaged cell water other than as an ordinary solvent with homogenous properties, in other terms as common water (bulk water). The work of biochemists and biophysicists, particularly specialists of protein structure, drew attention to the interaction between biomolecules and the surrounding water. The organization of an interface between a liquid and a surface is a general phenomenon. But in the case of water, for reasons that we shall explain, this phenomenon plays a capital role in biology.

The modelling of interfacial water and the simulation of its participation to biological processes are still at their beginning. Water molecules cannot be considered any longer as simple balls to fill up the space inside and around biomolecules. Indeed, the relationship of water to biomolecules is now perceived as increasingly complex and, given the present limits of computers, this makes its simulation difficult.

Our training as biologists does not permit to us to evaluate the feasability of such simulation at the present time. We wish only to discuss some important properties of *interfacial water* that are frequently involved, simultaneously or consecutively, in biological processes.

1. Interfacial water polymorphism: the key to its modulator role

The forces responsible for the organization of water in the vicinity of surfaces are principally electrostatic. Water molecules are electrical dipoles (Fig. 1). Moreover, because of the strong electrostatic field developed by the widely exposed protons of their hydrogens, they can form special bonds called hydrogen bonds (H-bonds; Fig. 2; we recall that the electrostatic field developed by a particle of radius r and charge q varies as q/r^2). In bulk water, on average, the water molecules are H-bound with four neighboring molecules forming tetrahedra (Fig. 3). These edifices are dynamic: the lifetime of H-bonds in *bulk water* (common water) is in the range of 10^{-12} s.

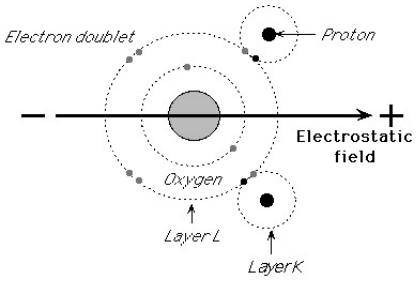


Figure 1. The water molecule (see plate 7)

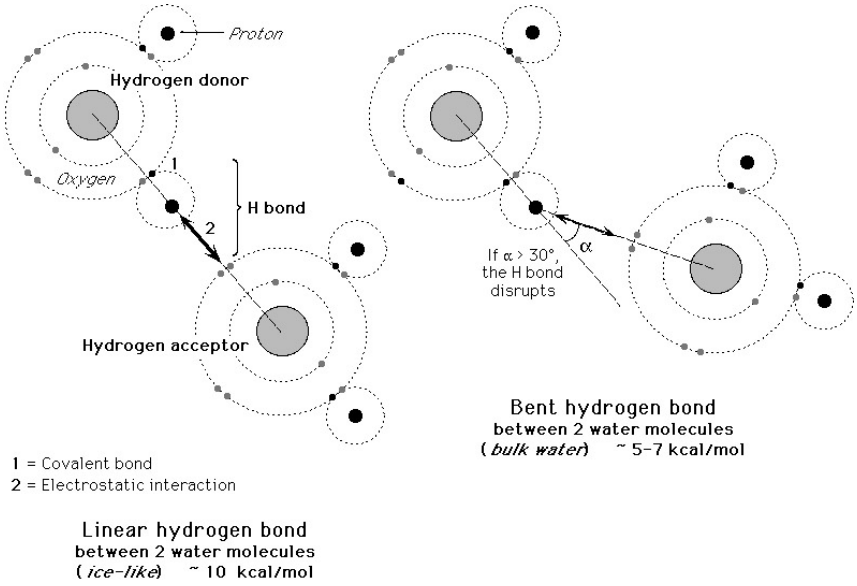


Figure 2. The hydrogen bond in ice-like structures and in bulk water. (see plate 8)

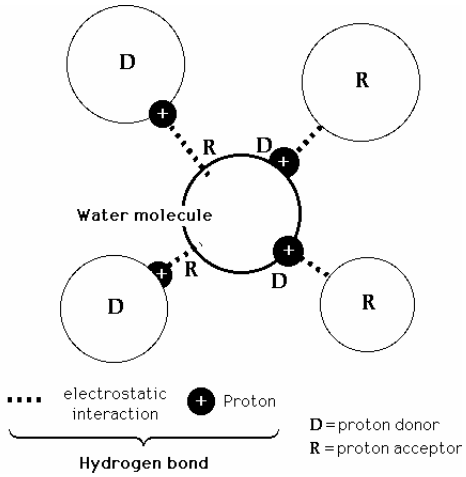


Figure 3. The tetrapolar character of the water molecule (see plate 9)

Statistically, in bulk water, there is no privileged direction. However, in *ice* and *ice-like* water, water molecules can be H-bound into lines, forming strongly cooperative edifices (Fig. 2; § 1.2, 2.1, 2.2).

In fact, the strength of H-bonds depends on the nature of the bound domains: ~10 kcal/mol in ice and *ice-like* structures (linear H-bonds), ~5 kcal/mol in bulk water (bent H-bonds), ~1 kcal/mol between water and CH aliphatic radicals (—CH···H_2O; § 1.1),).

The introduction of an extraneous substance into water ruptures the equilibrium. In contact with it, water molecules are submitted to electrical constraints –°depending on the nature of the substance°–, which privilege certain orientations: this involves the formation of a zone, called *interfacial water*, which is fundamentally anisotropic and which can extend beyond the thickness of four layers of water molecules (one-layer ≈°0.3°nm). Interfacial water is also known as *structured water* but must generally be distinguished from bound water [14].

The structure of interfacial water associated with biomolecules, and therefore its properties, are very varied and reflect the diversity of the biomolecule domains with which it is in contact [14].

1.1. Water associated with hydrophilic domains

These domains attract water because they generate electrostatic forces. They comprise *ionized* domains and *polar* domains (Fig. 4). Hydrophilic substances can be either merely wetted (e.g. glass) or dissolved (e.g. salt, sugar). In both cases, water forms a *hydration shell* on contact with them. This shell is responsible for their maintaining solutes in solution because it screens the electrostatic forces.

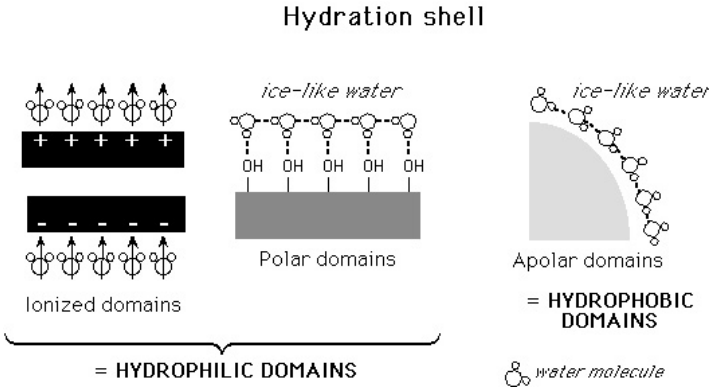


Figure 4. Principal types of organization of water on surfaces (see plate 10)

Water dipoles are attracted by *ionized domains* and arrange normally to their surface, forming *electrostricted water* in which pressure can reach 34 kbar and density 1.2 [24] (see plate 11). Water molecules form H-bonds with *polar domains*, as hydrogen donors or acceptors, forming a kind of water that is less dense than bulk water [16].

It must be made clear that molecules or radicals generally considered as apolar can also form H-bonds with water. Water molecules can thus bind with the clouds of π electrons surrounding the conjugated double bonds in hydrocarbons [25]. This could help stabilize the position of certain aromatic aminoacid residues at the surface of proteins. Moreover, carbons of aliphatic-chains, acting as hydrogen donors, can bind with water molecules. It has been shown that —CH···H_2O assemblies play an important role in the structure of ADN – forming

bridges between bases within a same strand or between two strands – and proteins – stabilizing β-sheets and α-helices [29].

1.2. Water associated with hydrophobic domains

Contrary to popular belief, hydrophobic domains do not repel water molecules. But, not being polar, they do not attract them either. Therefore they do not have solvation shells. This absence of solvation shells explains the non-miscibility of hydrophobic molecules with an aqueous phase: they can approach each other close enough to be aggregated by van der Waals forces [27]. However, water molecules in contact with hydrophobic surfaces adopt a particular arrangement, called *hydrophobic hydration*, to compensate for the rupture of equilibrium due to the presence of these extraneous surfaces. They align with H-bonds tangential to the surfaces, which results into a particular form of water often named *ice-like water* and even *icebergs* [5]. This water does indeed resemble ice in some of its properties such as particularly its formation which is exothermic with a negative variation of entropy. That said, it is not really ice. Its density is globally higher than 1, (for a review: [16]), although probably lower than 1 in the first layer at the direct contact of the hydrophobic surface [31]. The lifetime of its H-bonds is in the range of 10^{-6} s.

It is well known that hydrophobic hydration plays a key role in the stabilization of the structure of biomolecules [19].

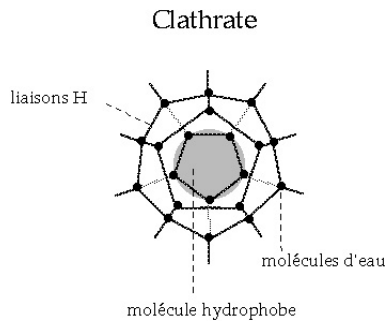


Figure. 5. Clathrate A clathrate results from the encaging of an apolar domain by water molecules (see plate 12).

Interestingly, the arrangement of water molecules in contact with apolar substances depends narrowly on the geometry of the exposed surfaces. In the case of surfaces with small radii of curvature (small molecules: methane, ethane, etc, or radicals: methyl, ethyl, etc; rare gases), water molecules do not have any possibilities of making H-bond with the surfaces. The result is the formation around the surfaces of little pentagonal-faced structures, whose vertices are water molecules and sides H-bonds. This arrangement is called *clathrate* (Fig. 5, [8]). There are many variants of this structure, according to the form and the curvature of the hydrophobic surfaces [3,26]. These variations could make hydrophobic hydration an excellent way of transferring signals through phospholipid membranes [28].

Moreover, a recent study by simulation suggests that the narrow spaces between hydrophobic domains could allow water to circulate rapidly through the crowd of macromolecules in the cell. Indeed, simulation shows that water molecules diffuse more rapidly through a hydrophobic nanotube of carbon, only 8 Å in diameter, than in bulk water [7].

1.3. Water confined inside biomolecules

Water molecules can be trapped in the interior of macromolecules with lifetimes in the order of 10^{-3}°s or more. They generally form strongly organized water inside hydrophobic recesses but they can also be associated with hydrophilic domains. This water contributes to maintaining the three dimensional structure of macromolecules and is also likely to be involved in the coordination of different domains of the macromolecule, for example, by proton currents or propagation of mechanical signals (§ 2.2, 2.4).

1.4. Cell water : a thin film of highly polymorphous interfacial water

Water constitutes about 80% of the total cell mass. But for a simple reason of surface/volume ratio it is distributed into only two layers of molecules around every biomolecule – °statistically speaking. If h is the mass of the first layer of hydration and M the mass of the biomolecule, $h/(h+M)$ is about 35% in the case of globular proteins and more than 60% in the case of elongated molecules (such as nucleic acids and fibrous proteins). These values, which were obtained by a variety of experimental methods, are corroborated by a very simple calculation (Fig. 6: the thickness of one layer of water molecules is ~0,3 nm and the average diameter of a globular protein 3.5 nm).

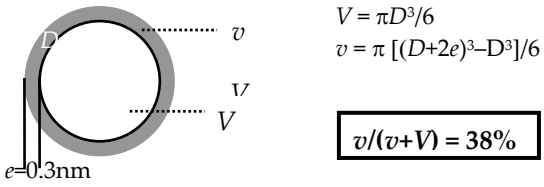


Figure 6. The average diameter of a protein is 3.5 nm and the thickness of a layer of water molecules is 0.3 nm. Consider a sphere, 3.5 nm in diameter covered by a 0.3 nm-thin layer (the thickness of a). V and v being respectively the volumes of the sphere and the layer, a very simple calculation shows that $v/(v+V)$ is 38% (see plate 13).

Therefore the quasi-totality of cell water is interfacial and is distributed in the form of a thin film, ~1.1°nm thick, that percolates through the crowd of biomolecules [12,14,15] and that has a structure reflecting the heterogeneity of these molecules.

2. Key properties of the modulator role of interfacial water in cell

It is well established that the behavior of water in contact with hydrophilic and hydrophobic domains is one of the keys to the three dimensional structure of biomolecules. It now appears that the variations of structure of interfacial water –°whose the list does not cease increasing°– confers a degree of sophistication to the relations between water and biomolecules that was unsuspected until recently.

2.1. Cooperativity

Hydrogen bonds are cooperative: H-bonding *encourages* water molecules to make stronger and shorter H-bonds with other polar molecules [21]. Cooperativity is maximum when water molecules are arranged into lines: H-bond energy ~ 10 kcal/mol in ice and ice-like water instead of ~5 kcal/mol in bulk water.

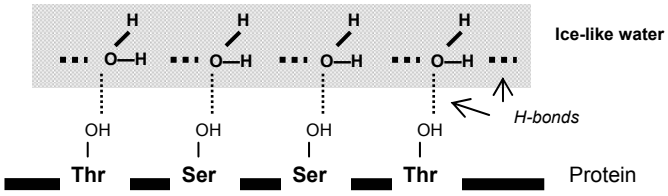


Figure 7. Ice –like water along a peptidic chain.

Relatively long lines of water molecules are not rare in interfacial water: along hydrophobic surfaces (§ 1.2) or along sequences of polar aminoacids (Fig. 7).

Interfacial water has a higher heat capacity than bulk water because more energy is necessary to break its H-bonds. This involves a higher thermal stability of the biomolecules and biomolecular complexes. Similarly, interfacial water is difficult to freeze since to be reorganized into ice it should be simultaneously heated, to break its H-bonds, and cooled, to produce ice crystals. Interfacial water, thus, through its high cooperativity, provides an efficient thermal protection against heat and chill.

Cooperativity is the key of many regulatory processes which have a sigmoid "all or nothing" profile. They display a critical threshold that corresponds to a sudden reorganization of H-bonds in interfacial water and that is correlated with a change in the underlying structure of the biomolecule; such cooperativity is therefore involved in allostery and in the stability of proteins below a critical temperature, etc. [4,19].

2.2. Proton currents

Ice-like lines of water molecules are good proton conductors. The H-bonds hop successively from the first to the last water molecule, simulating a displacement of protons (Grotthus' mechanism; Fig. 8, [6,8,34]).

It has been shown that the activity of lysozyme coincidates with the circulation of protons along water molecules associated with its surface [2]. Biomolecules could also exploit proton currents along ice-like water trapped in hydrophobic domains (§ 1.2, [7,23]).

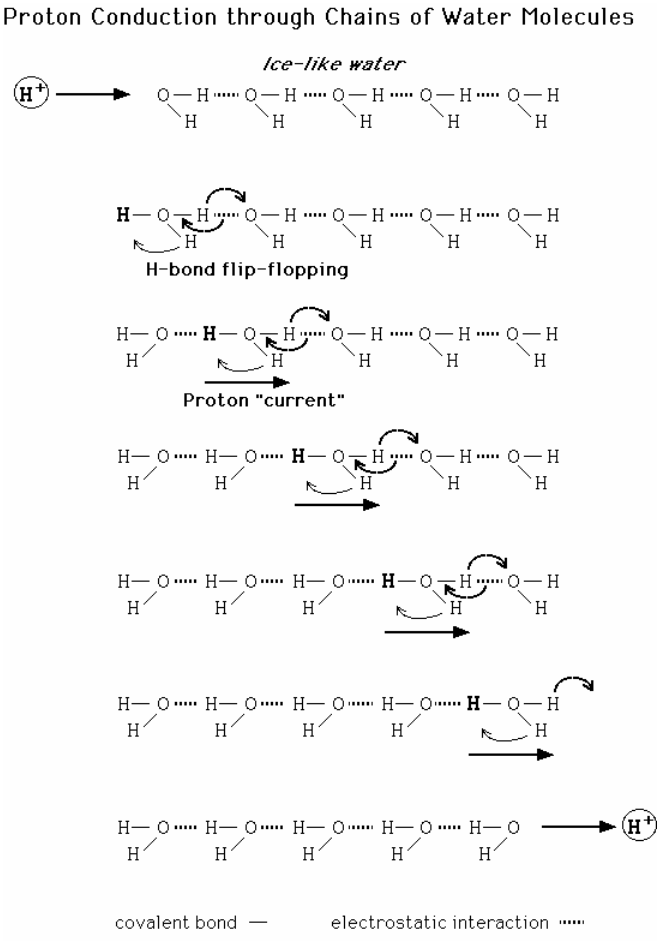


Figure 8. Grotthus' mechanism of proton conduction along ice-like water

2.3. Effects of osmotic pressure and hydrostatic pressure

Interfacial water is relatively insensitive to variations of osmotic pressure, as has been demonstrated by many experiments using chemically inert osmotic agents (osmolites) such as ethylene, glycol, glycerol, etc.). An effect is observed only when the osmolites are sufficiently concentrated (1 M = 22,4 atm) to enter into competition with the powerful local constraints produced by the surfaces of biomolecules. Under pressure of 20-100°atm, interfacial water is extracted, at least partially, which favours changes of configuration and interactions between biomolecules [20].

On the contrary, under high hydrostatic pressure, in agreement with the principle of Le Chatelier, the proportion of interfacial water increases because its density is globally higher than 1 [16]. The consequence is the weakening of ionic and hydrophobic interactions in biomolecules and biomolecule complexes (§ 1.1 and 1.2; Fig. 9). Under high hydrostatic pressure (10^2-10^3°atm), cytoskeleton, viruses, antigen-antibody complexes dissociate, reversibly (for a review: [16]).

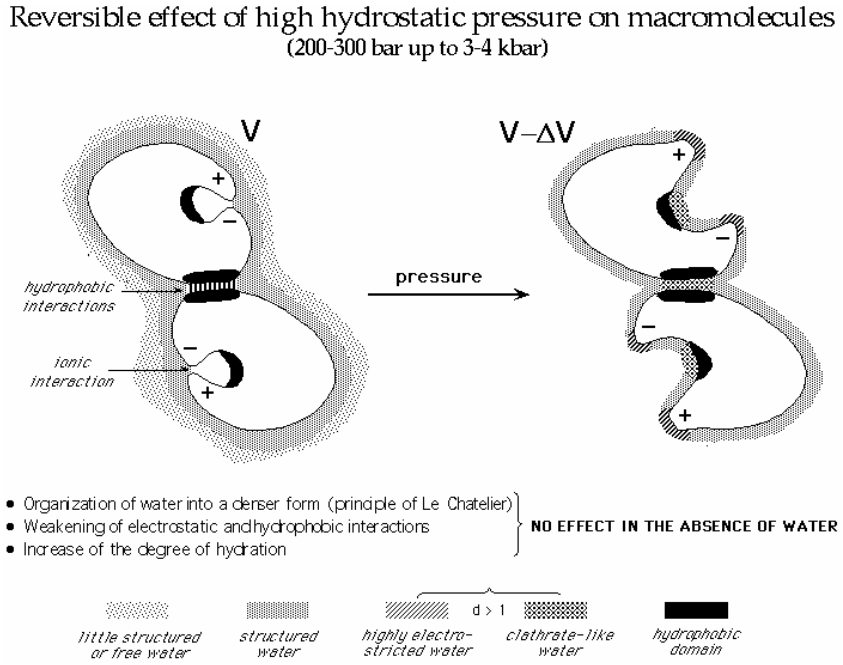


Figure 9. According to the principle of Le Chatelier, high hydrostatic pressure favorizes the transitions which are accompanied with a negative variation of volume (ΔV). Therefore biomolecules and biomolecule assemblies tend to expose more surface, because interfacial water, principally in its form associated to ionized and hydrophobic domains, is denser than bulk water. The consequence is the weakening of ionic and hydrophobic interactions.

Robinson and Sligar [20] used the antagonist properties of osmotic and hydrostatic pressures to analyze the recognition of DNA by enzymes of restriction and the Star activity (Fig.10).

Osmotic pressure of cell, or cell compartments, is a common notion. But, of course, the values obtained by the classical macroscopic methods are only *global*. It is difficult to believe that a cell possesses a uniform osmotic pressure since it is filled with a huge diversity of biomolecules which do not have, at a given time, the same hydration requirement to function properly. It would be better to imagine that osmotic pressure varies from a point to another one, even in the absence of any membrane. As was explained by Wiggins [30], osmotic equilibria from one layer of interfacial water to the next can be as different as though these layers were separated by a membrane.

We therefore believe that local variations in osmotic pressure in a compartment can be considerable and may be in the range of the variations which produce effects in vitro (up to 10^3 bar). Indeed, the very high pressures which are required in vitro to obtained effects identical to those observed in living cells, support the idea that these conditions must exist in cells, *locally*, to modulate the activity of biomolecules. It must not forgotten that at the nanometer scale, the distribution of the thermodynamical values is far more dispersed than at our scale [1,12].

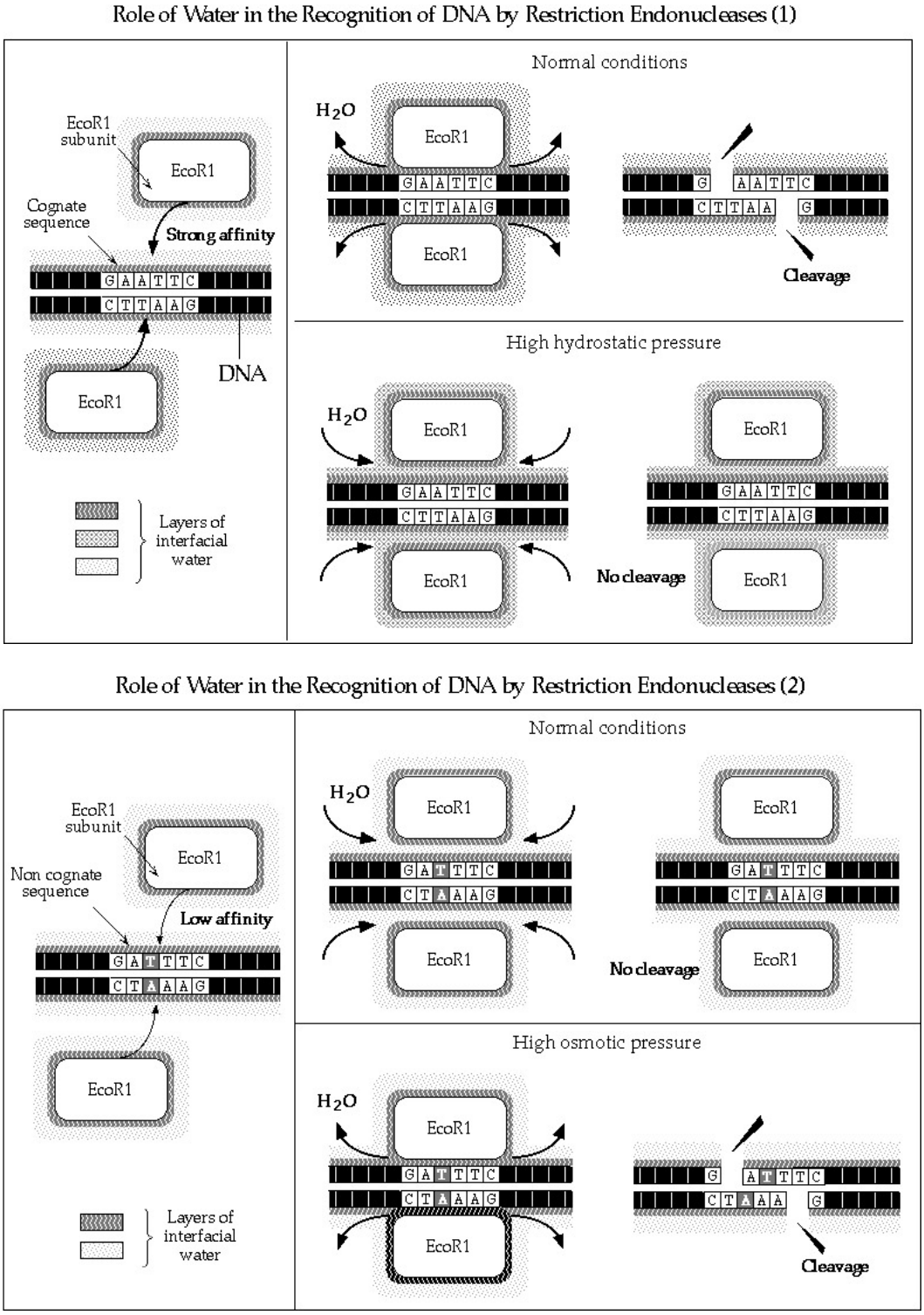


Figure 10. Role of water in the recognition of DNA by restriction endonucleases. (1) DNA with a cognate sequence which is normally recognized and cleaved by EcoR1. Under high hydrostatic pressure interfacial water between DNA and EcoR1 cannot be expulsed and there is no cleavage. (2).DNA with an erroneous sequence which is not recognized by EcoR1. But under high osmotic pressure (in the range of 10^{2}°bar) interfacial water between DNA and EcoR1 can be expulsed enough to permit the interaction and the cleavage. This phenomenon is called STAR effect (Scheme from the experiments of Robinson and Sligar [20]).

2.4. Variations of density

Biomolecules and biomolecular complexes expose and shield many of their domains whilst working and this involves modifications of the surface they expose to their aqueous environment. As interfacial water can have different densities in contact with different domains (§ 1.1, 1.2), these modifications produce variations of volume and therefore offer the possibility of mechanical effects. These effects are exploited in molecular motors [17]. It could also be envisaged that variations of volumes are involved in the transduction of signals along chains of biomolecules: variation of volume accompanying change of configuration of a biomolecule inducing change of configuration of the following one, etc. This process could produce "ion currents" without real transport of ions, by propagating a ion liberation-sequestration mechanism (Fig. 11).

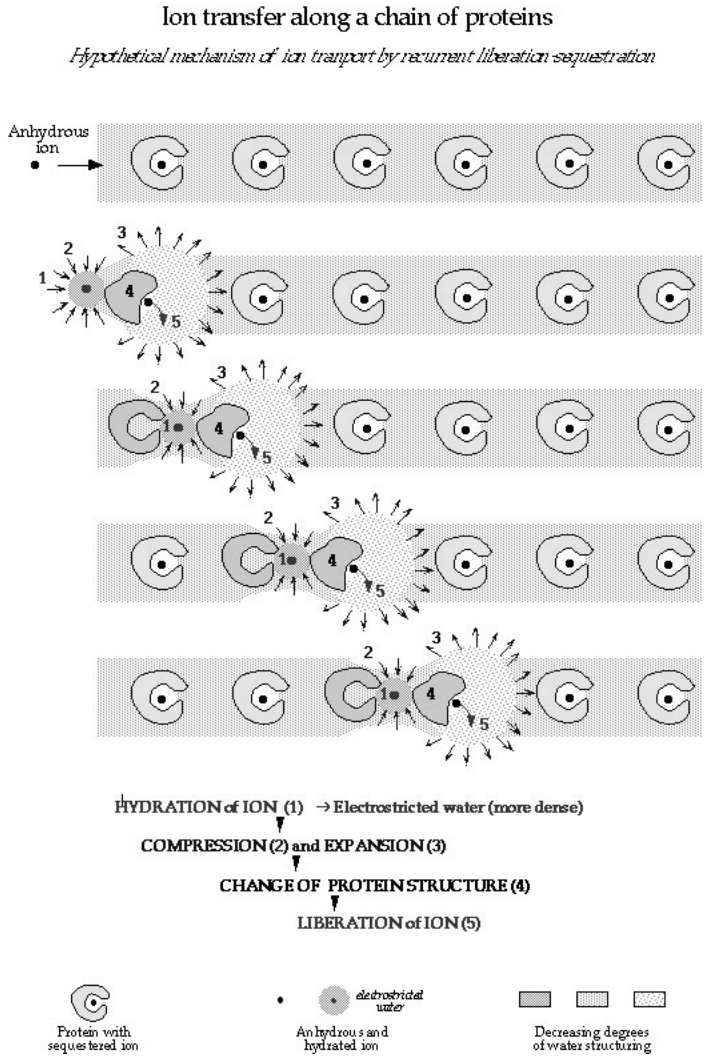


Figure 11. Hypothetical mechanism of an "ion current".

2.5. Selective exclusion of ions

Interfacial water excludes ions, selectively, according to the series of Hofmeister [9,18,33]. Figure 12 shows the Hofmeister's series in the case of cations. The left part of the series corresponds to the ions which are the most excluded because they are the biggest in their hydrated form. As a consequence of their higher charge and/or their smaller size, they develop a more intense electrical field (§ 1) and therefore retain a bigger amount of water. They powerfully order water molecules around them. For this reason, they are called *structure-making* or *kosmotropic ions.* The right part of the series corresponds to the ions which are the smallest in the hydrated state and therefore the less excluded. Far from ordering water around them, they *break the structure* of interfacial water by their thermal motion. They are called *chaotropic ions*.

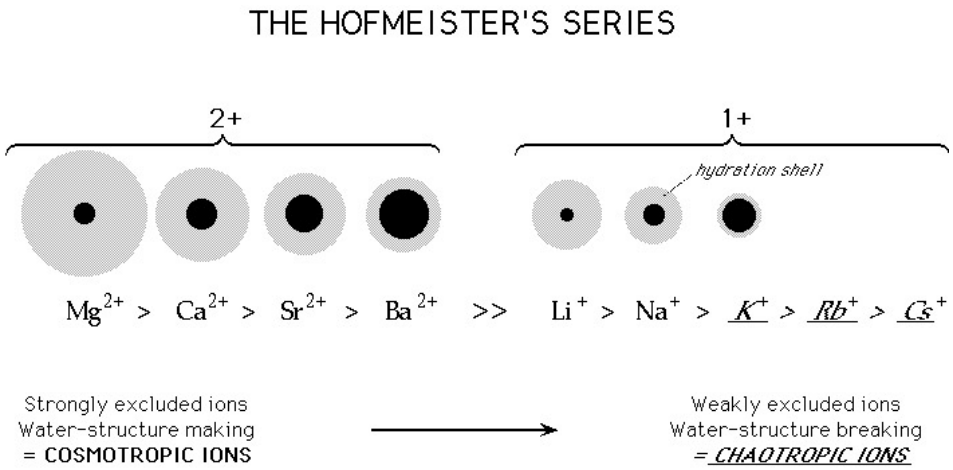


Figure 12. The Hofmeister's series.

The antagonist effects of potassium and sodium on interfacial water would play a key role in many processes. Particularly, their alternate entry in the hydrophobic enzymatic clefts of Na-K dependent enzymes, by involving respectively water-structure breaking and making, could produce the alternate opening and closing of these pockets [32].

For more than half a century, exclusion of ions by cell water has been the starting point of a quarrel between the supporters of the sodium pump theory and its opponents. The cell is rich in potassium and poor in sodium. Conversely, extracellular medium is poor in potassium and rich in sodium. The supporters of sodium pump claim that ion balance is maintained in cell by an ATPase, located in plasma membrane, called a sodium pump, which exchanges sodium and potassium, against their concentration gradients, between cell and extracellular medium, with an expenditure of energy in the form of ATP [22]. Their opponents think that the cell does not need any such pumps since the sodium ion is sufficiently to the left in the Hofmeister's series to be spontaneously excluded from cell water [10,11,18].

Conclusion

Interfacial water is an integral partner of biomolecules, the modulator of their activity. The quasi-totality of cell water is interfacial. Study by simulation of the role of water in biological processes must not ignore the properties of interfacial water, which are numerous and complex. *The obvious question is whether it is now feasible to undertake this study*. It depends essentially on the performances of computers. Interesting results have already been obtained for submolecular domains. We think that levels of higher complexity will be accessible soon (relations between domains inside of the same biomolecule, from one biomolecule to its neighbour, from a biomolecular complex to another one, etc) and may help us understand the orchestration of cellular processes by mechanical signals or proton currentss via the interfacial-water film which percolates the whole cell.

References

1 Ball P. When size does matter. *Nature* 1991, **349:** 101-102

2 Careri G, Peyrard M. Physical aspects of the weakley hydrated protein surface. *Cell Mol Biol* 2001, 47: 745-756

3 Cheng YK, Rossky PJ. Surface topography dependence of biomolecular hydrophobic hydration. *Nature* 1998, **392:** 696-699.

4 Colombo MF, Rau DC, Parsegian VA. Protein solvation in allosteric regulation: a water effect on hemoglobin. *Science* 1992, **256:** 655-659.

5. Franks HS, Evans MW. Free volume and entropy in condensed systems. III. Entropy in binary liquid mixtures, partial molar entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *J Chem Phys* 1945, **13:** 507-532.

6 Gascoyne PR, Pethig R, Szent-Gyorgyi A. Water structure-dependent charge transport in proteins. *Proc Natl Acad Sci U S A* 1981, **78:**261-265

7 Hummer G, Rasalah JC, Noworyta JP. Water conduction through the hydrophobic channel of a carbon nanotube. Nature 2001, 414: 188-190.

8 Klotz IM. Water. In: Horizon in biochemistry. Albert Szent-Györgyi dedicatory volume. Kasha M, Pullman B, eds, Academic Press, NY, pp 523-550.

9 Ling GN. A physical theory of the living state: application to water and solute distribution. *Scanning Microsc* 1988, **2:** 899-913.

10 Ling GN. The membrane pump theory. In: *A revolution in the physiology of the living cell*. Ling NG, Ed, Krieger Publishing Company, Malabar, Florida, 1992a, pp 9-30.

11 Ling GN. Cell water. In: *A revolution in the physiology of the living cell*. Ling NG, Ed, Krieger Publishing Company, Malabar, Florida, 1992b, pp 69-110.

12 Mentré P. Remise en question de la diffusion dans la cellule. In: *L'eau dans la cellule. Une interface hétérogène et dynamique des macromolécules*. Masson, Paris, 1996, pp 159-183.

13 Mentré P. L'eau, une interface hétérogène et dynamique des macromolécules. In: *L'eau dans la cellule. Une interface hétérogène et dynamique des macromolécules*. Masson, Paris, 1996, pp 267-275.

14 Mentré P. An introduction to « Water in the Cell »: tamed Hydra ? *Cell Mol Biol* 2001, **47**: 709-715.

15 Mentré P. L'eau dans la cellule vivante. Etat physique, fonctions. In : *Propriétés de l'eau dans les aliments*. Le Mesle M. ,Lorient D, Simatos D, eds., Tec & Doc Lavoisier, Paris, 2002, pp 85-109.

16 Mentré P, Hui Bon Hoa G. Effects of high hydrostatic pressures on living cells: a consequence of the properties of macromolecules and macromolecule-associated water. *Int Rev Cytol* 2001, 201: 1-84.

17 Oplatka A. The role of water in the mechanism of muscular contraction. *FEBS Lett* 1994, **355:** 1-3.

18 Pollack GH. *Cells, gels and the engines of life: a new unifying approach to cell function*. Ebner and Sons, Seattle, 2001.

19 Privalov PL. Thermodynamic problems of protein structure. *Annu Rev Biophys Biophys Chem* 1989, **18:** 47-69.

20 Robinson CR, Sligar SG. Hydrostatic and osmotic pressure as tools to study macromolecular recognition. *Methods Enzymol* 1995, **259:** 395-427.

21 Stillinger F.H. Water revisited. *Science* 1980, **209:** 451-457.

22 Skou JC. Overview: The Na,K-pump. *Methods Enzymol* 1988, **156:** 1-27.

23 Smondyrev AM, Voth GA. Molecular dynamics simulation of proton transport through the influenza A virus M2 channel. *Biophys J* 2002, **83:**1987-1996.

24 Stranks DR. The elucidation of inorganic reaction mechanisms by high pressure. *Pure Appl Chem* 1974, **38:** 303-323.

25 Suzuki S, Green PG, Bumgarner RE, Dasgupta S, Goddard WA, Blake GA. Benzene forms hydrogen bonds with water. *Science* 1992, **257:** 942-945.

26 Udachin K.A., Ripmeester J.A. A complex clathrate hydrate structure showing bimodal guest hydration. *Nature* 1999, **397** : 420-423.

27 van Oss CJ. *Forces interfaciales en milieu aqueux*. Masson, Paris, 1996.

28 Vigh L, Maresca B, Harwood JL. Does the membrane's physical state control the expression of heat shock and other genes? *Trends Biochem Sci* 1998, **23:** 369-374.

29 Wahl MC, Sundaralingam M. C-H¨¨O hydrogen bonding in biology. *Trends Biochem Sci* 1997, **22:** 97-102.

30 Wiggins PM. Osmosis and micro-osmosis. *New Zealand Sci. Rev*. 1994, **51:** 79-84.

31 Wiggins PM. High and low density intracellular water. *Cell Mol Biol* 2001, **47:** 735-744.

32 Wiggins PM, MacClement BAE. Two states of water found in hydrophobic clefts: their possible contribution to mechanisms of action of pumps and other enzymes. *Int Rev Cytol* 1990, **108:** 249-304.

33 Wiggins PM, van Ryn RT. Changes in ionic selectivity with changes in density of water in gels and cells. *Biophys J* 1990, **58:** 585-596

34 Williams RJ. The problem of proton transfer in membranes. *J Theor Biol* 2002, **219:** 389-396.

Hybrid system to model biological objects: application to mitochondria organelle.

Marie Beurton-Aimar[1,2], Sabine Pérès[1], Pascal Ballet[3], Abdallah Zemirline[3], Christine Nazaret[4], Jean-Pierre Mazat[1].

[1]Laboratoire physiologie mitochondriale, Université Bordeaux 2, 146 rue Léo-Saignat, F-33076, Bordeaux Cedex
[2]LaBRI, Université Bordeaux 1, 351 cours de la Libération, F-33405, Talence Cedex
[3]Departement Informatique, EA 2215, Université de Bretagne Occidentale, BP 809, F-29285 Brest Cedex
[4]Ecole Supérieure de Technologie des Biomolécules de Bordeaux, 146 rue Léo-Saignat, F-33076, Bordeaux Cedex

Abstract

Nowadays, many researches are focus on biological objects design in computer programs. It is obvious that the production of a wide quantity of data in biology and their complexity, impose to develop new methods where computing procedures are not only calculus procedures but a whole environment where semantics and complex links between data can be expressed. Hybrid systems are able to take into account the design of complex data [1] and to apply on it many different procedures: linear or differential equations, regulatory graphs or simulations from multi-agent models. Object oriented models are well known to be adapted to design such systems and so object oriented methods have been used to design the architecture of the system. We have used a modified version of SBML (a structured language for biology) to describe data. The obtained modelel has been applied on mitochondria data. This organelle which is the main research subject of our laboratory, is not very well represented in the different international projects more focused on the whole cell. Our goal is to build a system where we can put as more as possible all the knowledge about mitochondria (from our own experimentation or from literature) inside our system and to propose to use it with some known mathematical models or with new methods like interaction graphs or multi-agents systems. The first processes which are implemented describe the activity of the TriCarboxyllic Acid Cycle and the respiratory chain.

1 Introduction

Two main families of software are distinguished to model and manipulate biological objects, those to design structural and static information (mainly Data Base Manager System), and those to offer creating computing procedures (differential equations, stochastic process, inter action graphs, etc) to simulate experimental results. The main goal to hybrid systems is to merge these two approaches and to offer integrated systems which either manage static descriptive information and either allow to infer several kind of treatments.
The first thing to do is to choose a data language description, we will present SBML, a XML-like language and we will show modifications needed to satisfy our constraints. After that, we will discuss why the integration of several computing procedures types can improve simulations. Finally, we will show how to build composition of modules with object oriented methods, and so, to design architecture for such hybrid systems.

2 Modelling language

Usual simulation softwares are based on mathematical equations and the interface of such programs allows users to insert their own formula and/or matrix of data, lists of parameters and so on.

But generally, each treatment has its own design and it is needed to input data in a new format when the computing procedure changes. This operation is time consuming and moreover, it is possible to make mistake during re-copying operation. It is important to note that even if the transfer between procedures is correct, there are no guarantee that same data have the same semantics in the whole set of procedures because, in general, semantic is not explicit. In fact, data semantic is hardly ever explicitly given and the users must perform it themselves. This point is rarely treated but in our mind, it is a major point of mis-working when different results are obtained from the same set of data manipulated by different procedures. One of the solutions can be to express semantic with data types and structured objects.
The need to describe data (structures and value) is addressed by XML-like languages. XML (eXtended Markup Language) allows to put into a text file all the information about the data. It is a structured language with marks (TAGS) defining concepts or object classes. SBML [2] (System Biology Markup Language) is an augmented version of XML including TAGS for biology and specifically for describing metabolic process, kinetic equations ...

The SBML architecture get as the root frame the TAG `model` and all the components are associated to this frame. A `model` is defined by a name, a set of species which participates to the reactions and a set of mathematical formula (with its parameters). In fact the `model` is a specific version of a procedures set only. So the lack of abstraction like generic biological function forbids to re-use definition model and to instantiate it with another computing procedures.

Our suggestion is to modify some parts of SBML architecture model and to add some new concepts to be able to make composition of modules. Major modifications are introduction of a hight level component to contain data hierarchy and set of computing models. The TAG **enzyme** has been added, with such entity it is possible to give characteristics of a specific enzyme or to reason from enzyme class (with EC classification). We have also added some components to make easier procedures composition. The new hierarchy is presented in figure 1.

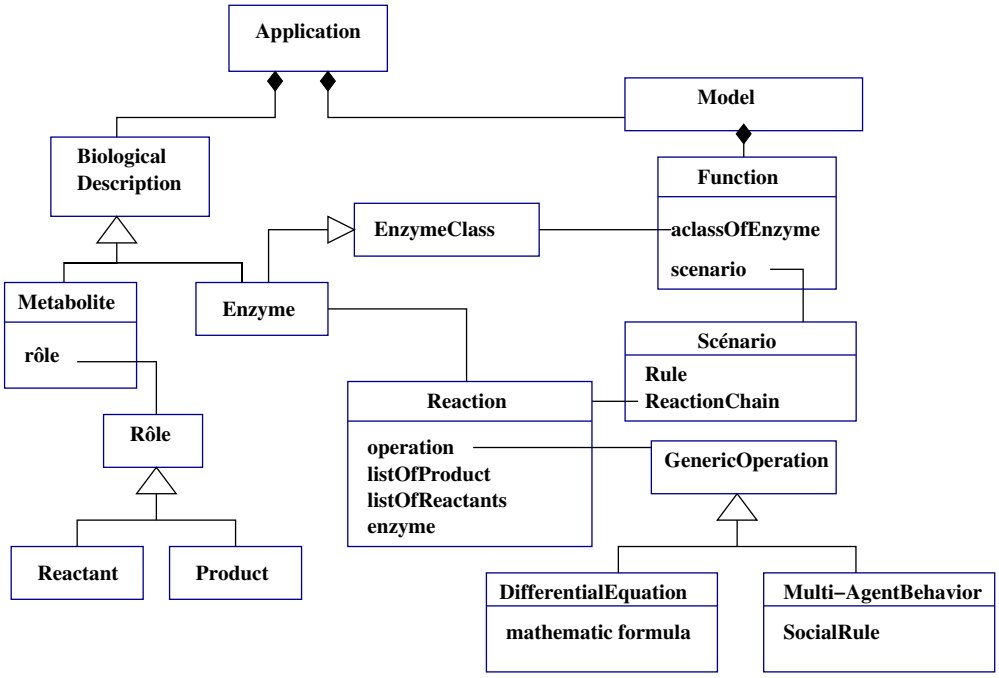


Figure 1: the whole object oriented model

3 Why to use several methods to compute

Nowadays, many kinds of treatments exist: differential equations or stochastic process, to simulate results of *"in vitro"* experimentations, graphs and feedback circuits to model regulations networks, social models like multi-agent systems to simulate behavior of complex systems. A better understanding of a whole biological process can be obtained by a composition or a collaboration between several levels of Modelling.
Object oriented models are well-adapted frames to design data and treatments at a sufficient level of abstraction and to use data in a different way (mathematic, qualitative ...) inside the same system. Languages like SBML fall to help designer to build very complex systems. Their representation is *flat* and they have no tools to model behaviors. So, we look forward UML (Unified Modelling Language) which have many diagrams to express constraints in time, control structures and collaboration between entities of the system. UML is the newer object oriented tool to standardize descriptions and give a rich interface to express complex behaviors. We have choose to use it in a first step of creating architecture and design classes for all biological objects. In a second step, the external description is exhibit in a modified SBML taking into account our modifications. The weaver of these tools and models would give us new hybrid systems more adaptive to the biologist needs. The figure 1 shows several classes which are implied in the composition of treatments. We have added the concept of `scenario`. An instance of `scenario` contains a list of `chain of reactions` which is the realization of a particular metabolic network. A `reaction` can be instantiate by different type of `operation` following the type of used calculation. The `function` concept allows to think over the *enzyme class* level.So it is possible to describe a generic function like *kynase* and to change some parts of the reaction following the species or the tissues. The introduction of these levels allows to compose new models by composition and re-using of existent ones.

4 Application to mitochondria

Mitochondria is an intra cellular organelle with its own metabolic networks, DNA and specific biological functions: oxydative phoshorylation, respiratory chain ... We have begun to design TCA cycle with our model. TCA is an example of multi-composition procedure requirements. Literature can give us almost 20th possible equations inside TCA (see KEEG web site [3]) and we can test several compositions of these equations. Such models like Wright model [4] use linear equations. But now we want to add a description of system behaviors with multi-agent tools. J. Ferber defined two level of description: physical like capability to an agent to aggregate itself with another, and social to describe constraints between two basic agents. Multi-agent systems are well-known in description of population behaviors, we want to adapt them to model intern behavior of an organism)or organelle). All the biological objects can be viewed as agent which communicate and exchange information with the others depending on their environment. Parameters from mathematical models can be traduce in value for agents. Running multi-agents systems can show emergence in system behaviors which cannot be directly deduce from basic agent behaviors. All these procedures must be tested in future. In a second time, our goal is to add respiratory chain to the TCA cycle and so observed how the two sub-systems are competitive or collaborative in the mitochondriale metabolism.

5 Conclusion

The great complexity of biological systems impose to design adaptive systems. hierarchy of components and tree structures are the first elements to take into account complexity of compositional

behaviors or organization. Separation between data and treatments are the second principle to respect and Object Oriented Methods are advised tools to reach this goal. The use of marked-up language like SBML to go out the description of the whole knowledge and put it into text file without any link with programming language is an evident way to allow exchange and multiplicity of treatments. So, from these bases we can include several treatments and test how they can collaborate.

References

[1] M. Beurton-Aimar, B. Le Blanc.
A Decision Making System in Inflammatory Rheumatology using Different Reasoning Methods . International Journal of Computers and their Applications, 6 (2): 73-77, 1999.

[2] M. Hucka1, A. Finney1, H. M. Sauro1, et al.
The Systems Biology Markup Language (SBML): A Medium for Representation and Exchange of Biochemical Network Models, Bioinformatics, 2002.

[3] KEEG web site: http://www.keeg.org

[4] Albe KR, Wright BE
Systems analysis of the tricarboxylic acid cycle in Dictyostelium discoideum. II. Control analysis., J. Biol. Chem. 267: 3106-3114, 1992.

Questions for cell cyclists

Vic Norris[1], Anthony Delaune[2], Cedric Derue[1], Armelle Cabin-Flaman[1], Maurice Demarty[1], Yohann Grondin[2], Guillaume Legent[1], Chantal Monnier[1], Helene Pollard[4] and Derek Raine[2]

[1]Laboratoire des Processus Intégratifs Cellulaires, CNRS UMR 6037, Faculté des Sciences et Techniques, Université de Rouen, 76821, Mont-Saint-Aignan Cedex, France
[2]UMR 6522 Polymeres, Biopolymeres, Membranes, université de Rouen, 76821 Mont-Saint-Aignan Cedex, France
[3]Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK
[4]Genopole Recherche, 2 rue Gaston Crémieux, CP 5723, 91057 Évry Cedex, France

Author for correspondence: Vic Norris, vjn@univ-rouen.fr

There are respectable questions that are asked about the bacterial cell cycle, though not as frequently as one might like. What is the nature of the signal that initiates chromosome replication? What ends the period of sequestration of newly replicated origins of replication? What couples chromosome replication/segregation with cell division? What initiates cell division? What is responsible for timing the formation of the FtsZ ring and for positioning it at the equator or elsewhere? Is there an inactivation mechanism that prevents multiple rounds of cell division? Biochemists and molecular biologists often try to answer these questions in terms of biochemistry and molecular biology. Such attempts will fail if the concepts needed are to be found in other disciplines. There are also less respectable questions that are not generally asked. What is the cell cycle *for*? Does it have some other function than simply increasing the number of genetically identical cells? If cells initiate chromosome replication at a particular mass (origin/mass etc.), why one particular mass rather than another? Are 'key' regulator proteins simply the messenger boys instructed by the dynamics of structures? Are they just part of a molecular overlay that behaves as a coupled oscillator? What *is* a cell? Satisfactory answers to the problems of the bacterial cell cycle might be expected to be answers to other problems in cell biology and to elucidate for example the eukaryotic cell cycle, nutritional shiftup, or the origin of life. Deeply satisfying answers might be founded in the nature of cells themselves. Here we conduct a magical mystery tour of a few speculative ideas.

Introduction

The nature of the regulation of the cell cycle has been one of biology's more profound mysteries. For several years, it seemed that the bacterial cell cycle had been largely understood. Initiation of chromosome replication resulted from either the dilution of a repressor or the accumulation of an activator. Chromosomes were segregated by the growth of the peptidoglycan wall between the origins of replication that attached them to this envelope. Cell division resulted from invagination of the wall. The discovery that cell could happily replicate scores of minichromosomes should have sunk the simple models for initiation of replication. Worse, the signal responsible for the initiation of the replication of minichromosomes continues 'as normal' even when the replication of the chromosome itself is random (Eliasson and Nordstrom, 1997). Despite this, these early models have not been replaced. The mysteries underlying chromosome segregation and cell division have fared better due to the discoveries of proteins with similar activities and/or structures to their eukaryotic equivalents; indeed a new paradigm for division site selection is now based on the oscillations of the Min proteins (Meinhardt and de Boer, 2001). The readiness with which this Min hypothesis has been

embraced may herald a serious interest in the value of theory and in the possible relevance of concepts found in physics and physical chemistry although it may of course simply reflect a lack of critical interest. Here, we give examples of concepts and approaches that may be relevant by trying to answer a few simple questions.

What is a bacterial cell?

It has been argued that the cell cannot be understood adequately as anything other than itself (Norris, 1998b). In other words, the cell cannot be reduced to a single simplified model system without losing its essence. Although we have to simplify our visions of the cell in order to study it, it must be stressed that the cell is far more than any of the definitions below. The cell is not a neural net or an oscillating system of diffusible enzymes or a dissipative system or a set of phase-separated membranes and cytoplasm etc. Rather, the cell is the creator and the creation of an extraordinarily high density of different organizing processes that have autocatalytic relationships with one another. It is a system that produces self-organization and assembly by recruiting and dismissing a multitude of processes and molecules. Whatever does this is a cell and, for the moment, we only know one example: the biological cell. In other words, the cell is its own metaphor (Norris, 1998b).

1/ An autocatalytic network? All cells that grow have to be autocatalytic networks in which the elements catalyse their own synthesis from a starting foodset (Kauffman, 1993). In fact, the bacterial cell may contain several autocatalytic networks – could they interfere with one another? One network might, for example, consume a key element in another and, if so, how are they reconciled? Is there a biologically realistic way in which different intracellular networks might undergo a phase separation based, for example, on connectivities (see **4/**)?

2/ A device to explore phenotype space? Put differently, a cell is a compromise solution between robustness to maximise survival and efficiency to maximise growth. Cells have to both endure long periods in Hell and profit from brief periods in Heaven. To survive in Hell, they must adopt strategies that do not depend on the supply of energy whilst to flourish in Heaven, they must be prepared to squander it (and to grow rapidly rather than efficiently). This compromise solution involves, in our view, equilibrium structures, which resist dissociation in the absence of a flux of energy/nutrients, and non-equilibrium structures, which do require such a flux. One possibility is then that (1) to survive difficult times cells contain equilibrium structures that allow the resumption of key functions for growth when times improve (Minsky *et al*., 2002) and (2) to grow rapidly and distance competitors, cells contain non-equilibrium structures (Norris *et al*., 2002). Equilibrium structures generate non-equilibrium ones as cells go from a survival to a growth regime and, *vice versa*, non-equilibrium structures generate equilibrium ones as cells go from a growth to survival regime. In this approach, the cell can be described as a ratio of equilibrium and non-equilibrium structures (or hyperstructures, see below) and the question of exploring phenotype space becomes "How are equilibrium and non-equilibrium functions balanced?"

3/ A molecular form of competitive coherence? The functioning of many organisations is characterised by a succession of states; the elements of each state are generated by a competition for inclusion in the new state between two sets of influences dependent on (1) the members of the old state and (2) the progressive coherence of those selected (Norris, 1998c). Consider the problem of selecting an amateur football team each week from a larger group of potential players. One influence in deciding who plays next week is who plays this week (e.g.

due to shared transport arrangements) and the other influence is the coherence of the team (must have a goalkeeper, midfield player and striker).

4/ A particular pattern of connectivity. A cell – indeed any object – has a frontier with the exterior that can be defined by the drop in average connectivity at this frontier (that is, the bits inside the cell are more connected with one another than they are with the bits outside the cell). What is the dynamic pattern of connectivity that describes cellular processes and the relationships between them (Raine *et al*., 2003; Raine and Norris, 2001)? Is it related, for example, to the small world of metabolism and to transcription characterised by a power law with a particular slope (Guelzim *et al*., 2002). And do cells that are replicating their chromosome (or that have already replicated it) have *two* distinct patterns of connectivity corresponding to the "exponential and stationary phase" destinies of the future daughter cells?

5/ A tensegrity structure? In this architectural approach, the integrity of the eukaryotic cell is maintained by elastic filaments that ensure continuous tension throughout and rigid struts to resist local compression (Maniotis *et al*., 1997). In an approach to the origins of life in which a capacity to resist turgor and mechanical stresses is given importance, cells originated as tensegrity structures and transcription and translation originated as processes that maintain cellular integrity (Norris *et al*., 1999). In our approach to modern bacteria in which "exponential and stationary phase" daughter cells are generated by the cell cycle, we might expect correspondingly different patterns of tensegrity – but what are these patterns?

6/ A unit of subjective experience? What exactly is 'subjective experience' and how can it be related to life and indeed to the material world? How far down the scale does subjective experience go (Nagel, 1974; Norris, 1998a)? And could it play any role in intracellular organisation? What experiments could be done to decide whether such questions are stupid?

What are the functions of the cell cycle?

This depends on what a cell is. Here, we consider just two of the above definitions, namely that a cell is 1/ an autocatalytic network(s) and 2/ a device to explore phenotype space.

Following studies of artificial autocatalytic networks, it has been proposed that cell division originated in its capacity to separate different autocatalytic networks that may interfere with one another (Demarty *et al*., 2003; Segre *et al*., 2000). There may be more to it than this. In our hypothesis, firstly, the cell cycle has a major role in the generation of the phenotypic diversity that allows a bacterial *population* to be ready for a wide variety of challenges, stresses and opportunities because the members of this population are born with the appropriate phenotypes. Secondly, the cell cycle responds to an excessive development of non-equilibrium activities/structures and restores a robust ratio of equilibrium to non-equilibrium structures and functions; in other words, the cell cycle adjusts the economy of the cell to prevent it getting "over-heated" and falling victim to minor fluctuations in the environment.

There is some evidence for the natural heterogeneity of bacterial populations even in conditions of unnatural homogeneity (Kaprelyants *et al*., 1996; Tolker-Nielsen *et al*., 1998) (for other references see (Booth, 2002)). One way to generate this heterogeneity is via stochastic variation in the small numbers of certain regulatory elements (Elowitz *et al*., 2002). Another way is via the successive events of the cell cycle (Norris *et al*., 2002; Segre *et al*., 2000) and, in our hypothesis, the existence of two chemically identical chromosomes in the same cytoplasm

allows intracellular differentiation (Norris and Madsen, 1995). This is essentially because there is competition between genes for access to RNA polymerase and between mRNAs for access to ribosomes. Hence positive feedback circuits can operate whereby the expression of one of the two copies of a gene can increase its expression at the expense of the other copy (consider the analogy of two neighbouring laboratories competing for limited funding ...). Factors responsible for linking the expression of genes that serve related functions (e.g. the functions related to growth in heaven) can lead to one coherent pattern of expression associated with the one daughter chromosome whilst another pattern of functions (e.g. related to survival in hell) is associated with the other daughter chromosome (Norris and Madsen, 1995). It is then the task of chromosome segregation and cell division to put these differentially expressed chromosomes into separate cells.

What is the nature of the signal that initiates chromosome replication?

One of the key criteria to use in evaluating candidates for the signal initiating chromosome replication is whether it is easy to see how it arose. For example, a signal would be evaluated highly if it were to have the physico-chemical characteristic of pushing the dNTP monomer to DNA polymer reaction to the right because then one could see how it might have arisen. The candidates given below should be similarly evaluated.

Transcriptional sensing? If the function of DNA is to be transcribed, then we might expect a relationship between replication and transcription. For example, it might be supposed that the cell senses when DNA risks becomes limiting and therefore initiates chromosome replication. This is of course a vision of the cell in which selection is at the level of performance in an Olympics in paradise rather than in the grimmer game of survival in hell (see **Robustness sensing** below). It turns out that this simple version of transcriptional sensing is – along with a whole class of other models – wrong as shown by a key experiment in which the synchronous initiation of replication of scores of minichromes is unperturbed in cells in which the initiation of replication of the chromosome itself is seriously perturbed (Eliasson and Nordstrom, 1997).

Diversity sensing? An alternative, based on definitions 1/ and 2/ is that the growing cell senses when its diversity has diminished and initiates chromosome replication to restore this diversity (Norris *et al*., 2002). (consider the analogy of what can happen in a conservative family business when an additional boss arrives). How could such sensing of diversity (alias complexity?) occur? One way would be to have a transcription factor that could also serve as an initiation factor, for which the DnaA protein in *E. coli* would be a candidate. Liberation of the transcription factor by a reduction of the number of functions in which it is involved (i.e. reduction of diversity) could free the factor to trigger replication.

Robustness sensing? A related idea is that the cell growing in paradise invests increasingly in non-equilibrium functions and structures at the expense of equilibrium ones. Replication then restores the equilibrium structures required for survival. Again, sensing the change in intracellular conditions could be done by the availability of a transcription/initiation factor (see above). Interesting alternatives include the ratio of the constituents of RNA and DNA (or precursors of these constituents) based on the idea that fast growth is accompanied by (1) a characteristic investment in rRNA and certain species of mRNA and (2) an increasing consumption of GTP for protein synthesis. There is of course the possibility that the signal is of an entirely different nature involving, for example, a change in water structure brought about by increasing ATP hydrolysis.

What is the nature of the signal that controls cell division?

Just as with the initiation of chromosome replication, candidates for the signal should be evaluated according to the criterion of whether it is easy to see how the signal might have evolved. In virtually all cells, division is an affair of membrane dynamics and often seems to involve tubulin or tubulin-like proteins such as the bacterial version, FtsZ. Although FtsZ is the key protein in cell division, attempts to find a protein acting upstream of FtsZ have failed.

Proteolipid domains? The bacterial membrane has both a structure and a structure that changes during the cell cycle. These changes are related to the presence (Fishov and Woldringh, 1999) and the activity of the chromosome (Binenbaum *et al*., 1999). A physico-chemical basis for the polymerisation of FtsZ at the membrane in cell division might therefore reside in proteolipid domains created by the coupled transcription-translation-insertion or *transertion* of proteins into and through membrane (Binenbaum *et al*., 1999; Fishov and Woldringh, 1999; Mileykovskaya and Dowhan, 2000; Norris and Fishov, 2001; Woldringh, 2002). For example, the phospholipids not bound to nascent proteins being inserted into the membrane around the chromosomes might congregate in the centre of the cell and create a membrane domain to which FtsZ would bind preferentially (Figure 1).

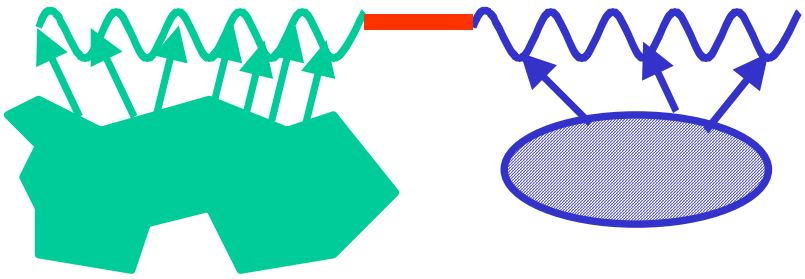


Figure 1: Transertion (arrows) from chromosomes with different activities and structures creates membrane domains around them (wavy lines) creates a division domain of specific phospholipid composition and structure between them (thick line).

Metabolic sensing? As argued above for robustness sensing, it might be advantageous for a cell to sense its metabolic state. Consider, for example, a cell that has its resources shared out between two sets of functions related to fast growth and to survival; such a cell could use division to concentrate its resources in a daughter equipped for fast growth. There would be no point doing this if its metabolism were relatively untaxed. One way metabolic sensing could occur would be if there were an enzoskeleton involving FtsZ such that FtsZ polymers in the cytoplasm associated with enzymes of the general metabolism would, at a critical level of metabolic activity, break down to release FtsZ to initiate cell division (Norris and Fishov, 2001). This would constitute a *structural sensor*. It is therefore significant that there is a fine FtsZ skeleton in moss chloroplasts (Kiessling *et al*., 2000) and an association between the tubulin cytoskeleton and glycolytic enzymes in eukaryotes (Walsh *et al*., 1989).

Are we asking the questions at the right level of intracellular organisation?

A hyperstructure is a collection of diverse molecules (genes, mRNAs, proteins, ions, lipids) that is associated with a function (figure 2).

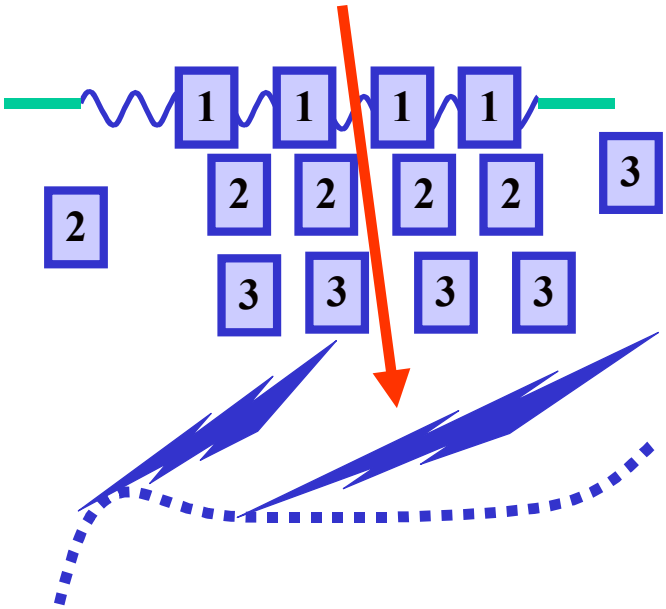


Figure 2: Formation of a hyperstructure involving genes, enzymes and the membrane that is dependent on the presence of substrate. Enzymes (rectangles 1, 2 and 3) in the pathway assemble in the presence of substrate to transport and metabolise it (red arrow). Enzyme 1, a membrane protein, structures the membrane (wavy line). The genes encoding these enzymes are attached to the hyperstructure by coupled transcription-translation (jagged shapes) (see plate 14).

Certain hyperstructures may therefore structure cytoplasm, chromosome and membrane. In particular, a non-equilibrium hyperstructure is assembled into a large, spatially distinct structure to perform a function and is disassembled, wholly or partially, when no longer required (Amar *et al*., 2002). We have proposed previously that in the case of the phosphotransferase system in *E. coli* a non-equilibrium hyperstructure may form due to an increase in the affinity of its constituent enzymes for one another in the presence of the sugar. Briefly, enzymes E1 can only diffuse in the plane of the membrane whilst the other enzymes, E2 to E7 diffuse in the cytoplasm. The binding of the sugar substrate to the E1 enzymes leads to an increase in their affinity for one another and their assembly into an E1 domain. On binding its substrate, each enzyme in the pathway acquires an increased affinity for the following enzyme. This results in the assembly of the hyperstructure which may be helped by the transcription of the genes encoding E1 to E7 and the simultaneous translation of the resultant mRNA.

In the hyperstructure approach, cell cycle progress is considered to be a state cycle of hyperstructures and the generation of phenotypically different daughter cells results in part from interactions between hyperstructures (figure 3). Such interactions can be invoked to explain how another cell cycle event, chromosome segregation, allows hyperstructures to create phenotypic diversity. For example, a *strand-specific* model can be based on each of the future daughter chromosomes being associated with a different set of hyperstructures in an asymmetric cell (Rocha *et al*., 2003).

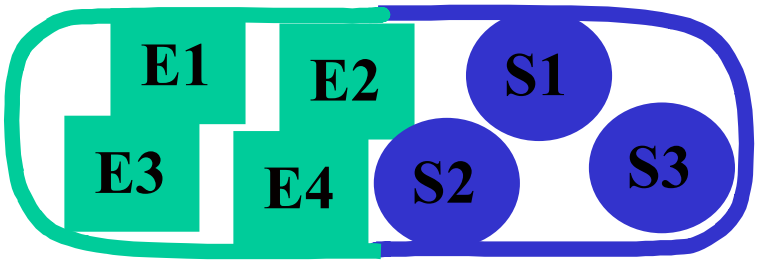


Figure 3: Segregation of hyperstructures. Green rectangles lettered E correspond to hyperstructures that have common preferences for ions, lipids, binding proteins, water structures etc. and that ensure the "exponential phase" set of phenotypes whilst blue circles lettered S correspond to hyperstructures that ensure the "stationary phase" phenotypes and that have a different set of preferences (see plate 15).

The essence of this segregation mechanism is that the genes on the same strand in the parental cell that are expressed together in a hyperstructure continue to be expressed together and segregate together in the daughter cell (figure 4).

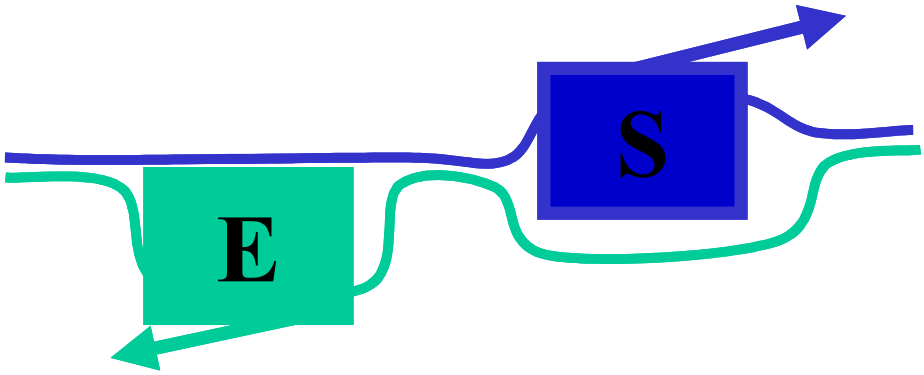


Figure 4: Strand-specific segregation mechanism. The hyperstructures S for the "stationary phase" set of functions are encoded by genes located on the blue strand of DNA whilst those for the "exponential set" E are on the green strand (see plate 16).

The model requires an asymmetric distribution of *classes* of genes and of binding sites and other structures on the strands of the parental chromosome. There is some evidence for this distribution insofar as highly expressed genes encoding functions needed for rapid growth are preferentially located on one strand whilst those needed for resistance to stresses are located on the other. This model can be usefully married to one in which different cholesteric pitches of the condensed daughter chromosomes are proposed to facilitate separation (Bouligand and Norris, 2001). In this marriage, the daughter chromosome with the "stationary phase" pattern of expression tends to a condensed, liquid crystal structure whilst the other daughter chromosome has "exponential phase" pattern of expression leading to a dynamic structure that is immiscible with the condensed structure of the other. There is actually evidence for the simultaneous presence of daughter chromosomes with different structures in the radiation-resistant bacterium, *Deinococcus radiodurans* (Minsky *et al*., 2002).

Discussion

For far too long in studies of the bacterial cell cycle, absence of evidence has been construed as evidence of absence. The absence of techniques (of the absence of *use* of techniques) that might have revealed heterogeneity at the level of structures and functions has been construed as evidence for the absence of heterogeneity. This has led to hypotheses tinged with molecular vitalism and dominated by a vision of the bacterium as a cell with an unstructured cytoplasm surrounded by an unstructured membrane in which intelligent proteins diffuse freely. In fact this vision has fitted in well with a number of other reductionist visions that include the RNA world in the case of the origin of life and the dance of sigma factors and other magic molecules in the case of differentiation. The last couple of decades have seen a gradual retreat from the idea of homogeneity under the pressure of increasing evidence for structured cytoplasm and membranes. This retreat has, however, yet to be translated into a readiness to generate or consider seriously hypotheses for cell cycle regulation that are based on dynamic structures (as opposed to those based on the simple diffusion of molecules). Here, we have advocated approaches to developing new hypotheses that are based on physics and physical chemistry. We have explored candidate hypotheses in the highly speculative context where we suggest that one partial answer to the question "what is a cell?" is "two autocatalytic networks with different connectivities that, in the form of equilibrium and non-equilibrium hyperstructures, explores phenotype space so as to both survive in Hell and multiply in Heaven". We argue that hypotheses for cell cycle regulation should be evaluated by the insights they provide into other fundamental problems. On this note, we suggest that hyperstructures are not too dissimilar from the composomes (Segre *et al*., 2000) and other structures proposed to be the selectable units in the origin of life (Norris and Raine, 1998).

Acknowledgements

We thank Christelle Koundibia for organisational help and the Ile de France, the Region de Haute Normandie, and Genopole Recherche for funding.

References

Amar, P., P. Ballet, G. Barlovatz-Meimon, A. Benecke, G. Bernot, Y. Bouligand, P. Bourguine, F. Delaplace, J.-M. Delosme, M. Demarty, I. Fishov, J. Fourmentin-Guilbert, J. Fralick, J.-L. Giavitto, B. Gleyse, C. Godin, R. Incitti, F. Képès, C. Lange, L. Le Sceller, C. Loutellier, O. Michel, F. Molina, C. Monnier, R. Natowicz, V. Norris, N. Orange, H. Pollard, D. Raine, C. Ripoll, J. Rouviere-Yaniv, M. Saier jnr., P. Soler, P. Tambourin, M. Thellier, P. Tracqui, D. Ussery, J.-P. Vannier, J.-C. Vincent, P. Wiggins and A. Zemirline (2002). Hyperstructures, genome analysis and I-cell. Acta Biotheoretica 50: 357-373.

Binenbaum, Z., A.H. Parola, A. Zaritsky and I. Fishov (1999). Transcription- and translation-dependent changes in membrane dynamics in bacteria: testing the transertion model for domain formation. Molecular Microbiology 32: 1173-1182.

Booth, I.R. (2002). Stress and the single cell: intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. International Journal of Food Microbiology 78: 19-30.

Bouligand, Y. and V. Norris (2001). Chromosome separation and segregation in dinoflagellates and bacteria may depend on liquid crystalline states. Biochimie 83: 187-192.

Demarty, M., B. Gleyse, D. Raine, C. Ripoll and V. Norris (2003). Modelling autocatalytic networks with artificial microbiology. Comptes Rendus de l'Academie des Sciences : in press.

Eliasson, A. and K. Nordstrom (1997). Replication of minichromosomes in a host in which chromosome replication is random. Molecular Microbiology 23: 1215-1220.

Elowitz, M., B. , A.J. Levine, E.D. Siggia and P.S. Swain (2002). Stochastic gene expression in a single cell. Science 297: 1183-1186.

Fishov, I. and C. Woldringh (1999). Visualization of membrane domains in *Escherichia coli*. Molecular Microbiology 32: 1166-1172.

Guelzim, N., S. Bottani, P. Bourgine and F. Kepes (2002). Topological and causal structure of the yeast transcriptional regulatory network. Nature Genetics 31: 60-63.

Kaprelyants, A.S., G.V. Mukamolova, H.M. Davey and D.B. Kell (1996). Quantitative analysis of the physiological heterogeneity within starved cultures of *Micrococcus luteus* by flow cytometry and cell sorting. Applied Environmental Microbiology 62: 1311 – 1316.

Kauffman, S. (1993). The origins of order., Oxford University Press, Oxford.
Kiessling, J., S. Kruse, S.A. Rensing, K. Harter, E.L. Decker and R. Reski (2000). Visualization of a cytoskeleton-like FtsZ network in chloroplasts. Journal of Cell Biology 151.

Maniotis, A.J., C.S. Chen and D.E. Ingber (1997). Demonstration of mechanical connections between integins, cytoskeletal filaments, and nucleoplasm that stabilize nuclear structure. Proceedings of the National Academy of Science U.S.A. 94: 849-854.

Meinhardt, H. and P.A.J. de Boer (2001). Pattern formation in *Escherichia coli*: a model for the pole-to-pole oscillations of Min proteins and the localization of the division site. Proceedings of the National Academy of Science U.S.A. 98: 14202-14207.

Mileykovskaya, E. and W. Dowhan (2000). Visualization of phospholipid domains in *Escherichia coli* by using the cardiolipin-specific fluorescent dye 10-N-nonyl acridine orange. Journal of Bacteriology 182: 1172-1175.

Minsky, A., E. Shimoni and D. Frenkiel-Krispin (2002). Stress, order and survival. Nat Rev Mol Cell Biol 3: 50-60.

Nagel, T. (1974). What is it like to be a bat? The Philosophical Review 83: 435-450.

Norris, V. (1998a). Bacteria as tools for studies of consciousness. In Toward a Science of Consciousness II: The Second Tucson Discussions and Debates. (Hameroff, S., Kaszniak, A. and Scott, A., eds.), pp. 397-405. Cambridge, USA. MIT Press,

Norris, V. (1998b). The cell as a metaphor. In Enseignements Thematiques (Tracqui, P., Demongeot, J. and Baconnier, P., eds.), Vol. 3, pp. 299-314. 3 vols. Saint Flour. Societe Francaise de Biologie Theorique,

Norris, V. (1998c). Modelling *E. coli*: the concept of competitive coherence. Comptes Rendus de l'Academie des Sciences 321: 777-787.

Norris, V., M. Demarty, D. Raine, A. Cabin-Flaman and L. Le Sceller (2002). Hypothesis: hyperstructures regulate initiation in *Escherichia coli* and other bacteria. Biochimie 84: 341-347.

Norris, V. and I. Fishov (2001). Division in bacteria is determined by hyperstructure dynamics and membrane domains. Journal of Biological Physics and Chemistry 1: 29-37.

Norris, V. and M.S. Madsen (1995). Autocatalytic gene expression occurs *via* transertion and membrane domain formation and underlies differentiation in bacteria: a model. Journal of Molecular Biology 253: 739-748.

Norris, V., T. Onoda, H. Pollaert and G. Grehan (1999). The mechanical origins of life. BioSystems 49: 71-78.

Norris, V. and D.J. Raine (1998). A fission-fusion origin for life. Origins of Life and Evolution of the Biosphere 28: 523-537.

Raine, D.J., Y. Grondin, M. Thellier and V. Norris (2003). Networks as constrained thermodynamic systems. Comptes Rendus de l'Academie des Sciences 326: 65-74.
Raine, D.J. and V. Norris (2001). Network Structure of Metabolic Pathways. Journal of Biological Physics and Chemistry 1: 89-94.

Rocha, E., J. Fralick, G. Vediyappan, A. Danchin and V. Norris (2003). A strand-specific model for chromosome segregation in bacteria. Molecular Microbiology : in press.

Segre, D., D. Ben-Eli and D. Lancet (2000). Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. Proceedings of the National Academy of Science U.S.A. 97: 4112-4117.

Tolker-Nielsen, T., K. Holmstrom, L. Boe and S. Molin (1998). Non-genetic population heterogeneity studied by *in situ* polymerase chain reaction. Molecular Microbiology 27: 1099-1105.

Walsh, J.L., T.J. Keith and H.R. Knull (1989). Glycolytic enzyme interactions with tubulin and microtubules. Biochimica Biophysica Acta 999: 64-70.

Woldringh, C.L. (2002). The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. Molecular Microbiology 45: 17-29.

Part 3

CONFERENCE SCHEDULE

Conference schedule

	Morning	Afternoon
Monday 12	François Képès Pascal Ballet et Abdallah Zermiline	Bernard Vandenbunder Eduardo Rocha Vincent Schachter Bertrand Séraphin
Tuesday 13	Arndt Benecke Patrick Amar Georgia Barlovatz-Meimon	Danièle Hernandez-Verdun Bertrand Séraphin Chaouqi Misbah Daniel Isabey
Wednesday 14	Daniel Claude Gilles Bernot Janine Guespin-Michel	Jean-Pierre Vannier Michel Thellier Daniel Claude
Thursday 15	Hans Meinhardt	Jan Traas Alexander Bockmayr
Friday 16	Marie Dutreix Christine Froidevaux Éric Fourmentin	Conrad Woldringh Marie-Claire Bellissent-Funel Pascale Mentré

Detailed Programme

Monday 12

Cellular and molecular biology / Simulation of discrete dynamic systems

9:00	Registration	
10:00	François KÉPÈS	From cells to bioinformatics
11:00	Coffee break	
11:30	Pascal BALLET and Abdallah ZERMILINE	Multiagents systems
12:30	Lunch	

Interpretation of the data

14:00	Bernard VANDENBUNDER	From transcriptome to transcriptional regulatory networks
15:00	Eduardo ROCHA	Order and disorder in bacterial genomes
16:00	Coffee break	
16:30	Vincent SCHACHTER	The double hybride
17:30	Bertrand SÉRAPHIN	Organisation of the yeast proteome
18:30	Thematic discussions	

Tuesday 13

Membranes and intra cellular hyperstructures

9:00	Arndt BENECKE	Organisation of the nucleous
10:00	Patrick AMAR	Cellular automata and proteins assemblies dynamics
11:00	Coffee break	
11:30	Georgia BARLOVATZ-MEIMON	Cellular migration and extra cellular remodelling
12:30	Lunch	

Functional nuclear hyperstructures

14:00	Danièle HERNANDEZ-VERDUN	The nucleolus
15:00	Bertrand SÉRAPHIN	Integration of the splicing process of ARNs in the cell dans the nucleous
16:00	Coffee break	
16:30	Chaouqi MISBAH	Dynamics of membranes
17:30	Daniel ISABEY	Cellular tensegrity
18:30	Thematic discussion lead by Vic NORRIS and Patrick AMAR	

Wednesday 14

Observability and theoretical models validation in biology

9:00	Daniel CLAUDE	About commandability, observability and identifiability of systems
10:00	Gilles BERNOT	Introduction to temporal logic and model-checking
11:00	Coffee break	
11:30	Janine GUESPIN-MICHEL	Epigenesis and validation
12:30	Lunch	
14:00	**Tour on the *Côte Normande***	

Biological examples

17:00	Jean-Pierre VANNIER	Haematopoiesis
18:00	Michel THELLIER	Storage and recall of morphogenetic signals in plants
19:00	Daniel CLAUDE	A PK/PD model of aminoglycoside nephrotoxicity
21:00	Thematic discussion lead by Gilles BERNOT and Janine GUESPIN-MICHEL	

Thursday 15

Spacial organisation in biology

9:00	Hans MEINHARDT	Models of biological pattern formation (part I)
11:00	Coffee break	
11:30	Hans MEINHARDT	Models of biological pattern formation (part II)
12:30	Lunch	
14:00	Jan TRAAS	The growth of the meristem and various models
15:00	Alexander BOCKMAYR	Modelling biological systems in constraint programming
16:00	Coffee break	
16:30	Students talks	
18:30	Thematic discussion lead by Jean-Louis GIAVITTO	
19:30	Conference dinner	

Friday 16

From facts to consensus in biology: physico-chemists points of vue

9:00	Marie DUTREIX	Resistance to radiations related to genes expression
10:00	Christine FROIDEVAUX	Using of an attributes selection method for the analysis of the transcriptome of yeast cells exposed to low levels of radiations
11:00	Coffee break	
11:30	Éric FOURMENTIN-GUILBERT	Simulation of the bacterial *in vivo*
12:30	Lunch	
14:00	Conrad WOLDRINGH	Physico-chemical aspects of the bacterial genome

The water

15:00	Marie-Claire BELLISSENT-FUNEL	Water at interfaces
16:00	Coffee break	
16:30	Pascale MENTRÉ	Interfacial water: a modulator of the biological activity
17:30	Thematic discussion lead by Pascale MENTRÉ	

Part 4

LECTURES / ABSTRACTS

From cells to bioinformatics

François Képès

ATelier de Génomique Cognitive CNRS UMR8071 / Genopole® Evry, France.
Francois.Kepes@genopole.cnrs.fr

Abstract

The goal of this introductory lecture is to demonstrate through two examples how cell biological knowledge allows asking fruitful questions, to which bioinformatics can efficiently answer.

Du transcriptome aux réseaux de régulation transcriptionnelle

Bernard Vandenbunder

Institut de Recherches Interdisciplinaire / Institut de Biologie de Lille, 1 rue du Professeur Calmette, BP 447 59021 Lille cedex

Résumé

Les progrès dans l'étude des réseaux régulation transcriptionnelle sont les fruits d'un dialogue entre biologistes, mathématiciens et informaticiens, entre les expérimentateurs qui mettent en oeuvre des approches locales et ceux qui utilisent des approches globales.

La première partie de ce cours concernera les résultats obtenus à ce jour sur la structure et la dynamique de ces réseaux chez E. Coli et S. Cerevisiae. C'est l'analyse informatique des données de puces à ADN et des séquences des promoteurs des gènes présentant des patrons d'expression similaires qui a d'abord servi de base pour déterminer l'architecture de ces réseaux. Cette analyse a révélé que les gènes ainsi regroupés sont souvent associés à une même fonction et que leurs séquences promotrices présentent des éléments capables de fixer le même facteur de transcription. Les expériences de précipitation de la chromatine qui permettent de localiser la présence d'un facteur de transcription sur des séquences promotrices et la manipulation des facteurs de transcription (inactivation, expression conditionnelle) ont apporté des données fonctionnelles qui ont permis de préciser la structure de ces réseaux de régulation transcriptionnelle. Dans ces réseaux il est possible de repérer des motifs ou des modules élémentaires et de simuler le comportement dynamique de ces motifs.

La seconde partie sera consacrée à l'étude de ces réseaux de régulation transcriptionnelle chez les organismes multicellulaires, la souris et l'homme. On évoquera quelques caractéristiques importantes de ces réseaux qui ont été mises en évidence par la biologie moléculaire classique et par la description de l'expression des gènes dans des cellules vivantes : le rôle de la structure de la chromatine dans la régulation de l'expression génique, la redondance des facteurs de transcription et le caractère stochastique de l'expression des gènes. On esquissera alors quelques modules de régulation qui apparaissent dans ces réseaux.

Order and disorder in Bacterial genomes

Eduardo P C Rocha

Atelier de BioInformatique, Université Pierre et Marie Curie et Unite Génétique des Génomes Bactériens, Institut Pasteur

Abstract

The availability of 100 bacterial genomes allows the comparative analysis of genome structure and composition in relation to bacterial lifestyle and evolution. Replication plays a major role among the elements structuralizing Bacterial genomes. The replication asymmetries induce biases at the levels of nucleotide composition and gene distribution that strongly fashion the chromosome. As a consequence the latter seems more structured than previously thought. However, Bacterial genomes also contain a very significant number of repeats that may shuffle the chromosome. Such repeats may be present for many purposes. Hence, a trade-off is established between the advantage of having the potential for sequence variation and the disadvantages it produces, by stimulating rearrangements that disrupt the replication chromosomal structure. For example, it has been observed that repeats tend to be placed in the chromosome in such a way that the rearrangements they may induce provoke smaller disruptions to the chromosomal replication structure than expected by chance. Further, there is an important negative correlation between the number of repeats and the replication strand biases or the synteny conservation. In short, this is a talk on biases and repeats, but especially on order and disorder in bacterial genomes.

Migration et remodelage extracellulaire

Georgia Barlovatz-Meimon

Équipe Dynamique du microenvironnement cellulaire Unité 492 INSERM, Créteil

Résumé

La migration est un processus complexe, multi-étapes, dont chacune des étapes est indispensable mais aucune ne peut constituer le point de départ. On peut néanmoins le décrire simplement comme la séquence répétée d'une phase d'adhésion (établissement d'un pont moléculaire entre la cellule et la matrice extra cellulaire), une phase d'ancrage (caractérisée par une réorganisation du cytosquelette), une phase de contraction (impliquant l'intervention de moteurs moléculaires et l'activation de voies de transduction de signaux) et enfin une phase de dé-adhésion (permettant le détachement de l'arrière de la cellule) et par voie de conséquence, une translocation résultante pour l'ensemble de la cellule. L'ensemble de ces phases est le plus souvent accompagné par une production accrue de protéinases.

Au cours de la formation d'un point d'adhésion cellulaire, un "continuum moléculaire" va se constituer et lier biochimiquement et mécaniquement la cellule à son microenvironnement matriciel (ou cellulaire dans le cas d'interactions cellules cellules) Ce continuum va induire l'activation de telle ou telle voie de signalisation et la cellule adaptera alors sa réponse. De cet échange, multiplié par autant de points d'adhésion, va résulter un ensemble de caractéristiques qui va promouvoir ou non, la migration proprement dite. Dans un éditorial de PNAS vol. 100, 2003, D. Ingber, pionnier de l'approche multidisciplinaire, dit que les cellules "agissent localement mais pensent globalement". Dans cette perspective nous utiliserons l'exemple d'un système protéinasique celui de l'activation du plasminogène, capable d'influencer chacune des étapes décrites, et qui permet d'avancer des hypothèses sur les mécanismes généraux de la migration cellulaire.

Sur la commandabilité, l'observabilité et l'identifiabilité des systèmes

Daniel Claude

Université Paris-Sud (Laboratoire des Signaux et Systèmes, CNRS-SUPELEC)
INRIA-Rocquencourt - Hôpital Paul Brousse, INSERM, Villejuif, France.

Résumé

Ce cours succinct a pour objet de présenter les notions clés de commandabilité, d'observabilité et d'identifiabilité des systèmes commandés.

Par souci de simplicité, la commandabilité et l'observabilité seront présentées dans le cadre de la représentation d'état des systèmes linéaires en temps continu et à coefficients constants.

La commandabilité se traduit par le fait que l'on peut se déplacer à son gré dans l'espace d'état ou, autrement dit, qu'il existe toujours une commande qui définit une trajectoire permettant de passer d'un état à un autre. L'observabilité se caractérise sur les observables du système qui en définissent les sorties. Elle est liée à la faculté de pouvoir distinguer des états initiaux différents en suivant l'évolution temporelle des observables considérées. Ces deux notions d'essence topologique ont en fait une correspondance algébrique qui permet de définir des tests de vérification à l'aide de simples calculs de rang de matrices.

De plus, il est aisé d'associer à ces deux notions des sous-espaces particuliers de l'espace d'état et on peut alors montrer comment ces sous-espaces varient lors de l'utilisation de feedbacks ou bien en cas de panne d'actionneurs relatifs à la commande du système ou de capteurs associés aux observables. Ceci permet de mieux appréhender la signification du comportement entrée-sortie du système.

D'autre part, on peut montrer que les feedbacks sur l'état permettent d'assigner d'une manière arbitraire les modes propres de tout système commandable et en particulier de le stabiliser. De plus, si le système a des sorties qui le rendent observable, il est aussi possible de reconstruire à l'aide d'un observateur une approximation de tout l'état et de réaliser en particulier tout feedback sur l'état sans pour autant avoir besoin de mesurer chacune de ses composantes.

La notion d'identifiabilité fait, elle, référence aux paramètres du système et à leur détermination à l'aide de la connaissance du comportement entrée-sortie du système. On dira qu'un système est identifiable s'il y a unicité de la valeur des paramètres, localement identifiable s'il n'y a qu'un nombre fini de valeurs possibles, et non identifiable si le nombre de solutions conduisant au même comportement entrée-sortie est infini. Le fait que les paramètres d'un système en temps continu soient constants peut s'exprimer en écrivant que leur dérivée par rapport au temps est nulle et si l'on rajoute ces équations au système d'équations différentielles définissant le système, qu'il soit linéaire ou non linéaire, on constate alors que la donnée des paramètres relève des conditions initiales du système étendu. On voit ainsi que la détermination des paramètres est associée à l'observabilité du système étendu.

Il est aujourd'hui possible de proposer un algorithme probabiliste qui distingue l'ensemble des variables observables d'un système et qui indique de plus le nombre de variables non observables qui doivent être supposées connues pour obtenir un système observable. La complexité arithmétique de cet algorithme est polynomiale et il est possible de minorer la probabilité d'obtenir une réponse correcte. De plus, dans certains cas, il est aussi possible de déterminer les groupes opérant sur certains sous-ensembles de paramètres dans le cas où le système est non identifiable, de proposer alors un nouveau paramétrage du système et de définir d'autres observables complémentaires.

Références

Tout livre d'Automatique sur les variables d'état. Par exemple :

D. Claude, Systèmes linéaires en temps continu ; représentation d'état, Paris-Onze Éditions - Éditions de Boeck, Bruxelles, 248 pages, 1995.

A. Sedoglavic, A probalistic algorithm to test local algebraic observability in polynomial time. Proceedings of the 2001 International Symposium on Symbolic and Algebraic Computation (London, Canada, July 22-25 2001), B. Mourrain, Ed. ACM press, 2001, pp. 309-316.

F. Ollivier, Identifiabilité des systèmes ; Applications à la biologie. Journée annuelle de la Société Mathématique de France, D. Claude, Ed., 2002, pp. 27-46.

Le site : http : //www.medicis.polytechnique.fr/sedoglav/productionLogiciel.html

Logique temporelle et vérification de modèles

Gilles Bernot[1,2]

[1]La.M.I. Université d'Évry Val d'Essonne & CNRS UMR 8042, Tour Évry 2, 523 Place des terrasses de l'agora, F-91000 Évry
[2]Genopole Recherche, 2 rue Gaston Cremieux, F-91057 Evry Cedex

Résumé

On montre ici comment une logique formelle dite temporelle peut être techniquement employée en conjonction avec les réseaux de régulation biologiques définis par René Thomas pour valider les liens entre modélisation et expérimentation en biologie. Après un rappel de notre méthode de travail pluridisciplinaire biologie/informatique, l'objet de ce cours est de décrire rapidement la modélisation discrète des réseaux de régulation biologiques, la logique temporelle CTL et le «Model-Checking».

Epigenèse et validation.

Janine Guespin-Michel

Laboratoire des Processus Intégratifs Cellulaires, UPRESA CNRS 6037, faculté des sciences et techniques, université de Rouen, 76821 Mont-Saint-Aignan Cedex, France

Résumé

Le terme épigenèse est très ancien, puisqu'il a été crée par Harvey (1651), pour décrire la formation progressive des différentes parties d'un embryon. Mais une nouvelle signification a émergé lorsque la génétique a développé les notions de génotype et de phénotype. Des modifications du phénotype sont dites épigénétiques lorsqu'elles sont héritables en dépit du fait qu'elles ne sont pas dues à des modifications génétiques (mutations ou remaniements du génome). Elles peuvent apparaître à la suite d'un signal de l'environnement, mais elles ne disparaissent pas avec ce signal. Autrement dit, avec le même génome, et dans des conditions identiques, des cellules ou des organismes peuvent avoir un phénotype différent, si leur passé a été différent. Ceci est la manière biologique de décrire ce que les physiciens appellent la multistationnarité, et qui résulte du fonctionnement de certains systèmes dynamiques présentant des interactions non linéaires.
L'épigenèse commence a être acceptée et reconnue chez les eucaryotes (Danchin 1999), mais la plus grande part des efforts de recherche porte à l'heure actuelle sur des mécanismes biochimiques, comme la méthylation de l'ADN. L'archétype des modifications épigénétiques étudiées selon ce type d'approche est l'extinction d'un des deux chromosomes X dans les cellules de mammifères femelles. Utiliser la notion de multistationnarité pour étudier l'épigenèse permet de découvrir des aspects différents et parfois inattendus de la réalité biologique. Dans cette optique, les phénomènes de différentiation sont épigenétiques, comme Delbrück l'avait génialement prévu en 1949 (Demongeot 1998, Thomas and Kaufman 2001a), de même que la mémoire (Cassadessus et d'Ari 2002, Demongeot *et al* 2000). Cependant, contrairement aux approches biochimiques, ou de biologie moléculaire, cette approche nécessite le recours à la modélisation, et pose donc d'entrée de jeu la question cruciale des rapports entre modélisation et expérimentation.
Le présent exposé va donc d'abord, sur un exemple très simple en microbiologie montrer ce que signifie un processus épigénétique dû à une multistationnarité, et les relations entre épigenèse et circuit de rétroaction positive.
Puis, à partir d'un exemple un peu plus compliqué, toujours en microbiologie, nous illustrerons la nécessité d'une modélisation, l'intérêt d'une méthode discrète pour étudier un tel modèle, et l'utilité d'une étape de validation.

Bibliographie

Casadesus J, D'Ari R. (2002) Memory in bacteria and phage. Bioessays ; 24512-8

Danchin, A. (1999) L'habit ne fait pas le moine. La Recherche 326: 28.

Delbrück M. (1949) Unités biologiques douées de continuité génétique. Colloques internationaux du CNRS

Demongeot, J. (1998) Multistationarity and cell differenciation. J.Biol.sys. 6: 1-2

Demongeot; J., Kaufman M. and Thomas, R. (2000) Positive feedback circuits and memory. Comptes Rendus de l'Académie des Sciences, Paris 323: 69-79.

Guespin-Michel J.F., and Kauffman M. (2001) Positive feedback circuits and adaptive regulations in bacteria. Acta biotheoretica 49: 207 - 218

Guespin-Michel J.F. (2000) Les boucles de rétroaction positives, une nouvelle approche intégrée de la physiologie microbienne. Bulletin de la société de microbiologie Septembre

Novick, A. and M. Wiener (1957) Enzyme induction is an all-or-none phenomenon. Proceedings of the National Academy of Sciences USA 43: 553-556.

Thomas R. and Thieffry D. (1995) Les boucles de rétroaction rouages des réseaux de régulation biologiques. Médecine/Science 11: 189-197

Thomas, R. (1998) Laws for the dynamics of regulatory networks. International Journal of Developmental Biology. 42: 479-485.

Thomas, R. (1980) On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. Springer Series in Synergies 9: 180-193

Thomas, R. and d'Ari R. (1990) Biological feedback. CRC press, Boston.

Thomas, R. and Kaufman M. (2001a) Multistationarity, the basis of cell differentiation and memory. I. Structural conditions of multistationarity and other nontrivial behavior. Chaos 11: 170-179.

Thomas, R. and Kaufman M. (2001b) Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits. Chaos 11: 180-195.

Organisation du protéome de la levure

Bertrand Séraphin

Centre de Génétique Moléculaire, CNRS, Avenue de la Terrasse, 91198 Gif sur Yvette
seraphin@cgm.cnrs-gif.fr

Résumé

La plupart des processus cellulaires sont effectués par des complexes protéiques macro-moléculaires. Nous avons développé une méthode générique pour purifier de tels assemblages que nous avons appelée méthode TAP. Combinée à la spectrométrie de masse, cette stratégie permet l'identification des protéines interagissant avec une protéine cible donnant ainsi des informations sur la manière dont l'ensemble des protéines d'un organisme (protéome) est organisé en unités fonctionnelles. L'utilisation à grande échelle de cette stratégie permet l'analyse systématique des réseaux d'interaction protéique. Une telle étude effectuée sur la levure *Saccharomyces cerevisiae* a permis la définition de nombreux complexes. L'analyse des résultats de cette analyse et des réseaux d'interaction entre ces complexes seront décrits. Les problèmes liés à la validation des données et à leurs analyses en conjonction avec les données des résultats de puces à ADN seront discutés. Nos résultats indiquent que dans le futur d'autres réseaux complexes d'interaction, tels que ceux entre les protéines et les ARNs messagers, pourront aussi être analysés.

The nucleolus: functions and assembly

Danièle Hernandez-Verdun

Institut Jacques Monod, CNRS, Univerité Paris VI et VII, Paris

Abstract

The nucleolus is a large nuclear domain and the site of ribosome biogenesis. It is also at the parting of the ways of several cellular processes, including cell cycle progression, gene silencing, and ribonucleoprotein complex formation. Consequently, a functional nucleolus is crucial for cell survival. Recent investigations of nucleolar assembly during the cell cycle and during embryogenesis have provided an integrated view of the dynamics of this process. Moreover, they have generated new ideas about cell cycle control of nucleolar assembly, the dynamics of the delivery of the RNA processing machinery, the formation of prenucleolar bodies, the role of precursor ribosomal RNAs in stabilizing the nucleolar machinery and the fact that nucleolar assembly is completed by cooperative interactions between chromosome territories. This has opened a new area of research into the dynamics of nuclear organization and the integration of nuclear functions.

Intégration du processus d'épissage des ARNs dans la cellule et le noyau

Bertrand Séraphin

Centre de Génétique Moléculaire, CNRS, Avenue de la Terrasse, 91198 Gif sur Yvette
seraphin@cgm.cnrs-gif.fr

Résumé

L'expression des gènes eucaryotes implique l'épissage d'introns : des séquences transcrites qui ne se retrouvent pas dans l'ARN messager final. Pour plusieurs raisons, le processus d'épissage est complexe :

- La reconnaissance des introns peut varier d'une cellule à l'autre ou en fonction des conditions (épissage alternatif) ;
- Les séquences spécifiques des introns sont courtes, rendant leur reconnaissance difficile tant par la machinerie d'épissage que par les programmes d'analyse de séquences ;
- Le processus d'épissage est lié à la synthèse de l'ARN en amont (transcription) et affecte son devenir (transport vers le cytoplasme, traduction, dégradation, ...) ;
- La cinétique des divers processus intervient dans la compétition entre diverses voies d'épissage.

Il est donc clair que la compréhension de la structure du noyau est au coeur des analyses de l'épissage et de ses liens avec les autres fonctions cellulaires et que des analyses quantitatives et/ou de modélisation seront importantes dans ce domaine.

Dynamique des membranes

Chaouqi Misbah

LSP , GREPHE, Univ. J. Fourier et CNRS, Grenoble I, 38402 Saint Martin d'Hères, France

Résumé

Nous présenterons des travaux théoriques et observations expérimentales sur (i) la dynamique de migration de vésicules biologiques, (ii) la cinétique de fluctuations de membranes en présence de macromolécules. Dans la première partie nous mettons l'accent sur les lois de migration en haptotaxie et sous cisaillement hydrodynamique (force de portance et mouvement de bascule de vésicules). Deux méthodes de modélisation sont utilisées : la méthode intégrale basée sur les fonction de Green, et la méthode de frontière diffuse (champ de phase, similaire au 'level set'). Sur le plan expérimental, les méthodes d'observation sont optiques (contraste de phase). Dans la seconde partie, on mettra en évidence le spectre de fluctuations de la membrane en présence de macromolécules, et discuterons les informations que l'on peut tirer sur le couplage macromolécules/membranes. Sous certaines conditions, des instabilités de membranes sont prédites. Ces instabilités peuvent générer soit (i) des modulations 'douces' de la membrane, ou peuvent conduire à la « vésiculisation » des membranes (émission de vésicules à l'image de l'appareil de Golgii). De nouvelles approches optiques sont en cours de développement au laboratoire pour analyser ces questions sur le plan expérimental.

Références

Cantat et C. Misbah, *Lift* force and dynamical unbinding of vesicles in a shear flow, Phys. Rev. Lett. **83**, 880 (1999).

Cantat et C. Misbah, *Dynamics* and similarity laws of adhering vesicles in haptotaxis, Phys. Rev. Lett. **83**, 235 (1999).

Cantat, C. Misbah et Y. Saito, *Vesicle* propulsion in haptotaxis: A local model Euro. Phys. J. E, **3**, 403 (2000).

T. Biben et C. Misbah, *An* advected field method for deformable entities under flow, Euro. Phys. J. B, **29**, 311 (2002).

F. Divet, T. Biben, I. Cantat, A. Stephanou, B. Fourcade and C. Misbah, Fluctuation of a membrane interacting with a diffusion field, Euro. Phys. Lett. **60**, 795 (2002).

T. Biben et C. Misbah, *Tumbling* of vesicles within an advected field approach, Phys. Rev. E **67**, 031908 (2003).

M. Abkarian, C. Lartigue et A. Viallat, Motion of phospholipidic vesicles along an inclined plane: sliding and rolling, Phys. Rev. E **63**, 041906 (2001).

M. Abkarian, C. Lartigue et A. Viallat, Tanktreading and unbinding of deformable vesicles in shear flow: determination of the lift force, Phys. Rev. Lett. **88**, 068103 (2002).

Tensegrity behavior of cortical and cytosolic cytoskeletal components in twisted living adherent cells

[1]Valérie Laurent, [1]Patrick Cañadas, [1]Redouane Fodil, [2]Emmanuelle Planus, [3]Atef Asnacios, [4]Sylvie Wendling, [1]Daniel Isabey

[1]INSERM U492, Faculté de medicine, 8 rue du Général Sarrail, F-94010 Créteil Cedex
[2]Faculté des Sciences et Technologies, Université Paris XII, 61 avenue du Général de Gaulle, F-94010 Créteil Cedex
[3]Laboratoire de Bio-rhéologie et Hydrodynamique Physico-chimique Case 7056, Université Paris 7, F-75251 Paris Cedex 05
[4]Laboratoire de Mécanique et d'Acoustique, CNRS UPR 7051, 31 chemin Joseph-Aiguier 13402 Marseille Cedex 20

Abstract

The present study is an attempt to relate the multicomponent response of the cytoskeleton (CSK) - evaluated in twisted living adherent cells - to the heterogeneity of the cytoskeletal structure - evaluated both experimentally by means of 3D reconstructions, and theoretically considering the predictions given by two tensegrity models composed of (4 and 6) compressive elements and (respectively 12 and 24) tensile elements. Using magnetic twisting cytometry in which beads are attached to integrin receptors linked to the actin CSK of living adherent epithelial cells, we specifically measured the elastic CSK response at quasi equilibrium state and partitioned this response in terms of cortical and cytosolic contributions with a two-component model (double Voigt). These two CSK components exhibit a stress-hardening response and a prestress characteristics of tensegrity behaviour with however significant differences. Compared to the cytosolic component, the cortical cytoskeleton appears to be a faster responding component, being a less pre-stressed and easily deformable structure. Noteworthy, the discrepancies in elastic behaviour between the cortical and cytosolic CSK components may be understood on the basis of prestress tensegrity model predictions, given that the length and number of constitutive actin elements are taken into account.

Modélisation des phases précoces de l'hématopoïèse

Jean Pierre Vannier

Groupe de recherche sur « Micro-Environnement et Renouvellement Cellulaire Intégré » Faculté de Médecine-Pharmacie – Rouen

Résumé

Les éléments figurés du sang ont pour la plupart une durée de vie très limitée. L'hématopoïèse est l'ensemble des processus cellulaires et tissulaires qui en assurent le renouvellement. Comme les autres grandes fonctions de l'organisme, elle met en jeu des mécanismes cellulaires généraux qui permettent la croissance et la différenciation des organes ou des tissus. Cependant elle est remarquable par

- la stabilité du compartiment des cellules souches qui s'adapte à l'organisme qui les héberge,

- l'importance de la division cellulaire qui assure une production journalière intense mais contrôlée et adaptée aux besoins,

- le contrôle de la multiplication cellulaire en raison des risques majeurs d'échappement et transformation leucémique,

- le couplage dans un rapport variable mais inévitable *in vitro*, entre multiplication et différenciation cellulaire,

- l'importance des phénomènes de migration qui assurent le repeuplement de nouvelles zones d'hématopoïèse en fonction de la croissance et des éventuels besoins physiologiques.

Les études sur le tissu médullaire et *a fortiori* sur le sang ont été grandement facilitées par la possibilité d'obtenir aisément des échantillons sans compromettre l'intégrité de l'individu. Que ce soit au laboratoire ou en clinique humaine les greffes de moelle osseuse se sont rapidement développées et ont apporté une masse importante d'informations sur l'hématopoïèse.

Dès les premières études *in vitro* par culture cellulaire puis *in vivo* chez l'animal il est apparu que le développement de l'une ou l'autre des lignées hématopoïétiques dépendait de facteurs dit « de croissance ». Beaucoup d'entre eux ont été identifiés. Certains sont spécifiques d'une lignée d'autres non. D'autres encore agissent sur les phases précoces parfois seulement en présence des constituants de la matrice extracellulaire. Une interprétation un peu naïve de la régulation de l'hématopoïèse a alors vu le jour. Tout se passerait comme si à chaque besoin physiologique correspondait un facteur stimulant et un facteur inhibiteur. En fait les facteurs inhibiteurs sont peu nombreux, peu spécifiques et leur rôle est pour le moins très ambigu avec des activités biologiques très difficiles à cerner. En revanche la multiplication cellulaire dépend fortement de la présence de facteur de croissance dont l'absence conduit quasi inévitablement à la mort cellulaire par apoptose. L'adaptation aux besoins pourrait utiliser ce mécanisme dont on sait la responsabilité dans les hémopathies malignes.

Si les mécanismes conduisant à la production cellulaire dans l'une ou l'autre des « branches » de l'hématopoïèse apparaissent plus clairs, il n'en est pas de même de ceux qui concernent les phases précoces qui ont outre la fonction de produire des cellules engagées vers telle ou telle lignée, celle de maintenir un tissu contenant des cellules souches en nombre suffisant. Au vu de ce qui est observable *in vivo* et *in vitro* un certain nombre d'exigences théoriques ont été formulées très tôt. On a donc imaginé des cellules souches douées de deux aptitudes d'une part celle d'engendrer des cellules engagées vers l'une ou l'autre des voies de la différenciation, d'autre part celle de se reproduire identiques à elles-même.

Le passage d'un état stable de cellules souche quiescentes ou se reproduisant lentement en fonction des besoins ou des capacités de l'organisme, vers un état très dynamique de multiplication et de différenciation suppose la mise en œuvre de modifications de type épigénétique sans doute liées à des mécanismes de rétrocontrôle positif. La mise en évidence de sécrétions autocrines plaide pour l'existence de telles boucles.

Plusieurs modèles théoriques peuvent rendre compte de l'engagement d'une cellule vers telle ou telle lignée. Une orientation stochastique reste la plus vraisemblable. Quoiqu'il en soit la différenciation correspond à un véritable changement d'état qui une fois enclenché doit aller jusqu'à son terme. Ce phénomène est supposé lui aussi correspondre à une modification épigénétique qui suppose l'existence d'une boucle de rétroaction. L'adaptation de la production de telle ou telle lignée se ferait plus tard, indépendamment d'une lignée à l'autre, en fonction des besoins spécifiques, essentiellement par des mécanismes de stimulation positive.

La participation du microenvironnement cellulaire mais aussi de la matrice extracellulaire est certaine. En effet bien qu'il existe une recirculation des cellules souches, l'hématopoïèse est cantonnée dans des tissus bien définis tout au long de l'évolution supposant l'existence de véritables « niches ». Leur spécificité physiologique nous échappe encore en grande partie. N'ayant jamais pu créer jusqu'à ce jour un modèle expérimental de maintien de l'hématopoïèse *in vitro*, il est difficile aux hématologistes de proposer une modélisation suffisamment précise et fiable. Il faut donc travailler sur des hypothèses simplifiant vraisemblablement à l'extrême des mécanismes beaucoup plus fins. Notre hypothèse est que le maintien d'un pool de cellules souches est vraisemblablement assuré dans la moelle comme dans beaucoup d'autres organes par des mécanismes intracellulaires peu influençables par l'état physiologique ou pathologique. En revanche la taille de ce pool pourrait s'adapter au « volume » de l'hématopoïèse au niveau de l'organisme global par des signaux mettant en jeu directement ou indirectement le microenvironnement cellulaire. Dans nos modèles il faut donc faire entrer des fonctions de nature différente. L'impératif de sauvegarde des cellules souches qui est observable *in vivo* dans des conditions physiologiques et pathologiques très variées laisse imaginer un phénomène relativement robuste. Les hypothèses mettant en jeu différentes boucles de rétrocontrôle positif et/ou négatif connues n'en rendent pas compte à moins d'imaginer un ou plusieurs autres facteurs dont la découverte reste à faire.

Un modèle PK/PD de la néphrotoxicité des aminoglycosides

Daniel Claude

Université Paris-Sud (Laboratoire des Signaux et Systèmes, CNRS-SUPELEC)
INRIA-Rocquencourt - Hôpital Paul Brousse, INSERM, Villejuif, France.

Résumé

Les aminoglycosides sont des antibiotiques bactéricides dont l'activité est concentration-dépendante. La voie rénale est leur voie essentielle d'élimination. La principale limite à leur utilisation est leur accumulation dans le cortex rénal qui peut conduire à la survenue d'une néphrotoxicité. Une ototoxicité est aussi connue. Nous proposons une modélisation pharmacocinétique/pharmacodynamique (PK/PD) de la toxicité rénale des aminoglycosides en unifiant plus de quarante années de connaissances physiologiques. Ce modèle déterministe décrit successivement le comportement pharmacocinétique des aminoglycosides, leur accumulation dans le cortex rénal, leur effet sur les cellules rénales, leur conséquence sur la fonction rénale par l'intermédiaire du feedback tubulo-glomérulaire et finalement l'évolution de la concentration sérique de la créatinine considérée comme un marqueur de la toxicité. L'étude de l'identifiabilité du modèle permet de comprendre comment l'association des données de la littérature, des données cliniques et l'utilisation des techniques d'optimisation conduit à une estimation correcte des paramètres du modèle. En outre, l'étude de l'identifiabilité permet aussi de reparamétrer le modèle et de définir les observables utilisables en clinique humaine est à même de rendre le système globalement identifiable. Enfin, la simulation du modèle met en évidence l'influence prépondérante de la forme et des heures d'application du schéma d'administration dans la recherche d'une minimisation de la toxicité. Ainsi, une administration d'une même dose journalière minimise la toxicité lorsqu'elle est administrée en une seule fois ; résultat qui corrobore les études cliniques antérieures. Mais, pour un tel schéma, l'heure d'administration joue un rôle essentiel dans la vitesse de récupération de la fonction rénale après arrêt du traitement, l'heure théorique optimale se situant chez l'Homme à 13h 30, au moment du repas de milieu de journée. On se trouve ainsi en parfaite correspondance avec les résultats chronobiologiques connus chez le rat, à un décalage près de 12 heures.

Références

F. Rougier, D. Claude, M. Maurin, A. Sedoglavic, M. Ducher, S. Corvaisier, R.W. Jelliffe, P. Maire, Aminoglycoside Nephrotoxicity: Modeling, Simulation and Control,
Antimicrobial Agents and Chemotherapy, *47 (3): 1010-6, 2003.*

Organ initiation at the shoot apical meristem in higher plants: dynamics of a stable system

Jan Traas[1], Pierre Barbier de Reuille[2], Christophe Godin[2]

[1]INRA, Institut Jean Pierre Bourgin, Laboratoire de Biologie Cellulaire, Route de Saint Cyr, 78026 Versailles Cedex
[2]UMR Cirad/INRA/CNRS/Univ. Montpellier II AMAP, Botanique et Bioinformatique de l'architecture des plantes, Boulevard de la Lironde, 34398 Montpellier cedex.

Abstract

The aerial parts of the plant are generated by small groups of rapidly dividing cells called shoot apical meristems (SAMs). These are highly organized, stable structures divided into morphologically distinct domains. These domains appear to have different functions. Thus, the central zone (CZ) at the tip of the well characterized angiosperm meristem, is involved in meristem maintenance and provides a permanent source of stem cells. This group of cells is surrounded by the so-called peripheral zone (PZ), where continuously new primordia are generated in highly stereotypic patterns. Although the meristem itself is extremely stable, capable of functioning for prolonged periods, its individual cells are highly dynamic as they are constantly dividing, expanding and differentiating. To explain this apparent paradox, it has been proposed that cells within the SAM do not act autonomously, but that they continuously interact. By doing so, they create a network of interactions which provides the overall stability of the system. In this context, the plant hormone auxin is supposed to play a major role in the patterning of cell behaviour at the meristem. Auxin is transported through the plant by membrane linked proteins, the PIN-FORMED proteins. These proteins are often polarly localised within the cells, thus creating auxin fluxes within the plant. This so-called polar auxin transport (PAT) is very important for patterning at the shoot apex. Mutants in Arabidopsis, where PAT is impaired (the pin1 and pid mutants), have reduced leaf numbers and are unable to initiate flowers. As a consequence, these mutants form a naked inflorescence stem without any organ. When small quantities of auxin are applied on these mutant apices, organs grow out at the site of application. Together, the data suggest, that the PIN-FORMED proteins create local auxin maxima at the meristem which, in turn, induce organ initiation and outgrowth. To obtain further information on the role of auxin in the patterning process, we are following two approaches:

- First we are trying to understand the molecular processes associated with auxin induced patterning. In particular, we have developed an in vivo observation method, which allows us to follow auxin induced changes in cell proliferation, expansion and differentiation (gene expression patterns).
- Second, we are trying to develop a model in the form of a virtual meristem, based on our current knowledge. This second approach (explained in the poster by Pierre Barbier de Reuille et al.), should allow us to test a number of hypotheses regarding the role of auxin fluxes in patterning at the SAM.

References

Gälweiler L., Guan C., Mueller A., Wisman E., Mendgen K., Yephremov A., Palme K. 1998 Regulation of polar auxin transport by AtPIN1 in Arabidopsis vascular tissue. Science 282, 2226-2230.

Reinhardt D., Mandel T., Kuhlemeier C. 2000 Auxin regulates the initiation and radial position of plant lateral organs. Plant Cell 12, 507-518.

Steeves T.A., Sussex, I.A. 1989 Patterns in Plant Development, 2nd ed. Cambridge University Press

Traas J., Doonan J. (2001) Cellular Basis of Shoot Apical Meristem Development. Int. Rev. Cytol. 208: 161-206

Vernoux T., Kronenberger J., Grandjean O., Laufs P., Traas J., (2000) PIN-FORMED 1 regulates cell fate at the periphery of the shoot apical meristem. Development 127, 5157-5165

Modélisation de systèmes biologiques en programmation par contraintes

Alexander Bockmayr

Université Henri Poincaré, Nancy 1, LORIA (UMR 7503, CNRS/INRIA/Universités), B.P. 239, 54506 Vandoeuvre-lès-Nancy, France, Alexander.Bockmayr@loria.fr, http://www.loria.fr/~bockmayr

Résumé

La programmation par contraintes est un nouveau paradigme de programmation qui a été proposé dans les années 80 et qui est de plus en plus utilisé depuis. Une contrainte est une formule logique qui contient des variables et qui définit une relation qui doit être satisfaite par les valeurs de ces variables. Par exemple la formule $x+y \leq 1$ exprime que la somme des valeurs des variables x et y doit être inférieure ou égale à 1. En programmation par contraintes, l'utilisateur programme avec des contraintes, c.-à-d. qu'il décrit un problème avec un ensemble de contraintes qui peuvent être liées avec différents combinateurs : conjonction, disjonction, opérateurs temporels, etc. Chaque contrainte donne une information partielle sur l'état du système étudié. Les logiciels de programmation par contraintes permettent de déduire de nouvelles contraintes à partir des contraintes données et de calculer des solutions, c.-à-d. des valeurs pour les variables qui satisfont simultanément l'ensemble des contraintes.

Nous décrivons deux applications de la programmation par contraintes en biologie : détermination de la structure de macromolécules biologiques et modélisation de la dynamique de systèmes biologiques sur le plan moléculaire et cellulaire. En particulier, nous présentons des travaux actuels sur la modélisation de la régulation de l'épissage alternatif chez HIV.

Références

Bockmayr Alexander and Courtois Arnaud. Modeling biological systems in hybrid concurrent constraint programming (Poster). 2nd International Conference on Systems Biology, ICSB'01, Caltech, November 2001. http://www.icsb2001.org/Posters/096_bockmayr.pdf

Bockmayr Alexander and Courtois Arnaud. Using hybrid concurrent constraint programming to model dynamic biological systems. 18th International Conference on Logic Programming, ICLP'02, Copenhagen, July 2002. Springer, LNCS 2401, 85-99.

Bockmayr Alexander and Hooker John. Constraint Programming. In Handbook of Discrete Optimization (Eds. K. Aardal, G. Nemhauser, et R. Weismantel), Elsevier, 2003, To appear.

Bockmayr Alexander and Weispfenning Volker. Solving numerical constraints. In Handbook of Automated Reasoning (Eds. A. Robinson and A. Voronkov). Vol. 1, Chapter 12, 751 - 842, Elsevier, 2001.

Eveillard Damien, Ropers Delphine, de Jong Hidde, Branlant Christiane and Bockmayr Alexander. Multiscale modeling of alternative splicing regulation. Computational Methods in Systems Biology, CMSB'03, Rovereto, Italy, February 2003. Springer LNCS 2602, 75-87.

Lunin Vladimir Y., Urzhumtsev Alexandre and Bockmayr Alexander. Direct phasing by binary integer programming. Acta Crystallographica, A58, 283 - 291, 2002.

*Simulation de l'*in-vivo *bacterien*

Jean Fourmentin-Guilbert[1], Eric Fourmentin-Guilbert

[1]Président de la Fondation Scientifique Fourmentin Guilbert
fondation@fourmentin.guilbert.org

Résumé

Compte tenu des progrès continus de l'informatique, il est probable que d'ici une quinzaine d'années nous serons capable d'intégrer les informations nécessaires à la reconstruction in-silico d'une cellule à l'échelle moléculaire.

Le projet SIMEBAC (Simulation du Métabolisme BACtérien) vise à terme le développement d'une simulation 3D d'une bactérie (E. Coli). Dans un premier temps, nous cherchons à construire une simulation topologique d'une fonction simple comme la transcription de l'ADN. Un premier recensement des acteurs intervenant dans cette fonction montre de nombreuses lacunes dans nos connaissances et ce projet pose déjà des interrogations sur les technologies expérimentales à développer pour y répondre.

Le rôle de la Fondation est d'aider au travers de différentes actions le développement de ce projet.

Water at interfaces: from model systems to biomolecules

Marie-Claire Bellissent-Funel

Laboratoire Léon Brillouin (CEA-CNRS), CEA-Saclay, 91191 Gif-sur-Yvette Cedex
e-mail : mcbel@llb.saclay.cea.fr

Abstract

Structure and single-particle dynamics of bulk water have been extensively studied over a wide range of temperatures [1,2] and pressures [3]. However, in many relevant situations, water is not in its bulk form but instead attached to some substrates or filling some cavities. Common examples are water in porous media, water in the interior of living cells or on surfaces of proteins and lipid bilayers. We shall call water in the latter environment an « interfacial water » as opposed to a bulk water.

Water is essential for the stability and function of biological macromolecules [4]. In living systems, essential water-related phenomena occur in restricted geometries in cells, and at active sites of proteins and membranes or at their surface. In this review paper, we present the more recent up to date account of the structure and dynamics of interfacial water as compared with that of bulk water. Various techniques are used to study dynamics of confined water. Among them, quasi-elastic and inelastic neutron scattering is a powerful tool to study translational and rotational diffusion as well as vibrational density of states of confined water. Several examples involving water confined in porous media, in presence of organic solutes and at surface of biological molecules are presented. In particular, high resolution quasi-elastic neutron scattering studies of translational dynamics of water molecules on the surface of a deuterated C-phycocyanin protein [5,6] are presented and compared with that of water confined in the pores of a hydrophilic model system [5,7]. Vibrational density of states of interfacial water are presented as a function of temperature, and for two levels of hydration of the protein [5].

Furthermore, the combined effects of the hydration level and the temperature on both the change of structure of water and on the retardation of the water molecules motions are discussed. Some recent theoretical approach is also presented [6,7].

We are now able to propose a picture of room temperature interfacial water in terms of some increase of the extension of H-bond network of water as it occurs in supercooled water at a temperature some 25 K lower.

References

[1] M.-C. Bellissent-Funel, in *Hydrogen Bonded Liquids* (J. C. Dore and J. Teixeira, eds.), NATO ASI Series C, Kluwer Academic, Dordrecht, (1991).

[2] J. Teixeira, M.-C. Bellissent-Funel, S.H. Chen and A. J. Dianoux, Experimental determination of the nature of diffusive motions of water molecules at low temperatures, *Phys. Rev. A,* **31**, 1913 (1985).

[3] M.-C. Bellissent-Funel and L. Bosio, A neutron scattering study of liquid D_2O under pressure and at various temperatures, J. Chem. Phys., **102**, 3727 (1995).

[4] M.-C. Bellissent-Funel, "Hydration Processes in Biology: Experimental and Theoretical Approaches"NATO ASI : Life Sciences Series (Publisher IOS Press), Vol. 305, 1999.

[5] M.-C. Bellissent-Funel, J.-M. Zanotti and S.H. Chen, Slow dynamics of water molecules on surface of a globular protein, *Faraday Discussions,* **103**, 281 (1996).

[6] S. Dellerue and M.-C. Bellissent-Funel, Relaxational dynamics of water molecules at protein surface, Chemical Physics, **258**, 315 (2000).

[7] J.-M. Zanotti, M.-C. Bellissent-Funel and S.H. Chen, Relaxational dynamics of supercooled water in porous glass, *Phys. Rev*. E, **59**, 3084 (1999).

Part 5

POSTERS

Modélisation de la voie des MAPK par un graphe d'inférences flou pour les simulations multi-agents

Gabriel Querrec, Vincent Rodin, Jean-François Abgrall, Jacques Tisseau

Laboratoire d'ingénierie informatique, Technopôle Brest-Iroise, BP 30815, F-29608 Brest cedex , France
e-mail : gabriel.querrec@enib.fr

Résumé

La recherche médicale est grande pourvoyeuse de données. La modélisation et la simulation informatique sont alors devenues des outils indispensables pour comprendre ces phénomènes et pour aider les biologistes à orienter leurs recherches. Nous développons des modèles multi-agents de cellules pour les biologistes. Cependant, les agent-cellules doivent avoir un comportement cohérent. Nous proposons ici l'utilisation d'un graphe d'inférences flou pour la modélisation et la simulation des réseaux biochimiques régulant le comportement cellulaire. Dans cette étude, nous avons simulé deux modèles de la voie des MAPK dont un modèle impliquant une rétro-action négative et un comportement cyclique. Le comportement de cette voie varie d'une cellule à l'autre mais reste cohérent. Par contre, la simulation à grande échelle montre que les comportements globaux de ces modèles se rapprochent de ceux obtenus par équations différentielles. Par conséquent, cette méthodologie semble être appropriée pour donner à nos agent-cellules, un comportement basé sur les réseaux biochimiques de régulation intra-cellulaires.

Towards a dynamic model of the Arabidospsis' Meristem

Pierre Barbier de Reuille[1], Jan Traas[2], Christophe Godin[2]

[1]INRA, Institut Jean Pierre Bourgin, Laboratoire de Biologie Cellulaire, Route de Saint Cyr, 78026 Versailles Cedex
[2]UMR Cirad/INRA/CNRS/Univ. Montpellier II AMAP, Botanique et Bio-informatique de l'architecture des plantes, Boulevard de la Lironde, 34398 Montpellier Cedex.

Abstract

The shoot apical meristem (SAM) is a population of undifferentiated cells which initiates all aerial parts of the plant. The SAM is a highly organized structure, which is divided into distinct zones and layers. It constantly produces organs at its periphery: organ position (or phyllotaxis) is highly stereotypic and in many species, like Arabidopsis, organs are positioned in spirals.

The SAM is composed of highly dynamic cells. In this context, the cell-cell interactions are essential. Auxin, a plant hormone, is supposed to be a very important signal. We know that auxin is transported through the plant by plasmamembrane linked proteins of the so called PIN-FORMED family. Moreover, the PIN proteins are often polarly localized in the cells. This causes oriented fluxes of auxin through the plant.

The goal of this thesis project is to build a virtual meristem from the existing biological knowledge. In the first phase of this project, two major issues have been addressed:

The first issue is the representation of the meristem itself. At the tissue level, the meristem can be represented as a pack of interacting cells. To avoid to explicit the topology and geometry of each cell, we are considering a representation of the cell population based on Voronoi complexes, in which each cell is characterized by a single point.

The second issue is the formalization of the dynamic of such a system. As the number of cells increases with time and the state of the cells themselves changes, we have to describe a dynamical system with a dynamical structure (DS2). To model the evolution of such a system, we have selected a modelling language, MGS, a rewriting system developed by Jean-Louis Giavitto (CNRS, Evry) and Olivier Michel to tackle different types of dynamical structures (in particular Voronoi complexes).

We will present in this poster how we are planning to use these techniques and the first results we have concerning them.

References

Douady, S. & Couder, Y. (1996) Pytllotaxis as a self organizing process Part II: The whorled modes and their relation to the spiral ones. Journal of theoretical biology, 178, 275-294

Honda, H. (1981) Geometrical Models for Cells in Tissues. International review of cytology, 81, 191-248

Autran, D. & Traas, J. (2001) Organisation et fonctionnement des cellules souches ve'ge'tales : le me'riste`me apical d'Arabidopsis. me'decine/sciences, 17, 836-844

Giavitto, J-L. & Michel, O. (2001) MGS: a Ruled Base Programming Language for Complex Objects and Collections. Electronic Notes in Theoretical Computer Science, 59, No.4

Modèles d'horloges circadiennes

Paul François, Vincent Hakim

Laboratoire de Physique statistique, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05
francois@lps.ens.fr

Résumé

La plupart des modèles analytiques d'oscillateurs circadiens proposent des mécanismes complexes, reposant sur un grand nombre d'étapes réactionnelles explicites -afin d'introduire des délais- et implicites - multimérisations qui permettent d'introduire des fonctions de Michaelis avec des coefficients de Hill parfois très élevés. Cependant, la complexité mathématique de ces modèles ne permet pas d'appréhender de façon globale les caractéristiques intrinsèques des oscillateurs circadiens. Afin de pallier à cette difficulté, nous proposons deux modèles minimaux d'horloges circadiennes, l'un pour Neurospora, l'autre pour les mammifères. Ces modèles, uniquement décrits à l'aide de lois d'action de masse, ont l'ambition de mettre en évidence les mécanismes fondamentaux de ces oscillateurs. Malgré leur simplicité, les oscillateurs ainsi construits présentent les caractéristiques principales en terme de déphasage notamment des oscillateurs réels. Nous mettons également en évidence des comportements a priori contre-intuitifs conformes aux observations expérimentales.

A Formal Framework for the Modelling of Biological Regulatory Networks

Vincent Bassano, Gilles Bernot

La.M.I. Université d'Évry Val d'Essonne & CNRS UMR 8042, Tour Évry 2, 523 Place des terrasses de l'agora, F-91000 Évry
{vbassano,bernot}@lami.univ-evry.fr

Abstract

In the field of biological regulation, models obtained from experimental biology are usually complex networks of induction and repression between genes. In R.Thomas modelling, a biological regulatory network can be seen as a discrete model allowing representing qualitative evolution of a system. Due to the development of high throughput genomic, it is now necessary to treat large scale networks. The challenge is to automatically analyze their behaviour. Here, we propose a formal framework to model Biological Regulatory Networks as a computer science interpretation of generalised logical model of genetic network.

In our framework, we have introduced transitions to represent interactions between entities of biological system. Transitions clarify the biological model, reduce the number of parameters taken into account, enhance the expressivity power and give the model a strong modularity. A software environment to support this framework is also described. This software, developed for biologists, can be used to rapidly obtain a prototype describing the behaviour of the system, to produce simulations and to automate proofs of different system properties.

A model of genetic regulation for the variation of mRNA expression

Derek J Raine, Yohann Grondin

Department of Physics and Astronomy, University of Leicester, Leicester, LE1 7RH, UK
jdr@leicester .ac.uk

Abstract

We consider a model genetic regulatory network in which the nodes represent the protein production machinery for each protein and directed links represent regulation by (out-link) or on (in-link) the associated node. Each link may be activating (+) or repressing (–). General properties of the distribution of links that are of interest include the node distribution of in- and out- links and the number of repressing links as a fraction of the total. Genetic networks are studied indirectly by mRNA arrays, which give the mRNA production of the cell over the residence time of an mRNA molecule. Comparison with model networks has focussed on the power law tails in the mRNA distributions for less expressed genes, but another important feature is a peak at intermediate values and a possible second power-law tail. One generic explanation for such distributions is the observation at random times of an exponential growth process, as has been argued in a variety of cases. Here we explore the effect of varying the fraction of repressor proteins on the mRNA distributions of model regulatory networks under some simple assumptions for the network architectures. The model predicts a link between the slope of the tail and the number of repressed genes. There is also phase change in the network dynamics as the percentage of repressor links is increased.

Part 6

HIGH LEVEL COURSES

Hyperstructures

Patrick Amar[1,2], Pascal Ballet[3], Georgia Barlovatz-Meimon[4], Arndt Benecke[5], Gilles Bernot[1], Yves Bouligand[6], Paul Bourguine[7], Franck Delaplace[1], Jean-Marc Delosme[1], Maurice Demarty[8], Itzhak Fishov[9], Eric Fourmentin-Guilbert[10], Joe Fralick[11], Jean-Louis Giavitto[1], Bernard Gleyse[12], Christophe Godin[13], Roberto Incitti[1], François Képès[14], Catherine Lange[15], Lois Le Sceller[8], Corinne Loutellier[15], Olivier Michel[1], Franck Molina[16], Chantal Monnier[8], René Natowicz[17], Vic Norris[8], Nicole Orange[18], Helene Pollard[19], Derek Raine[20], Camille Ripoll[8], Josette Rouviere-Yaniv[21], Milton Saier jnr. [22], Paul Soler[19], Pierre Tambourin[19], Michel Thellier[8], Philippe Tracqui[23], Dave Ussery[24], Jean-Pierre Vannier[25], Jean-Claude Vincent [26], Philippa Wiggins[27] and Abdallah Zemirline[3]

[1]La.M.I. Université d'Évry Val d'Essonne & CNRS UMR 8042, Tour Évry 2, 523 Place des terrasses de l'agora, 91000 Évry, France
[2]Laboratoire de Recherches en Informatique, Université Paris Sud & CNRS UMR 8623, 15 avenue George Clémenceau, 91405 Orsay Cedex, France
[3]Département d'Informatique, EA2215 Université de Bretagne Occidentale BP 809, 29285 Brest Cedex, France
[4]INSERM U492, Faculté de Médecine, 8, rue du Général Sarrail, 94010 Creteil, France
[5]Institut des Hautes Etudes Scientifiques, Bures-Sur-Yvette, France & INSERM U417, Paris, France
[6]Histophysique (EPHE) & Faculté de Pharmacie, 10 rue A.-Bocquel, 49100 Angers, France
[7]CREA, Ecole Polytechnique, 1 rue Descartes, 75005 Paris, France
[8]Laboratoire des Processus Intégratifs Cellulaires, CNRS UMR 6037, Faculté des Sciences & Techniques, Université de Rouen, 76821, Mont-Saint-Aignan Cedex, France
[9]Department of Life Sciences, Ben-Gurion University of the Negev, POB 653, Be'er Sheva, 84105 Israel
[10]Fondation Scientifique Fourmentin-Guilbert, 2 Avenue du Pavé Neuf, 93160 Noisy-Le-Grand, France
[11]Health Science Center, Texas Tech University, Lubbock, TX, USA
[12]Laboratoire de Mathématiques de l'INSA de Rouen, Université de Rouen, 76821 Mont-Saint-Aignan Cedex, France
[13]CIRAD, Laboratoire de modélisation des plantes, Blvd de la Lironde, 34398 Montpellier Cedex, France
[14]ATelier de Genomique Cognitive, Département de Mathématiques, Université d'Évry, 91025 Evry Cedex, France
[15]Spectrométrie de Masse Bio-organique, UFR des sciences, Université de Rouen, 76821 Mont-Saint-Aignan Cedex, France
[16]Faculté de Pharmacie, IBPh - CNRS UMR 5094, 15 Av. Charles Flahault, B.P. 14491, 34093 Montpellier Cedex, France
[17]ESIEE, Laboratoire d'Informatique, Cité Descartes, BP 99, 93162 Noisy le Grand Cedex, France
[18]Laboratoire de Microbiologie du Froid, Université de Rouen, Evreux, France
[19]Genopole Recherche, 2 rue Gaston Cremieux, 91057 Evry Cedex, France
[20]Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK.
[21]Institut de Biologie Physico-Chimique, CNRS UPR 9073, 75005 Paris, France
[22]Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116, USA
[23]Laboratoire des Techniques Imagerie Modélisation Cognition, CNRS UMR 5525, Faculté de Médecine, 38706 La Tronche, France
[24]Center for Biological Sequence Analysis, Dept. of Biotechnology, The Technical University of Denmark, 2800 Kgs, Lyngby, Denmark
[25]Hématologie-Oncologie Pédiatrique, Faculté de Médecine, CHU de Rouen, 76813 Rouen Cedex, France
[26]Polymères, biopolymères et membranes, UMR 6522 CNRS, Université de Rouen, 76821 Mont-Saint-Aignan Cedex, France
[27]Genesis Research and Development Corporation Limited, Parnell, PO Box 50, Auckland, New Zealand

Author for correspondence: Vic Norris, email: vjn@univ-rouen.fr, Telephone 00 33 235 14 6908, Fax 00 33 235 14 7020

Abstract

New concepts may prove necessary to profit from the avalanche of sequence data on the genome, transcriptome and proteome and to relate this information to cell physiology. Here, we focus on the concept of hyperstructures in which a variety of types of molecules are brought together to perform a function. The processes responsible for hyperstructure formation include changes in enzyme affinities due to metabolite-induction, transertion, and elevated local concentrations of proteins and their binding sites on DNA and RNA. We review the evidence for the existence of hyperstructures responsible for the initiation of DNA replication, the sequestration of newly replicated origins of replication and for cell division. We interpret cell cycle progression in terms of hyperstructure dynamics. Experimental techniques exist that can be used to study hyperstructures and we review some of the ones less familiar to biologists. Finally, we speculate on how a variety of *in silico* approaches could be combined to develop new concepts in the form of an *Integrated* or *Imaginary* cell – *I-cell* – which would undergo selection for growth and survival in a world of artificial microbiology.

1 Introduction

Molecular biology and biochemistry have provided a wealth of information about how RNA polymerases transcribe DNA into RNA and how ribosomes then translate mRNA into proteins, about the nature of those proteins and lipids that form membranes, and about other important molecules such as polyamines, polyphosphates and poly-β-hydroxybutyrate. Model organisms such as the bacterium *Escherichia coli* are invaluable in making sense of this information. The 4.6 Mb genome of *E. coli* K-12 MG1655 has been sequenced (Blattner *et al*., 1997) and was found to have 4288 protein-coding genes of which 38% had no attributed function. Other strains of *E. coli* have genes that range in number from 4085 (K-12 W3110) to 5361 (O157 RIMD) whilst the bacterium *Streptomyces coelicolor* has 7846. These numbers are in the same range of those of unicellular eukaryotes such as the fission yeast, *Schizosaccharomyces pombe*, with 4824 genes and the budding yeast, *Saccharomyces cerevisiae*, with 5885. But even when all genes are ascribed functions, how are we to interpret this information and use it to predict phenotypes? The challenge is to understand how cells organise their myriad constituents and processes. To explain how the concept of hyperstructures may help us, here we briefly review the bacterial cell cycle, focussing on the problem of division, and then discuss hyperstructures. We do this in the light of different questions: Why might a hyperstructure language be useful? What are hyperstructures? How do they form? How do they interact? How might they guide cells through state space to control growth, adaptation, differentiation and the cell cycle? We then discuss how the hyperstructure concept may help in exploiting the information provided by genome sequencing and how it may be tested using a combination of cellular automata and multi-agent systems. Finally, we advocate an approach to the study of biological complexity *via* construction of cell *in silico* based on hyperstructures. Such a cell would be an important step towards the construction of an integrated cell, *I-cell*, that would bring together many of the processes believed to determine the structure and phenotype of real cells.

2 The bacterial cell cycle

The principal events in the bacterial cell cycle include:

- *Initiation of chromosome replication from a single origin of replication*
- *The sequestration of newly replicated origins of replication*
- *Chromosome separation*

• *Chromosome segregation*
• *Cell division*
• *Inactivation of the division site*

In the case of cell division, it is still not clear how this event is timed, positioned and coupled to other events. The earliest known protein to act in *E. coli* is the tubulin-like FtsZ which migrates from the cytoplasm to a mid-cell location on the membrane where it assembles into a ring-like structure and where it recruits other division proteins. What lies upstream of FtsZ? Is it yet another protein or is it something else? We have shown that FtsZ can interact directly with phospholipid membranes in the absence of other proteins (Alexandre *et al* ., 2002). This is consistent with a major role for membrane dynamics in the regulation of the cell cycle as is the finding that membrane domains around the chromosomes differ from the domain at the future site of division (Fishov and Woldringh, 1999), this latter presumably being related to the large domains of cardiolipin observed at the division sites and poles (Koppelman *et al* ., 2001; Mileykovskaya and Dowhan, 2000).

The constraints on a solution to the division problem for *E. coli* are that the division site must be:

• *in the right place – midcell – to give daughters of similar sizes*
• *between chromosomes – to avoid producing a DNA-less cell*
• *formed at the right time in the cycle – perhaps to give the right DNA/mass ratio?*
• *formed at the right rate – to avoid, for example, cells getting bigger and bigger*
• *of the right nature – to allow membranes to curve and fuse whilst controlling ion and lipid fluxes*

It is in the context of trying to find a solution to this problem that we present hyperstructures.

3. Why invoke a hyperstructure language?

Cells survive and sometimes grow by somehow orchestrating millions of molecules of thousands of types to adapt to the environment and to proceed through the cell cycle. This entails cells solving the combinatorial problem of negotiating the immensity of state space since if each gene in *E. coli* were in either an on (transcribed) or an off (untranscribed) state, there would be 2^{4000} or 10^{1200} on-off patterns of gene expression (Kauffman, 1996). But there is more than this, there is also the epigenetic trap – cells in a population should not all have the same phenotype (else, for example, a single catastrophe would be more likely to wipe them all out). Exploring state space effectively boils down to: How can cells be both efficient and robust? We argue that the answer is that cells rely on an intermediate level of organisation – *hyperstructures.*

4. What are hyperstructures?

The sort of hyperstructure that concerns us here is the non-equilibrium type which is a large, extended structure of diverse molecules – genes, mRNAs, proteins, ions, lipids – that depend on a flow of energy/material for their existence. These hyperstructures are assembled to serve a specific function and are disassembled when no longer functional (Figure 1).

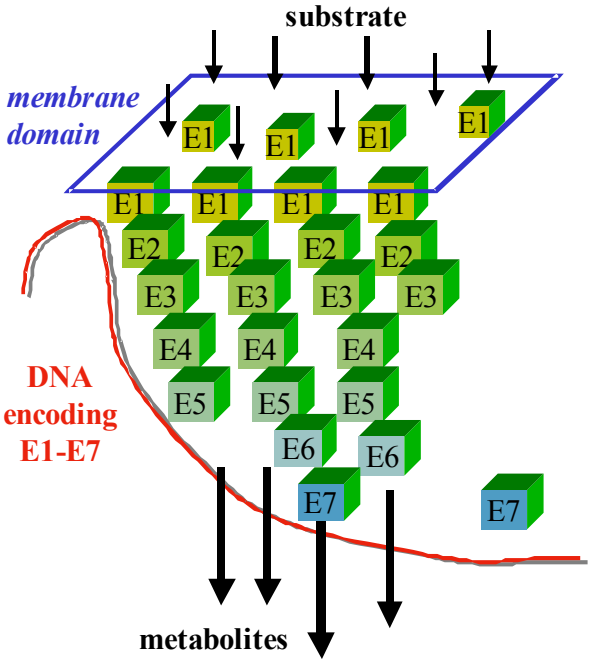


Figure 1: Formation of a non-equilibrium hyperstructure due to changes in the affinity of its constituent enzymes for one another. Enzymes E1 can only diffuse in the plane of the membrane whilst the other enzymes, E2 to E7 diffuse in the cytoplasm. The binding of a substrate, such as a sugar, to the E1 enzymes leads to an increase their affinity for one another and their assembly into an E1 domain. On binding its substrate, each enzyme in the pathway acquires an increased affinity for the following enzyme. This results in the assembly of metabolons E1 to E7 and the assembly of the hyperstructure (here, a group of metabolons). Note that transcription of the genes encoding E1 to E7 and the simultaneous translation of the mRNA may help the assembly of the hyperstructure (see plate 17).

They are essential for growth and for progression through the cell cycle. (There are also equilibrium or quasi-equilibrium hyperstructures that are not dependent on a flow of energy for their existence, that are important for survival in harsh conditions when growth is not possible and that can have an intimate relationship with non-equilibrium hyperstructures. They include condensed structures in chromosomes needed for bacteria to withstand irradiation (Wolf *et al.*, 1999) and to separate chromosomes (Bouligand and Norris, 2001).) The hyperstructure concept is relevant to the organisation of both prokaryotic and eukaryotic cells but here we confine ourselves to bacteria and in particular *E. coli*. The examples below are drawn primarily from the cell cycle and from metabolism. It should become clear that the archetypal bacterial hyperstructure contains genes being transcribed, their mRNAs being translated and the nascent proteins being inserted into the membrane to form a domain with a characteristic lipid composition (Figure 1).

Examples of possible non-equilibrium hyperstructures include:

4.1 Cell cycle.

The cell cycle comprises the initiation of DNA replication and the temporary sequestration of the newly replicated origins of replication (to prevent multiple rounds of replication), the separation and segregation of the chromosomes into the future daughter cells, and the division of the parental cell between the chromosomes, probably followed by the inactivation of the division apparatus.

An initiation/origin of replication hyperstructure. This comprises the DnaA protein and certain of the sites on DNA to which it binds. In *E. coli*, the ATP-bound form of DnaA is required for initiation of replication *in vitro* whilst the ADP-bound form is inactive (Castuma *et al*., 1993). Both ATP-DnaA and ADP-DnaA bind to 9mer DnaA boxes, TTA/TTNCACA, but only ATP-DnaA protein binds in addition to a 6mer site, AGATCT (Speck *et al*., 1999). DnaA is associated with the membrane *in vivo* (Newman and Crooke, 2000) and initiation requires a membrane with a particular phospholipid composition (Fralick and Lark, 1973) and domain structure to activate DnaA (Castuma *et al*., 1993; Xia and Dowhan, 1995) as well as the polymerisation of DnaA (Weigel *et al*., 1999). In one hypothesis, this putative hyperstructure has a transient existence that depends on the dynamics of other hyperstructures. It is clearly a non-equilibrium structure, indeed, it is disassembled after accomplishing the act of initiation (Norris *et al*., 2002a).

A DNA replication hyperstructure. This comprises the SeqA protein, key enzymes in DNA replication, and the genes that encode them (Norris *et al*., 2000). SeqA sequesters newly replicated origins and is found in clusters (Onogi *et al*., 1999). It binds to GATC sequences and it polymerises. The non-equilibrium nature of this hyperstructure results from its dependence on the energy-consuming process of replication with SeqA binding preferentially to hemi-methylated GATC sites in genes encoding enzymes responsible for DNA replication, topology, repair and precursor synthesis such as ribonucleoside diphosphate reductase (Guzman *et al*., 2002).

A cell division hyperstructure. This comprises the ten or so division proteins plus enzymes involved in peptidoglycan synthesis plus the genes that encode them (Buddelmeijer *et al*., 1998; Norris and Fishov, 2001). Many of these genes are located together and are transcribed together in the *dcw* cluster at the 2 min position on the chromosome. Recently, we have speculated that cell division may involve the conversion of lipid domains, positioned between the chromosomes, into tubes and vesicles in a dynamic, Golgi-like structure (Norris et al., in preparation). This is a non-equilibrium structure in that it requires (i) the energy-consuming processes of transcription and translation to bring these genes together at the membrane along with the nascent proteins (see below) and (ii) continuing and major changes in membrane structure.

4.2 Other hyperstructures.

These include a DNA compaction hyperstructure possibly involving the MukB protein which can form foci (Ohsumi *et al*., 2001); a nucleolus-like hyperstructure for ribosome synthesis (Lewis *et al*., 2000; Woldringh and Nanninga, 1985); a chemotaxis hyperstructure comprising chemotactic receptors such as Tsr with the kinase CheA and the transducing protein CheA (Bray *et al*., 1998; Stock and Levit, 2000) plus, we propose, the genes encoding the abundant chemotaxis proteins (and hence the necessity for hydrolysis of ATP and GTP during transcription and translation). Factors in the formation of possible hyperstructures for transport and glycolysis (Mitchell, 1996; Velot *et al*., 1997) are discussed below.

5. How do non-equilibrium hyperstructures form?

There are several complementary possibilities:

5.1 Metabolite-induction

• Non-equilibrium hyperstructures form when the cell is actively engaged in processing substrates and disappear when they are not

• These hyperstructures include enzymes in the same pathway and their genes
• Formation of certain of these hyperstructures may involve an interplay between diffusion in 2-D and 3-D in the sense that enzymes confined to domains in the 2-D membrane interact with other enzymes or groups of enzymes diffusing in the 3-D cytoplasm (Figure 1)

The evidence consistent with this scenario has been advanced for the existence of metabolons which are assemblies of the enzymes that act in succession in a pathway (Velot et al., 1997). Of course, such metabolons may themselves associate into larger hyperstructures.

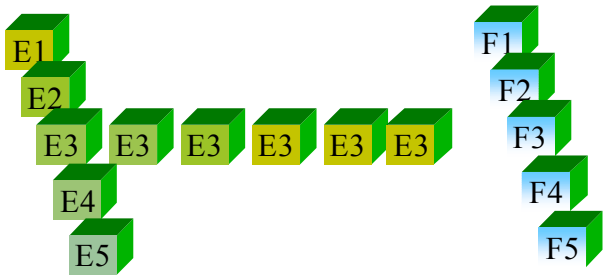


Figure 2: Horizontal links aid the assembly of a hyperstructure. Oligomeric protein E3 may bind together two identical metabolons (E1-E5 to E1-E5) or two different ones (E1-E5 to F1-F5). In the former case, E3 plays a role in the assembly of an individual hyperstructure whilst in the latter case E3 plays a role in the interaction between two different hyperstructures (see plate 18).

In the case of secretion, substrate binding promotes assembly of the 3 components of the ABC exporters of Gram negative bacteria e.g. in *Erwinia chrysanthemi* the substrate (protease) binds to PrtD (an ABC protein) which then binds to PrtE (membrane fusion protein) and which binds to PrtF (outer membrane protein) (Letoffe *et al*., 1996). In the case of glycolysis, the glycolytic pathway can be extracted as an equimolar complex of 1.65 MDa that reveals compartmentation of substrates (Mowbray and Moses, 1976). Evidence has also been obtained for the existence of metabolons which are assemblies of the enzymes that act in succession in a pathway (Mitchell, 1996; Velot *et al*., 1997). Of course, such metabolons, which may or may not be non-equilibrium structures, may themselves associate into larger structures or hyperstructures. In the case of import, sugar-specific phosphotransferase system permeases consist of EIIC and EIID in the membrane and EIIA and EIIB in the cytoplasm; EIIA is phosphorylated by HPr in a reaction catalysed by EI with P from phosphoenolpyruvate; E2s+E1+HPr are proposed to form a complex in response to the presence of the appropriate substrate as part of a 'metabolite-induced metabolon' event (Norris *et al*., 1999). An extension of this would be that a PTS hyperstructure forms when the cell is actively engaged in processing substrates and disappears when it is not. The idea here is that successive enzymes in the same pathway can be activated by their substrates to bind to one another in a heteropolymeric organisation that is sequential (or *vertical* insofar metabolic pathways are often drawn vertically with metabolites entering at the pathway at the top of the page and products as leaving at the bottom). A complementary idea is a single species of enzyme can be activated to oligomerize by substrate (Torshin, 1999); indeed, the full enzymatic activity of glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate mutase and enolase – all glycolytic enzymes – results from their association. Again, this homopolymeric or *horizontal* organisation could help nucleate and stabilise hyperstructures (Figure 2).

5.2 Local concentrations

The phenomenon of oligomeric proteins binding to specific sites on DNA has been invoked to explain the operation of the *lac* and lambda repressors (Revet *et al*., 1999). It might also be invoked to explain the sequestering of newly replicated origins of replication by the protein

SeqA (Onogi *et al*., 1999) and the binding of DnaA to its sites. There are variations of this theme with, for example, the possibility that proteins such as the histone-like protein HU, which binds to both RNA and DNA (Balandina *et al*., 2001), could play important roles (see below) (Figure 3).

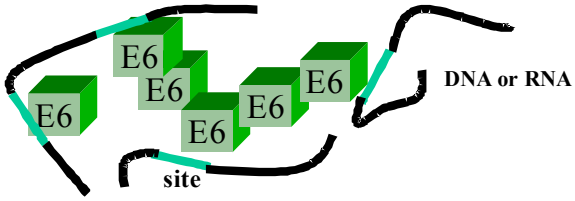


Figure 3: Local concentrations of oligomeric proteins can promote hyperstructure assembly. Protein E6 binds to its site (green) present in DNA or RNA to produce a region of the cytoplasm enriched in both E6 and its sites (see plate 19).

5.3 Transertion

Transertion is the coupled transcription, translation and insertion into and through membranes of proteins (Binenbaum *et al*., 1999; Lynch and Wang, 1993). The cytoplasmic membrane is composed of a wide variety of lipids and proteins so if these proteins have lipid affinities small proteolipid domains form.

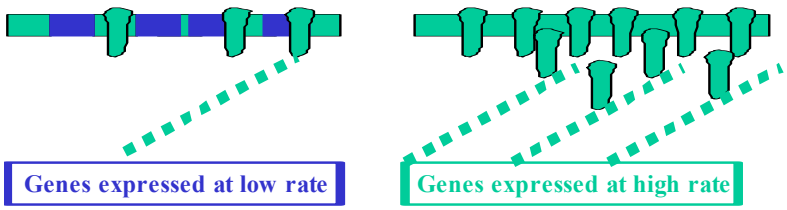


Figure 4: Transertion can nucleate hyperstructure assembly. Transertion, alias the coupled transcription, translation and insertion into and through membranes of proteins, may enrich a region of the membrane in the lipids (green) for which the proteins have an affinity. At a critical density of inserted nascent proteins, small proteolipid domains fuse into large ones and so nucleate hyperstructure assembly (see plate 20).

High rates of transertion may create a critical density of inserted nascent proteins that is sufficient for small proteolipid domains to fuse into large ones and so nucleate hyperstructure assembly (Norris, 1995) (Figure 4). For example, it might be supposed that the high density of transertion of the ATP synthetase components, which have lipid affinities (Arechaga *et al*., 2000; Ksenzenko and Brusilow, 1993), would result in assembly of an ATP synthesis hyperstructure.

5.4 Translated mRNA is protected from RNases and enzymes in metabolons are protected from proteases

It has been suggested that enzymes in complexes are more likely to escape proteolytic degradation than when not in complexes (Miller, 1996). An extension of this idea is that the partitioning of enzymes into a hyperstructure protects them from proteases – providing the latter are excluded from the hyperstructure. Hence an enzyme which has been assembled into a hyperstructure *because* of its activity is thereby preserved (i.e. active enzymes are preferentially protected). A similar argument is that mRNA translated within a hyperstructure could be preferentially protected from RNases on the outside of the hyperstructure.

6. How do hyperstructures interact?

6.1 Shared lipid affinities creates shared membrane domains

It can be argued that proteins with lipid preferences may congregate with those lipids in a positive feedback fashion to form the membrane domain part of a hyperstructure (see *5.3 Transertion*). Similarly, it might be expected that hyperstructures characterised by enrichment for a particular lipid would also tend to associate.

6.2 Shared binding proteins create shared cytoplasmic compartments

The idea is that certain abundant proteins may participate in the assembly of several different types of hyperstructures. This would enable a synergy whereby the progressive formation of a group of hyperstructures responsible for a set of functions would aid the recruitment of other related hyperstructures fulfilling complementary functions. Candidates for these proteins include IHF, FIS, and HU (for references see (Ussery *et al*., 2001)). IHF can modulate the transcriptional activity of promoters by influencing the looping of upstream DNA; the consensus site of IHF binding, YAACTTNTTGATTTW, lies within many repetitive extragenic palindromic sequences. FIS binding to upstream regions can enhance the transcription of highly expressed genes; the consensus for the FIS binding site is weak with estimates of its numbers ranging from 6 to 68000. HU binds to DNA with no evident sequence preference and, in so doing, influences the interaction of regulatory proteins with their specific sites on the DNA (Bonnefoy and Rouviere-Yaniv, 1992); HU also recognizes certain specific structures of both DNA and RNA with very high affinity and, for example, binds to the mRNA for RpoS (Balandina *et al*., 2001; Kamashev and Rouviere-Yaniv, 2000). In addition, there are over a 100 known activators and repressors of transcription in *E. coli* (Ouzounis *et al*., 1996) and it may be expected that these will control the synthesis of certain oligomeric proteins important in the assembly of different – but complementary – hyperstructures.

6.3 Gene distribution

There are strong compositional asymmetries in codon and amino acid usage depending on the orientation of the genes with respect to DNA replication and on the nature of the proteins encoded. This has led to predictions of different compartments for the syntheses of different proteins (Danchin and Henaut, 1997). A recent analysis of the distribution of genes in *E. coli* and *Bacillus subtilis* supports the hypothesis that one parental strand carries the set of genes required for growth in good conditions whilst the other strand carries those required for survival (resisting stress) in adverse ones; the idea is that one set of hyperstructures is associated with one strand whilst a different is associated with the other such that the segregation of these sets of hyperstructures underpins the segregation of the daughter chromosomes (Rocha *et al*., in preparation).

6.4 Water preferences

Water exists as nm-sized microdomains of differing structure and density that must have different solvent properties (Robinson *et al*., 1999; Wiggins, 1990). The difference in density appears to be of the order of 30% (Cho *et al*., 1997). Such differences should affect the distribution and activity of cellular constituents near, for example, the surfaces of membranes (Mayer and Hoppert, 1996). An important but difficult question is the extent to which the water preferences of the constituents of hyperstructures might determine hyperstructure formation and interaction.

6.5 Oscillations/vibrations

There are numerous oscillatory processes in eukaryotes of which the oscillation of protons and of NAD(P)H in neutrophils is particularly exciting (Petty and Kindzelskii, 2001). Such oscillations are candidates for playing global as opposed to local organising roles. Relating them to the dynamics of hyperstructures is a problem that has still to be addressed. In bacteria, several oscillatory systems have been discovered. For example, the Min system, which is involved in the selection or inactivation of the division site, oscillates with a periodicity of around 1 minute in *E. coli*; several explanations for these oscillations have been advanced (Meinhardt and de Boer, 2001; Norris *et al*., 2002b).

7. Cell division

The regulation of cell division can now be considered in terms of the dynamics of hyperstructures. It has been argued that one of the functions of the bacterial cell cycle is to generate daughter cells with different phenotypes since this would allow the population to both explore all the possibilities for growth offered by the environment and be ready for a sudden catastrophic change (Norris *et al*., 2002a; Segre *et al*., 2000). In this scenario, during the run-up to initiation, the mass to DNA ratio increases and certain hyperstructures become 'stronger' by attracting ever more of the cell's resources (such as the transcriptional and translational apparatus) whilst other hyperstructures are weakened and disappear (Norris *et al*., 2002a) (Figure 5). This results in a drop in the diversity of hyperstructures, some of which release DnaA as they dissociate, a DnaA-initiation hyperstructure forms, and replication of the chromosome begins. Now suppose that short FtsZ polymers are associated with glycolytic and other hyperstructures so that FtsZ is effectively sequestered (noting that, at least in eukaryotes, tubulin is associated with glycolytic enzymes (Lloyd and Hardin, 1999)). This leads us to consider three complementary possibilities. One is that the FtsZ-sequestering hyperstructures are temporarily disrupted by chromosome replication to release FtsZ which can then participate in division. Another possibility is that the changing activity of the phosphotransferase system/glycolytic hyperstructure directly leads to its own disassembly (for example, its capacity might exceed demand and lead to feedback inhibition) and releases FtsZ. This would be consistent with the advance in divisions in synchronous cultures of *E. coli* induced by addition of the non-metabolisable, glucose analogue α-methylglucoside (Fishov, 1994) and the delay in divisions induced by transfer to a rich growth medium (Kepes and Kepes, 1985). Finally, FtsZ may exist as a delicate cytoskeleton that extends throughout the cell, as described for chloroplasts (Kiessling *et al*., 2000), that may act as a metabolic sensor leading to its redistribution in cell division.

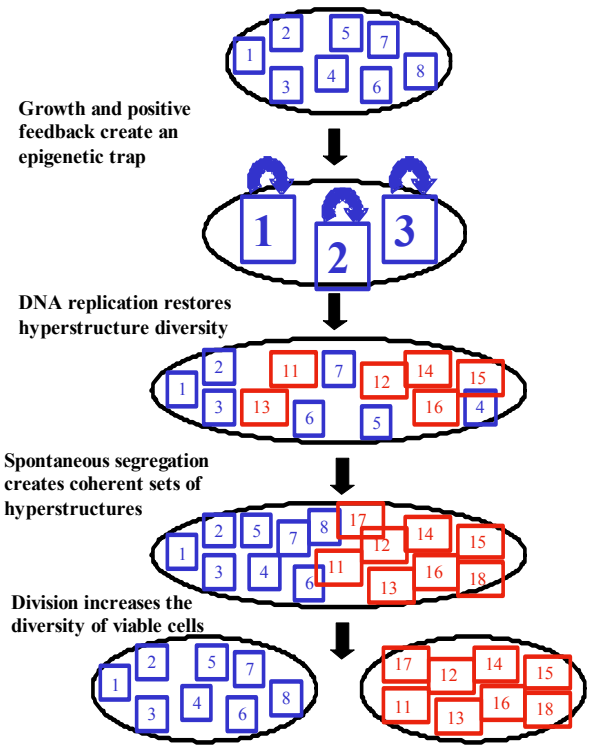


Figure 5: Cell cycle progress as a state cycle of hyperstructures. Rectangles represent non-equilibrium hyperstructures each performing one function. Blue rectangles correspond to hyperstructures with a common set of lipid (or other) preferences whilst red rectangles correspond to hyperstructures with a different set of preferences (see plate 21).

Before trying to put it all together, we should bear in mind that, all else being equal, the rates of transcription of two copies of the same gene diverge if this gene is –vely regulated *in trans* but +vely *in cis* (Norris and Madsen, 1995). The –ve regulation *in trans* could result from a repressor diffusing through the cytoplasm to each separate stretch of DNA whilst the +ve regulation *in cis* could result from an RNA polymerase transcribing a gene making this particular stretch of DNA accessible to another polymerase. This leads to the conclusion, surprising for many microbiologists, that two identical chromosomes in the same cytoplasm (which contain many such genes) *therefore* have different patterns of gene expression. The same argument can be made in terms of hyperstructures: a set of genes is expressed from one chromosome to form part of a hyperstructure; this assembly involves positive feedback between the constituent ions, lipids, proteins and nucleic acids since as the density of one constituent in a region increases, the probability increases that the density of another constituent will also increase. In the context of a cell in which hyperstructures compete for existence (that is, negative regulation *in trans*), the result is a highly structured, asymmetric cell in which each future daughter cell has a different set of hyperstructures associated with it and these sets differ in their composition of lipids, ions, water structures, proteins, mRNA and expressed genes (Figure 6). At present, it is difficult to discriminate between the different ways in which the principal proteolipid domains around the chromosomes could create a division site (red, dotted arrows in Figure 6). In one scenario, the site would simply consist of the *interface* between the two domains whilst in the other scenario, the site would consist of a distinct *domain* between the principle two domains. There are, of course, permutations of these possibilities.

The essence of our proposal is that hyperstructure dynamics could achieve:

• separation of chromosomes during replication

• differentiation of both chromosomes and membrane
• the right place for a site to attract and activate division enzymes (between the chromosomes)
• the right time for the creation of a division site (after chromosome segregation)
• the right nature for a division site - a potential non-bilayer
• coupling between replication, segregation and cell division
• a calcium flux (down the concentration gradient)
• orchestration of membrane-activated kinases, proteases *etc.*

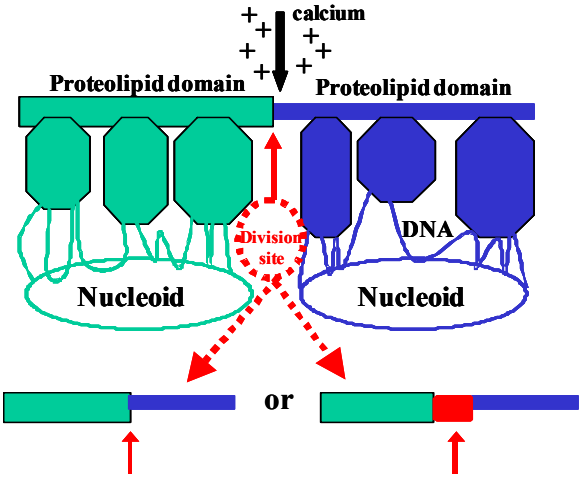


Figure 6: The spatiotemporal control of cell division by hyperstructures. The hyperstructures (the green or blue polygons) form one of two sets depending on the common preference within a set for lipids, ions, proteins etc. Each set is associated with a chromosome and is present in the future daughter cell. The division site is in the cytoplasmic membrane (thin rectangles) at the interface between these sets indicated by the arrow. Two possibilities for the structure of the division site (red arrows) which may be between the principal domains (blue and green) at either the interface or a separate, specific domain (red) (see plate 22).

8. The advantage of organisation at the level of hyperstructures

It has been observed that the difficulty of administering a laboratory is proportional to the square of the number of members of the laboratory, N^2 (Bok, 1983). This difficulty, D, is reduced if the individuals are put into N_1 groups such that D equals the square of the number of groups (to reflect group interactions) plus the square of the number of individuals in each group N_0^2 (to reflect interactions within groups) times the number of groups:

$$D = N_1^2+(N_0^2)N_1$$

Hence

$$D = (N/N_0)^2+(N_0^2)N/N_0$$

And

$$D = N^2/N_0^2+N_0N$$

To minimise D,

$$\delta D/\delta N_0 = -2N^2/N_0^3+N = 0$$

Hence the difficulty is at a minimum when

$$N_0 = (2N)^{1/3}$$

This formula helps to give us a feel for the numbers of hyperstructures that may exist in a cell. Just considering proteins, for example, a bacterium containing of the order of a million interacting proteins would be expected to have around a hundred hyperstructures. The existence of this intermediate level of organisation therefore means that the problem of generating a

limited number of coherent phenotypes that are adapted to survival and/or growth is greatly simplified. Navigation through the immensity of state space becomes a choice between 100 or so hyperstructures rather than 4000 plus genes – 2^{100} on-off combinations rather than 2^{4000}. To generate a coherent phenotype, for example, enzymes appropriate for growth in cold oxygenated conditions should not be synthesized in the same cell at the same time as those for growth in hot anaerobic conditions. Coherence can be achieved because cells can manage the relatively few common factors required to bring together a particular set of hyperstructures. The existence of hyperstructures also allows, we speculate, bacterial cells to regulate DNA replication and cell division so as to create heterogeneous populations that can both grow and survive unexpected challenges.

9. Using the hyperstructure concept to exploit sequence data

Of the numerous *in silico* approaches possible, we focus here on cellular automata which are used to model many physical and biological phenomena (Vichniac, 1984). Once the units that constitute the automata have been assigned initial states, the evolution of these states can then depend on both the previous history of the state and on the state of neighboring units. Hence, cellular automata can be particularly suitable for modeling the dynamics of interactions between molecules in 3 dimensions. We now use cellular automata to illustrate how they might be used to model the effects on hyperstructure assembly of the following:

9.1 Metabolite-induction

To determine the values of the parameters governing the formation of hyperstructures in bacteria, we have constructed a preliminary version of a cellular automaton program (with features of multi-agent systems) that simulates the dynamics of the localization of the PTS and glycolytic enzymes in both a 2 dimensional membrane and a 3 dimensional cytoplasm (Le Sceller *et al*., 2000). Each unit volume represents a 10nm x 10nm x 10nm cube in a cell that can have a maximum volume of 200 x 200 x 200 unit volumes or $8\mu m^3$. This is more than sufficient to represent *E. coli* which in certain growth conditions has a volume of 2 cubic microns. Each cubic unit volume in the membrane is surrounded by 8 other unit volumes and each unit volume in the cytoplasm is surrounded by 26 others. At each time step, all enzymes are considered in a random order. Each can move into a free neighboring unit volume. In this preliminary study, there was a structuring of both membrane and adjacent cytoplasm and hyperstructures were generated containing up to 500 enzymes.

9.2 Transertion

To model the anchoring effect of transertion on nascent proteins (Figure 4), a proportion of the PTS Enzymes II (for example) could be permanently confined *in silico* to a patch of the membrane. An important parameter may therefore be the *area* over which these proteins are inserted. It is not easy to obtain this area experimentally with current techniques (but see the NanoSIMS below). However, this may be an instance when the simulation reveals whether hyperstructure formation is very sensitive to the area of transertion and therefore whether energy should be invested in performing the relevant experiments.

9.3 Lipid preferences

The cosegregation of proteins with the lipids for which they have pronounced affinities is a potent way to produce domains. This process may be simulated in the 'membrane' of cellular automata given these affinities. Below (*10.3*), we suggest a series of experiments that could

lead to consensus sequences for lipid binding and hence a way, ultimately, to convert sequence information into the 'lipidome' and facilitate the simulation of the distribution of all membrane proteins.

9.4 Local concentrations

Using cellular automata to model local concentrations might exploit knowledge of DNA-binding proteins and their sites providing DNA can also be introduced into the model. One way to achieve this would be to divide the chromosome into chunks comparable in size to proteins. Each chunk would be constrained in its diffusion by a function inversely proportional to the distance between the chunk in question and another chunk. It may also prove necessary to make efforts to model reptation, the constrained movement of polymers in a crowded solution.

9.5 DNA distribution

DNA curvature, flexibility and stability have been analysed for 18 fully sequenced bacterial genomes (Pedersen *et al*., 2000). This reveals many significant structural features including a set of 20 regions with identical and extreme structural properties that are proposed to function as topological domain boundaries. These features are presumably related to the properties of proteins such as HU (see *6.2*) which binds preferentially to unusual structures such as kinked or cruciform DNA (Bonnefoy *et al*., 1994; Kamashev and Rouviere-Yaniv, 2000). The challenge is to translate this information into a dynamic 3-D model taking into account that much of the DNA is probably in a cholesteric form. One model that might be tested *via* cellular automata (as in *9.4*) is that HU both binds to these curved regions and self-associates such that curved regions are stacked at the edges of twisted liquid crystalline regions. In such a model, the terminus region, which has high curvature, low flexibility and low helix stability (Pedersen *et al*., 2000), might be expected to exhibit a distinctive packing.

9.6 Parallel approaches

In an activity-based vision of the cell, only a subset of its constituents is important in determining the phenotype of the cell at any one time (Norris, 1998). This subset comprises those constituents that are *active* where *active* is considered to mean being transcribed for a gene, being translated for a mRNA, and catalysing a reaction for an enzyme. Belonging to this active subset requires a competition between constituents that were active in the previous time period (the *status quo* factor) and constituents that act in synergy with one another (the coherence factor).

In this section, we describe a new implementation of cellular automata or units based on the related idea that only a few unit volumes are potentially active, that is either contain a molecule or a have a neighbouring unit containing a molecule. The advantage is that memory is not needed to store these empty units. This leads to a time and memory efficient approach for computing the successive generations of the units. The overall state of the system is determined by the content of all the units at a given time. Computing the next generation means determining the new state of the system after the application of all the local rules to each unit. This process must not depend on the order the units are examined and, ideally, each unit is treated independently of all the other units. The standard way to represent the 3-D space is to use a array of structures to address each unit that often contains only a number. Using this method, it is easy to determine the neighbourhood of a unit by a simple transformation of its coordinates, and then access the array to get the values of the neighbouring units. The major drawback is that we must store *all* the units, even the empty ones.

In our approach, we also represent the space by a three coordinate system, but we store in the computer memory *only* the active or potentially active units (i.e. those that are filled or next to filled units). This reduces the memory cost and allows us either to reduce the size of each unit to have a more accurate simulation, or to simulate a larger space. The potentially active cells are stored in a hash table which allows a very fast access time, comparable to the access time of a 3-D array, if a good hash function is used along with an adapted strategy to resolve collisions. This low cost implementation of the state of the system can be used to reduce the time used to compute each generation if an extra cost is paid by duplicating the representation of the space: the local rules are applied to each active unit using the values from the first, *current* space and the result is stored in the second, *new* space. After all the units of the first space have been processed and the second space is complete, the second space becomes the current space and the next generation can be computed.

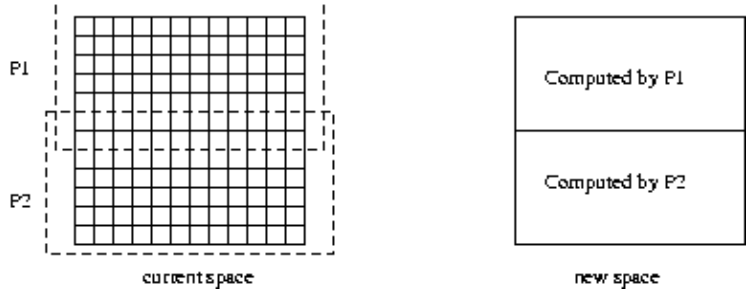


Figure 7: A two process example of parallel processing. Since each process does not need to access the space of the other one (except the boundary of the current space), each part can be stored locally in a multi-computer networked environment. The boundary of each part is the only information to be shared (i.e. transmitted between the computers).

Since the current space is only accessed for *reading* values whilst the new space is only accessed for *writing* results, the current space can be freely accessed by multiple processes without synchronisation. The new space can be split into parts that can be computed separately on a multiprocessor with a consequent dramatic reduction in computation time. Each process requires its own part of the current space but also acts on a surrounding layer of single units in the parts treated by other processes. Since each process only accesses the part of another process at the boundary, each part can be stored locally in a multi-computer networked environment.

In the 12 by 12, 2-D example (Figure 7), process P1 only needs to access the first 7 lines (0 to 6) of the current space to compute the first 6 lines of the new space, while process P2 needs to access the last 7 lines (5 to 11) of the current space to compute the other half of the new space. Since there is no read/write conflict between P1 and P2 no synchronization is needed. This is another advantage of the inherent parallelism of this implementation.

9.7 Hyperstructure movements and reactions

Interactions *between* hyperstructures are proposed to result in a pre-divisional cell with one set of hyperstructures in one half the cell and a different set in the other half. Such sets of hyperstructures may be formed on the basis of common lipids, ions, binding proteins and/or water properties. Movements of hyperstructures are nicely illustrated by the SeqA-replication hyperstructure that, during the cell cycle, goes from a single focus to two foci that then migrate to the one-quarter and three-quarter positions (Ohsumi *et al*., 2001; Onogi *et al*., 1999). To model how interactions between hyperstructures might lead to redistribution of hyperstructures within the cell, we consider a cellular automaton model in which several hyperstructures can be

represented simultaneously in a coarse-grained way (so that the units are bigger than single macromolecules).

The idea presented in this section entails providing local rules to reproduce molecular reaction and diffusion using cellular automata. The difference between our approach and typical reaction-diffusion processes is that the molecule concentration (in a specific position) is boolean: *true* if there is a set of molecules, *false* if there are none. One of the simplest systems has only one type of molecule on a 2-D grid (environment). Focusing on a particular molecule and its neighbourhood, it is clear that a unit plus the 8 adjacent units is a square of side 3 units. If we suppose that the molecule can move or stay in the same place, the molecule will have 9 possible positions (Figure 8).

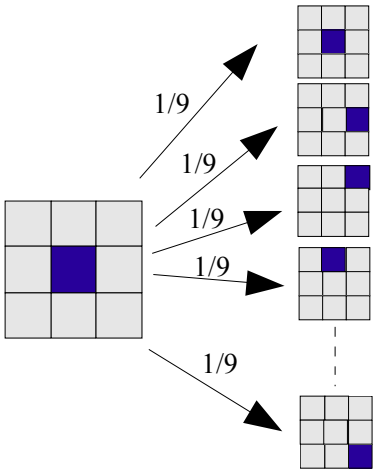


Figure 8: Movement on a 2-D grid. A molecule at time *t* can choose between 9 positions at time *t+1*. A molecule at time *t* can choose between 9 positions at time *t+1*.

With cellular automata, the state of a unit depends on its neighbourhood. Thus, a local rule must be used to determine whether a unit in the 2D-grid becomes *true* (has molecules) or *false* (is empty). The idea is to invert the arrow direction in the previous figure. Thus, if an empty cell is adjacent to one filled cell, it has a probability of $1/9^{th}$ to become filled. We can apply this to any neighborhood.

Given a unit in a 2D-grid, the probability of the unit to become *true* (filled) is $p = n / 9$ where n is the number of filled units into its neighbourhood. More generally, for a *dim*-dimensional environment (a grid of dimension *dim*), the probability of one cell to become *true* is:

$$p = n / N \qquad (1)$$

where $N = 3^{dim}$. Figure 9 shows the results of this local probabilistic rule (rule 1) with a *60x60* grid at three different times.

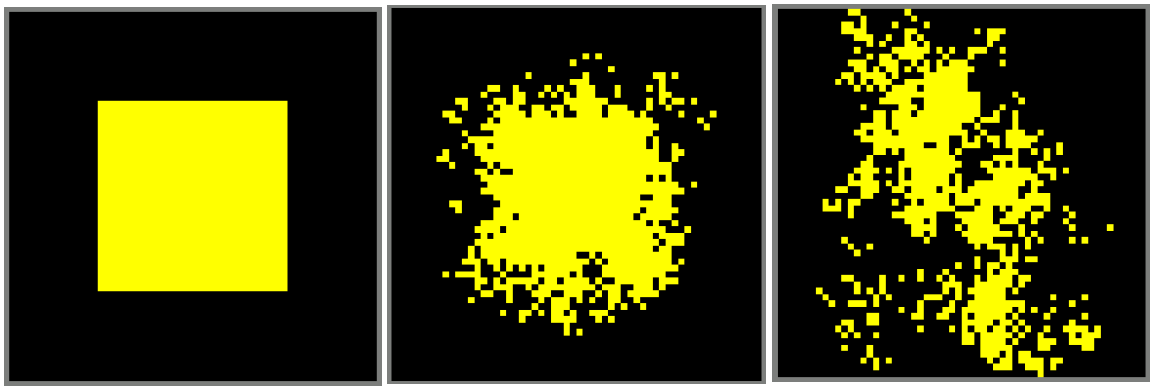


Figure 9: Diffusion in a cellular automata system. Empty units are black and filled units are yellow. States at successive times (t=0, 10 and 100) are shown.

To construct a multi-molecule hyperstructure, the next stage consists in putting together different types of molecules. In this case, the value of a unit is not a boolean but an integer included between *0* and *nb* (number of types of molecule). In this way, rule (1) becomes

$$\begin{aligned} p_0 &= n_0 / N \\ p_1 &= n_1 / N \\ p_2 &= n_2 / N \\ &\dots \\ p_{nb} &= n_{nb} / N \end{aligned} \qquad (2)$$

where n_i is the number of molecules of type i into the neighbourhood and $N = 3^{dim}$ into the *dim*-dimension grid. An empty cell is of type *0*.

To choose the future type of a cell among all the possibilities, we consider a real random number A bounded by *0* (included) and *1* (excluded). The decision rules are the following:

$$\begin{aligned} &\text{if } A \in [0, p_0[\text{ then the considered cell is of type } 0 \\ &\text{if } A \in [p_0, p_0+p_1[\text{ then the considered cell is of type } 1 \\ &\dots \\ &\text{if } A \in [p_0+p_1+\dots+p_{nb-1}, p_0+p_1+\dots+p_{nb}[\text{ then the considered cell is of type } nb. \end{aligned} \qquad (3)$$

Figure 10 shows the results of these local probabilistic rules with a *60x60* grid at three different times and for two types of molecule.

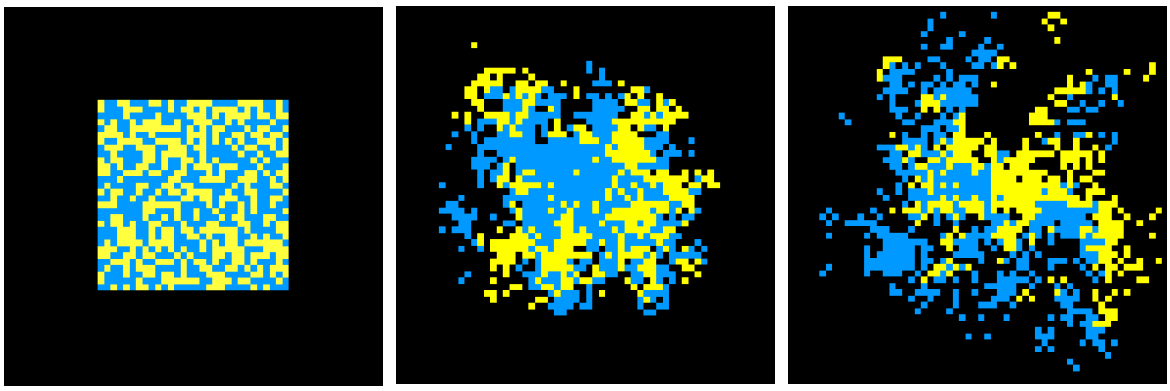


Figure 10: Diffusion of 2 types of molecule in a cellular automata system. States at successive times (t=0, 10 and 100) are shown (see plate 23).

To study hyperstructures containing many different molecules that can perform chemical reactions, we allow two different molecules to react to produce another type of molecule:

$$mol_i + mol_j \rightarrow mol_k \qquad (4)$$

To do this, we add the following simple rule (5) to rule (3):

if the *considered cell* is of type *i* (resp. *j*) and the type chosen thanks to rule 3 is *j* (resp. *i*) then the cell will have the type *k* (according to rule 4) (5)

Figure 11 shows the evolution of the system in a *60x60* grid at three different times, for two types of molecule and that react together according to rules (4) and (5) to produce a third type of molecule.

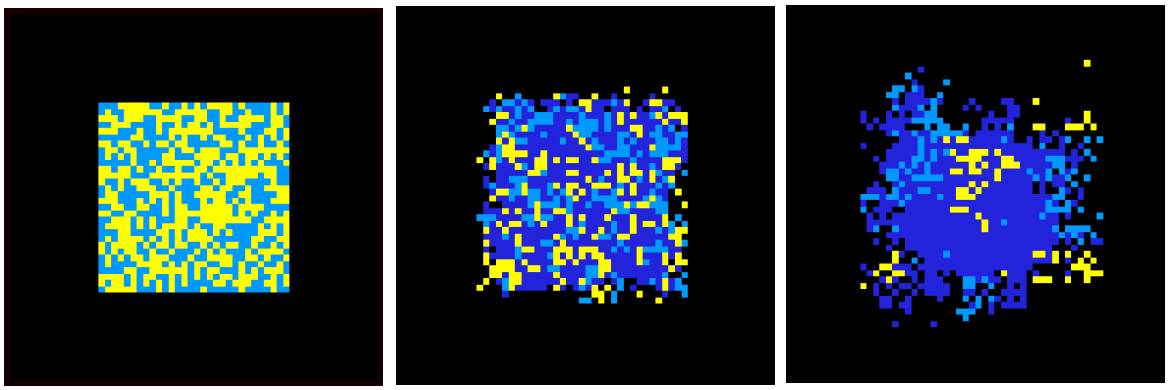


Figure 11: Diffusion of 2 substrates and a product in a cellular automata system. Molecules of type 1 and molecules of type 2 interact to produce type 3. At time t=0, there are only 2 types of molecules, type 1 (yellow) and type 2 (light blue). At time t=3, type 3 (dark blue) appears (see plate 24).

To observe the formation of and interaction between hyperstructures, we introduce the notion of affinity between molecules. In our example, molecules of type 1 are activated and can therefore bind one another. Molecules of type 2 and 3 have similar behaviours. Moreover, molecules of type 1 can react with molecules of type 2 to produce molecules of type 3.

- Molecules of type 1 come from the top of the cellular automata and bind together
- Molecules of type 2 come from the bottom of the cellular automata and bind together too
- Molecules of type 1 react with molecules of type 2 to produce molecules of type 3
- Molecules of type 3 bind together and with molecules of type 1 and type 2.

10 Experimental aspects

10.1 NanoSIMS

Visualising hyperstructures directly with conventional techniques has been difficult since it requires the co-localization of such disparate elements as proteins, mRNA, genes and lipids at the 50 nm scale. In secondary ion mass spectrometry, a section of biological material is subjected to a beam of ions that pulverizes it to release secondary ions that are filtered by mass spectrometry to allow an image to be obtained (Thellier *et al*., 1993). Recent developments in NanoSIMS technology are very promising since the new generation of machines provides

resolution at the scale required and allows detection of isotopically marked probes to proteins and nucleic acids. This opens up the exciting possibility of studying hyperstructures by imaging simultaneously both nucleic acids and up to 10 different proteins at a resolution intermediate between light and electron microscopy.

10.2 Optical waveguide lightmode spectroscopy (OWLS)

In the case of glycolysis, we lack details of the exact abundance of proteins such as phosphoglucose isomerase**,** fructose -1,6-P2 aldolase, triose-P isomerase, glyceraldehyde 3-phosphate dehydrogenase A complex, and phosphoglycerate kinase. Although we can obtain these *via* radioactive labeling and 2-D gel electrophoresis, there are attractive, recent techniques such as those based on isotope-coded affinity tags (Gygi *et al*., 1999). More seriously, we lack details of the constants of affinity of the PTS and glycolytic enzymes. These could be obtained using OWLS in experiments with purified proteins and substrates (Ramsden, 1993). By introducing and removing the substrates, it may also prove possible in these experiments to estimate the period of time for which an enzyme remains active (i.e. has a higher affinity constant) once its substrate has gone (Ricard *et al*., 1998).

10.3 MALDI-MS and ES-MS

We are presently using sensitive techniques of mass spectrometry to explore the possibility that concomitant with overproduction of a membrane protein is a compensatory overproduction of the lipid for which it has an affinity. If this approach is successful, a semi-automated, general strategy might be developed in which bacteria are transformed with plasmids each containing a different peptide (from a random library); the idea is to obtain thousands of colonies, each containing lipids resulting from the overproduction of a particular peptide. Mass spectrometry and sequencing would then match lipids and peptides. The data would be used to try to derive consensus sequences to be used to interpret the genome and construct a 'lipidome'.

10.4 Atomic Force Microscopy (AFM) and the Langmuir-Blodgett technique

Langmuir-Blodgett monolayers of phospholipids, which assemble at the air-water interface, followed by transfer to a solid support and inspection with AFM, provide a powerful combination of techniques for studying FtsZ interaction with membranes and may constitute the beginnings of an *in vitro* division system (Alexandre *et al*., 2002). The characteristics of the lipids used along with the values of parameters obtained for factors that interact with FtsZ, such as calcium, GTP and other division proteins, might be used to try to construct an *in silico* model of the division process.

11. I-cell

Developing new concepts may prove essential to a full understanding of how a cell works. To test and develop such concepts, we advocate the construction of an *Integrated* or *Imaginary* cell – *I-cell* (Amar *et al*., 2002). An I-cell would undergo selection for growth and survival in a world of artificial chemistry (Demarty *et al*., 2003; Dittrich and Banzhaf, 1998). The unit volumes that constitute an I-cell would be inspected at each time step and, according to the molecule(s) found, the appropriate entry would be consulted in a table containing a large number of 'biological' functions (Norris and Le Sceller, 2001). These functions would determine the interactions of the molecule with its neighbours and also, via global functions, with distant molecules. The I-cell would be fed according to different regimes and, depending on the functions implemented, would grow and eventually divide; I-cells would be analysed after

selection over several generations. Combinatorial problems would be reduced if an activity-based vision of the cell were adopted in which only a subset of constituents would be consulted at each time step; this subset would correspond to constituents that play an active role in coherent cell states via a mechanism based in part on global functions and termed competitive coherence (Norris, 1998). An I-cell might, for example, offer a way to discover the importance of a particular organising process, for example, one based on water structure or tensegrity. An I-cell might even be used to see whether new laws of complexity emerge as the number of organising processes in the system increases.

Acknowledgements

We thank Genopole® and the Conseil Regional de l'Ile de France for support.

References

Alexandre, S., G. Cole, S. Coutard, C. Monnier, V. Norris, W. Margolin, X. Yu and J.-M. Valleton (2002). Interaction of FtsZ protein with a DPPE film. Colloids and Interfaces B 23: 391-395.

Amar, P., P. Ballet, G. Barlovatz-Meimon, A. Benecke, G. Bernot, Y. Bouligand, P. Bourguine, F. Delaplace, J.-M. Delosme, M. Demarty, I. Fishov, J. Fourmentin-Guilbert, J. Fralick, J.-L. Giavitto, B. Gleyse, C. Godin, R. Incitti, F. Képès, C. Lange, L. Le Sceller, C. Loutellier, O. Michel, F. Molina, C. Monnier, R. Natowicz, V. Norris, N. Orange, H. Pollard, D. Raine, C. Ripoll, J. Rouviere-Yaniv, M. Saier jnr., P. Soler, P. Tambourin, M. Thellier, P. Tracqui, D. Ussery, J.-P. Vannier, J.-C. Vincent, P. Wiggins and A. Zemirline (2002). Hyperstructures, genome analysis and I-cell. Acta Biotheoretica 50: 357-373.

Arechaga, I., B. Miroux, S. Karrasch, R. Huijbegts, B. de Kruijff, M.J. Runswick and J.E. Walker (2000). Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale overproduction of the b subunit of F_1F_0 ATP synthase. FEBS Letters 482: 215-219.

Balandina, A., L. Claret, R. Hengge-Aronis and J. Rouviere-Yaniv (2001). The *Escherichia coli* histone-like protein HU regulates *rpoS* translation. Molecular Microbiology 39: 1069-1079.

Binenbaum, Z., A.H. Parola, A. Zaritsky and I. Fishov (1999). Transcription- and translation-dependent changes in membrane dynamics in bacteria: testing the transertion model for domain formation. Molecular Microbiology 32: 1173-1182.

Blattner, F.R., G. Plunkett III, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau and Y. Shao (1997). The complete genome sequence of *Escherichia coli* K-12. Science 277: 1453-1462.

Bok, J. (1983). Un modèle d'auto-organisation : le principe de moindre difficulté. In L'auto-organisation. De la physique au politique. (Dumouchel, P. and Dupuy, J.-P., eds.). Paris. Seuil,
Bonnefoy, E. and J. Rouviere-Yaniv (1992). HU, the major histone-like protein of *E. coli*, modulates the binding of IHF to *oriC*. EMBO Journal 11: 4489-4496.

Bonnefoy, E., M. Takahashi and J. Rouviere-Yaniv (1994). DNA-binding parameters of the HU protein of *Escherichia coli* to cruciform DNA. Journal of Molecular Biology 242: 116-129.

Bouligand, Y. and V. Norris (2001). Chromosome separation and segregation in dinoflagellates and bacteria may depend on liquid crystalline states. Biochimie 83: 187-192.

Bray, D., M.D. Levin and C.L. Morton-Firth (1998). Receptor clustering as a cellular mechanism to control sensitivity. Nature 393: 85-88.

Buddelmeijer, N., M.E.G. Aarsman, A.H.J. Kolk, M. Vicente and N. Nanninga (1998). Localisation of cell division protein FtsQ by immunofluorescence microscopy in dividing and non-dividing cells of *Escherichia coli*. Journal of Bacteriology 180: 6107-6116.

Castuma, C.E., E. Crooke and A. Kornberg (1993). Fluid membranes with acidic domains activate DnaA, the initiator protein of replication in *Escherichia coli*. Journal of Biological Chemistry 268: 24665-24668.

Cho, H.C., S. Singh and G.W. Robinson (1997). Understanding all of water's anomalies with a non-local potential. Journal of Chemical Physics 107: 7979-7988.

Danchin, A. and A. Henaut (1997). The map of the cell is in the chromosome. Current Opinion in Genetics and Development 7: 852-854.

Demarty, M., B. Gleyse, D. Raine, C. Ripoll and V. Norris (2003). Modelling autocatalytic networks with artificial microbiology. Comptes Rendus de l'Academie des Sciences : in press.

Dittrich, P. and W. Banzhaf (1998). Self-evolution in a constructive binary string system. Artificial Life 4: 203-220.

Fishov, I. (1994). Do oscillations control the bacterial cell cycle? In What is controlling Life? (Gnaiger, E., Gellerich, F.N. and Wyss, M., eds.), pp. 221-225. Innsbruck, Austria. Innsbruck University Press,

Fishov, I. and C. Woldringh (1999). Visualization of membrane domains in *Escherichia coli*. Molecular Microbiology 32: 1166-1172.

Fralick, J.A. and K.G. Lark (1973). Evidence for the involvement of unsaturated fatty acids in the initiation of chromosome replication in *Escherichia coli*. Journal of Molecular Biology 80: 459-475.

Guzman, E.C., J. Caballero, L. and A. Jimenez-Sanchez (2002). Ribonucleoside diphosphate reductase is a component of the replication hyperstructure in *Escherichia coli*. Molecular Microbiology 43: 487-495.

Gygi, S.P., B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb and R. Aebersold (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotechnology 17: 994-999.

Kamashev, D. and J. Rouviere-Yaniv (2000). The histone-like protein HU binds specifically to DNA recombination and repair intermediates. EMBO Journal 19: 6527-6535.

Kauffman, S. (1996). At home in the Universe, the search for the laws of complexity., Penguin, London.

Képès, F. and A. Képès (1985). Postponement of cell division by nutritional shift-up in *Escherichia coli*. Journal of General Microbiology 131: 677-685.

Kiessling, J., S. Kruse, S.A. Rensing, K. Harter, E.L. Decker and R. Reski (2000). Visualization of a cytoskeleton-like FtsZ network in chloroplasts. Journal of Cell Biology 151.

Koppelman, C.-M., T. Den Blaauwen, M.C. Duursma, R.M.A. Heeren and N. Nanninga (2001). Escherichia coli minicell membranes are enriched in cardiolipin. Journal of Bacteriology 183: 6144-6147.

Ksenzenko, S.M. and W.S.A. Brusilow (1993). Protein-lipid interactions of the proteolipid c subunit of the *Escherichia coli* proton-translocating adenosinetriphosphatase. Archives of Biochemistry and Biophysics 305: 78-83.

Le Sceller, L., C. Ripoll, M. Demarty, A. Cabin-Flaman, T. Nyström, M. Saier Jnr. and V. Norris (2000). Modelling bacterial hyperstructures with cellular automata. Interjournal Paper 366: http://www.interjournal.org.

Letoffe, S., P. Delepelaire and C. Wandersman (1996). Protein secretion in gram-negative bacteria: assembly of the three components of ABC protein mediated exporters is ordered and promoted by substrate binding. EMBO Journal 15: 5804-5811.

Lewis, P.J., S.D. Thaker and J. Errington (2000). Compartmentalization of transcription and translation in *Bacillus subtilis*. EMBO Journal 19: 710-718.

Lloyd, P.G. and C.D. Hardin (1999). Role of microtubules in the regulation of metabolism in isolated cerebral microvessels. American Journal of Physiology 277: C1250-1262.

Lynch, A.S. and J.C. Wang (1993). Anchoring of DNA to the bacterial cytoplasmic membrane through co-transcriptional synthesis of polypeptides encoding membrane proteins or proteins for export : a mechanism of plasmid hypernegative supercoiling in mutants deficient in DNA topoisomerase I. Journal of Bacteriology 175: 1645-1655.

Mayer, F. and M. Hoppert (1996). Functional compartmentalization in bacteria and archaea. A hypothetical interface between cytoplasmic membrane and cytoplasm. Naturwissenschaften 83: 36-39.

Meinhardt, H. and P.A.J. de Boer (2001). Pattern formation in *Escherichia coli*: a model for the pole-to-pole oscillations of Min proteins and the localization of the division site. Proceedings of the National Academy of Science U.S.A. 98: 14202-14207.

Mileykovskaya, E. and W. Dowhan (2000). Visualization of phospholipid domains in *Escherichia coli* by using the cardiolipin-specific fluorescent dye 10-N-nonyl acridine orange. Journal of Bacteriology 182: 1172-1175.

Miller, C.G. (1996). Protein degradation and proteolytic modification. In Escherichia coli and Salmonella. (Neidhardt, F.C., ed.), Vol. I, pp. 938-954. Washington D.C. American Society for Microbiology,

Mitchell, C.G. (1996). Identification of a multienzyme complex of the tricarboxylic acid cycle enzymes containing citrate synthase isoenzymes from *Pseudomonas aeruginosa*. Biochemical Journal 313: 769-774.

Mowbray, J. and V. Moses (1976). The tentative identification in *Escherichia coli* of a multi-enzyme complex with glycolytic activity. European Journal of Biochemistry 66: 25-36.

Newman, G. and E. Crooke (2000). DnaA, the initiator of *Escherichia coli* chromosomal replication, is located at the cell membrane. Journal of Bacteriology 182: 2604-2610.

Norris, V. (1995). Hypothesis: transcriptional sensing and membrane domain formation initiate chromosome replication in *Escherichia coli*. Molecular Microbiology 15: 985-987.

Norris, V. (1998). Modelling *E. coli*: the concept of competitive coherence. Comptes Rendus de l'Academie des Sciences 321: 777-787.

Norris, V., M. Demarty, D. Raine, A. Cabin-Flaman and L. Le Sceller (2002a). Hypothesis: hyperstructures regulate initiation in *Escherichia coli* and other bacteria. Biochimie 84: 341-347.

Norris, V. and I. Fishov (2001). Hypothesis: membrane domains and hyperstructures control bacterial division. Biochimie 83: 91-98.

Norris, V., J. Fralick and A. Danchin (2000). A SeqA hyperstructure and its interactions direct the replication and sequestration of DNA. Molecular Microbiology 37: 696-702.

Norris, V., P. Gascuel, J. Guespin-Michel, C. Ripoll and M.H. Saier Jr. (1999). Metabolite-induced metabolons: the activation of transporter-enzyme complexes by substrate binding. Molecular Microbiology 31: 1592-1595.

Norris, V. and L. Le Sceller. (2001). *International Conference of Systemics, Cybernetics and Informatics, Orlando, Florida, USA*.

Norris, V. and M.S. Madsen (1995). Autocatalytic gene expression occurs *via* transertion and membrane domain formation and underlies differentiation in bacteria: a model. Journal of Molecular Biology 253: 739-748.

Norris, V., G. Misevic, J.M. Delosme and A. Oshima (2002b). Hypothesis: a phospholipid translocase couples lateral and transverse bilayer asymmetries in dividing bacteria. Journal of Molecular Biology 318: 455-462.

Ohsumi, K., M. Yamazoe and S. Hiraga (2001). Different localization of SeqA-bound nascent DNA clusters and MukF-MukE-MukB complex in *Escherichia coli* cells. Molecular Microbiology 40: 835-845.

Onogi, T., H. Niki, M. Yamazoe and S. Hiraga (1999). The assembly and migration of SeqA-Gfp fusion in living cells of *Escherichia coli*. Molecular Microbiology 31: 1775-1782.

Ouzounis, C., G. Casari, A. Valencia and C. Sander (1996). Novelties from the complete genome of *Mycoplasma genitalium*. Molecular Microbiology 20: 898-900.

Pedersen, A.G., L.J. Jensen, S. Brunak, H.-H. Staerfeldt and D.W. Ussery (2000). A DNA structural atlas for *Escherichia coli*. Journal of Molecular Biology 299: 907-930.

Petty, H.R. and A.L. Kindzelskii (2001). Dissipative metabolic patterns respond during neutrophil transmembrane signaling. Proceedings of the National Academy of Science U.S.A. 98: 3145-3149.

Ramsden, J.J. (1993). Calcium-dependence of laminin binding to phospholipid membranes. Biopolymers 33: 475--477.

Revet, B., B. von Wilcken-Bergmann, H. Bessert, A. Barker and B. Müller-Hill (1999). Four dimers of λ repressor bound to two suitably spaced pairs of λ operators form octamers and DNA loops over large distances. Current Biology 9: 151-154.

Ricard, J., B. Gontero, L. Avilan and S. Lebreton (1998). Enzymes and the supramolecular organization of the living cell. Information transfer within supramolecular edifices and imprinting effects. Cellular and Molecular Life Sciences 54: 1231-1248.

Robinson, G.W., C.H. Cho and J. Urquidi (1999). Isobiestic points in liquid water: Further strong evidence for the two-state mixture model. Journal of Chemical Physics 111: 698-702.

Segre, D., D. Ben-Eli and D. Lancet (2000). Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. Proceedings of the National Academy of Science U.S.A. 97: 4112-4117.

Speck, C., C. Weigel and W. Messer (1999). ATP- and ADP-DnaA protein, a molecular switch in gene regulation. EMBO Journal 18: 6169-6176.

Stock, J. and M. Levit (2000). Signal transduction: hair brains in bacterial chemotaxis. Current Biology 10.: R11-4.

Thellier, M., C. Ripoll, C. Quintana, F. Sommer, P. Chevallier and J. Dainty (1993). Physical methods to locate metal atoms in biological systems. Methods in Enzymology 227: 535-586.

Torshin, I. (1999). Activating oligomerization as intermediate level of signal transduction: analysis of protein-protein contacts and active sites in several glycolytic enzymes. Front. Biosci. 4: D557-570.

Ussery, D., T.S. Larsen, K.T. Wilkes, C. Friis, P. Worning, A. Krogh and S. Brunak (2001). Genome organisation and chromatin structure in *Escherichia coli*. Biochimie 83: 201-212.

Velot, C., M.B. Mixon, M. Teige and P.A. Srere (1997). Model of a quinary structure between Krebs TCA cycle enzymes: a model for the metabolon. Biochemistry 36: 14271-14276.

Vichniac, G.Y. (1984). Simulating physics with cellular automata. Physica D 10: 96-116.

Weigel, C., A. Schmidt, H. Seitz, D. Tungler, M. Welzeck and W. Messer (1999). The N-terminus promotes oligomerization of the *Escherichia coli* initiator protein DnaA. Molecular Microbiology 34: 53-66.

Wiggins, P.M. (1990). Role of water in some biological processes. Microbiological Reviews 54: 432-449.

Woldringh, C.L. and N. Nanninga. (1985). Structure of the nucleoid and cytoplasm in the intact cell. In Molecular Cytology of *Escherichia coli*. (Nanninga, N., ed.), pp. 161-197. London. Academic Press,

Wolf, S.G., D. Frenkiel, T. Arad, S.E. Finkel, R. Kolter and A. Minsky (1999). DNA protection by stress-induced biocrystallization. Nature 400: 83-85.

Xia, W. and W. Dowhan (1995). *In vivo* evidence of the involvement of anionic phospholipids in initiation of DNA replication in *Escherichia coli*. Proceedings of the National Academy of Science U.S.A. 92: 783-787.

Emergent features of cell structural dynamics: a review of models based on tensegrity and nonlinear oscillations

Philippe Tracqui[1], Emmanuel Promayon[1], Patrick Amar[3,4], Nicolas Huc[1], Vic Norris[2] and Jean-Louis Martiel[1]

[1]Laboratoire TIMC-IMAG, CNRS, Instut A. Bonniot, F-38706 La Tronche Cedex
[2]Laboratoire des Processus intégratifs cellulaires, Faculté des Sciences et Techniques, Université de Rouen, F-76821 Mont-Saint-Aignan Cedex
[3]Laboratoire de Recherches en Informatique, Université Paris Sud & CNRS UMR 8623, 15 avenue George Clémenceau, F-91405 Orsay Cedex
[4]La.M.I. Université d'Évry Val d'Essonne & CNRS UMR 8042, Tour Évry 2, 523 Place des terrasses de l'agora, F-91000 Évry

1. Introduction

The key role of the coupling between mechanical forces and tissue growth and remodelling was suggested nearly twenty years ago (Trinkhaus, 1984), especially in the field of bone formation and remodelling. However, it is only recently that a large body of experiments highlighted the effects of physical forces such as tension, compression, gravity or shear stress at the cell level (Edwards et al., 1999; Huang and Ingber, 1999; Kaspar et al., 2000; Tabony et al., 2002). Indeed, direct application of mechanical stresses to cultured cells can induce or modify cell differentiation, growth and migration as well as gene expression (Schmidt et al., 1998; Fujisawa et al., 1999; Wang et al., 2000; Meyer et al., 2000; Chen et al., 2001). Despite this progress, we do not fully understand how individual cells perceive and orchestrate mechanical signals to respond either individually at the level of the cell itself or collectively, via cooperative processes, at the level of the tissue.

The challenge that confronts cells is how to integrate the information from many kinds of signals so as to permit a single appropriate response. Among these different cell signalling mechanisms, mechanotransduction, i.e. the conversion of a mechanical signal into a biological or biochemical response, enjoys a special status (Wang et al., 1993). This is because it relies on the structure of the cell, which is a global property, as opposed to its biochemistry, which is a local property (and which can be restricted for example, to the cytoplasmic membrane or an intracellular organelle). Mechanotransduction provides the only way to understand how many, simultaneous, extracellular, mechanical inputs (adhesion to extracellular matrix (ECM) proteins, junctions with other cells, …) combined with heterogeneous mechanical properties (local softening or hardening of the cytoskeleton) are integrated with other stimuli to allow a specific physiological or pathological cellular response (Lelièvre et al., 1996; Janmey, 1998). The importance of mechanotransduction makes it crucial to develop theoretical cytomechanical models of the living cell.

We explore here some of these cytomechanical models and pay special attention to models of cell tensegrity, as proposed by Don Ingber (1993; 1997), although we also examine alternative models considering either the cell are a mechanical (viscoelastic) continuum, or more simply the spatio-temporal organization of structural elements of the cell cytoskeleton (CSK) like filamentous actin. All models should help to understand how the cell architectural properties are dynamically regulated and provided a support to the transduction of intracellular signals and mechanical forces that lead to a wide range of integrated cellular responses. These modelling approaches are based on the conviction that it is essential to model how cells dynamically stabilise and self-organise their structure and shape if we are to

understand how they sense their physical microenvironment and respond to mechanical signals through bi-directional signalling pathways that connect the plasma membrane to the nucleus. In this framework, one expect a strong connection between the CSK remodelling and the control of gene expression.
With these considerations in mind, this paper reviews a representative range of discrete and continuous cytomechanical models of cells. We first explore the properties of purely structural models. We then present discrete and continuous models which investigate the self-organisation of the cell structure induced by nonlinear biochemical and biophysical processes, possibly coupled with changes of cell mechanical properties.

2. Discrete mechanical cell models

2.1. Cells as tensegrity structures

Since the eighties, D. E. Ingber and coworkers have promoted and architectural view of the living cell and of larger biological entities based on the concept of tensional integrity or tensegrity, according to which the structure of cells depends on tensile forces for its integrity (Ingber, 1993 ; 1997). As explained by Ingber, this proposition draws on the architectural principles developed for the construction of buildings by Buckminster Fuller and K. Snelson; these principles underlie the organisation of structures as diverse as geodesic domes and viral capsids.

It is worth emphasizing in this introductory section that the holistic approach proposed by Ingber has deep conceptual implications. First, it again draws attention to the need for unifying principles to interpret the plethora of biological data collected successfully by a reductionism that cannot explain them on its own. Second, it strengthened the view of biological systems as non-linear systems with functions that are more than the sum of its parts.

What is the definition of a tensegrity system? According to Ingber and Jamieson (1985) "*A tensegrity system is defined as an architectural construction that is comprised of an array of compression-resistant struts that do not physically touch one another but are interconnected by a continuous series of tension elements*". At the cell level, the tensegrity framework proposed by D. E. Ingber is based on several assumptions. The first assumption is that cell shape is stabilized by an internal, mechanically active, structure, the CSK. This assumption excludes both theoretical models treating the cell as a viscous fluid surrounded by a membrane as well as mechanical models treating the CSK as a passive structure. Indeed, it is well-documented that living cells (and even non-muscle cells like endothelial cells or fibroblasts) can generate active tension through an actomyosin filament sliding mechanism similar to the one used in the contraction of smooth muscles. This is easily visualised, either at the cell level by the wrinkling of malleable substrata by cells (Dembo et al., 1996 ; Dembo and Wang, 1999), or at the population level via measurements of isometric tension or compaction rate of viscoelastic biogels (Kolodney and Wysolmersky, 1992; Eastwood et al., 1996). This internal contraction capability creates within the cell a prestress upon which external mechanical loads are superimposed. Tensegrity cell models have thus been developed by considering elastic and contractile actin microfilaments as tension elements and microtubules as compression-resistant elements (Wang et al., 2001).

The second assumption is that forces are not transmitted continuously across the cell but rather that transfer of mechanical loads and stresses take place at points where the cell is anchored to the extracellular matrix (ECM) and to neighbouring cells. Indeed, the cell anchors itself to the ECM by physically binding CSK elements to specific focal adhesion complexes (FACs) that cluster within localized adhesion sites. Recent works have shown a strong

correlation (a potential regulation) between the surface of such contacts and the amplitude of the transmitted forces (Balaban et al., 2001; Tan et al., 2003). The FACs include not only transmembrane proteins, mostly integrins, but also a scaffold of actin-binding proteins (talin, vinculin, α-actinin, paxillin, …) which form a molecular bridge between the CSK and the intracellular part of the integrin. Similarly, specific cell-cell adhesion molecules (CAMs), like cadherins, selectins, cathenins, …) ensure the transmembrane coupling to the CSK of neighbouring cells at localised junctional sites (adherens junctions, desmosomes).

Cellular architecture seen in this way defines a mechanical network that provides a physical support to biochemical signal transduction pathways and that allows mechanical signals to be propagated from mechanoreceptors on the surface (in the form of cell adhesion molecules and transmembrane receptors) to targets deep within the cell.

2.2 Qualitative behaviour of cell tensegrity models

Some basic, qualitative properties of cell tensegrity models have been illustrated by Ingber and co-workers by constructing a physical structure made of multiple wood dowels interconnected with a series of elastic springs (Ingber, 1997). This "cellular toy" has interesting properties: if it is not subjected to an external force and if not attached to a rigid surface, it has a round shape because of the internal tension. However, it can spread out and flatten along a rigid surface if an external force is applied vertically. Moreover, if the toy is attached to a flexible surface while in its flattened configuration, and if the external force is then removed, the toy contracts spontaneously and returns to the round shape characteristic of its rest state, again because of the presence of internal tension. This relaxation is accompanied by a progressive wrinkling of the flexible substratum which closely mimics the formation of wrinkles observed when living cells are cultured on malleable substrates like silicon rubber.

Simulations of tensegrity via the construction of virtual cells may eventually prove indispensable to understanding how the cell interprets its genome. In this context, an additional and particularly interesting feature of models of cellular tensegrity is their ability to fit naturally into a structural hierarchy. One can then include within such virtual cells smaller intracellular structures with their own mechanical properties. Of course, the cell nucleus is the first candidate. The nucleus has its own internal structure, the nuclear matrix, and a body of experimental findings are consistent with the nucleus being mechanically coupled to the rest of the cell. For example, when a cell spreads on a rigid substratum, its nucleus extends in parallel, even if some delay can be observed in migrating cells. The tensegrity model explains how the nuclei of living cells can respond directly to mechanical stimuli that are applied to specific surface receptors as those involved in cell adhesion. It thus provides a basis to understand how extracellular mechanical stimuli can modify gene expression through mechanical deformations of the nucleus (Maniotis et al., 1997).

2.3. Theoretical properties of cell tensegrity models.

Different theoretical studies have been undertaken to analyse the mechanical properties of cell tensegrity models. A minimal tensegrity structure composed of 6 compressive elements (bars) and 24 extensible elements (cables) with frictionless joints has been analysed by Stamenovic et al. (1996). This study was extended by Wendling et al. (1999) who investigated whether such a pre-stressed geometric structure could account for the stiffening response observed in living cells. They showed that, under large deformations, the tensegrity structure exhibits a non-constant stiffening response which depends on the loading conditions (extension, compression or shear). They further demonstrated that, although the Young's elastic modulus

of each constitutive element stays constant, the apparent elastic modulus of the overall structure, the apparent elastic modulus of the overall structure does not stay constant.
They analysed the deformation of the 30-element cell tensegrity model governed by the constitutive equation:

$$\{\mathbf{F}\} = [\mathbf{K}]\{\mathbf{u}\}$$

which relates the vector of external forces {**F**} to the vector {**u**} defined by the displacement (distance between the deformed and the reference state) of each of the 12 nodes through the rigidity matrix [**K**] of the structure (cables and bars are assumed to be linearly elastic with Young's modulus E_c and E_b respectively, with $E_b >> E_c$) . The external forces are applied at three upper nodes of the structure, while the contact of the structure with the inferior plane occurs through three fixed nodes : this latter condition is chosen to simulate a weak cell adhesion of a round cell onto a rigid substratum.

At the unloaded (reference) state, mechanical equilibrium results from the balance between pre-stretching stress in the cables and in the bars. For the different loading conditions, the tensegrity model response is analysed by considering the apparent elasticity modulus derived from the stress-strain relation ship. With reference to continuous material, apparent elasticity modulus E_s and shear modulus G_s of the tensegrity structure have been defined as the derivative of the polynomial functions fitting the apparent normal and shear stresses to the apparent normal and shear strains respectively.

The numerical simulations of the tensegrity model response to extension and compression exhibit a non-linear mechanical response, characterised by linear relationship between the apparent elasticity modulus and the apparent strain. The modulus increases with strain during extension and is thus associated with a strain-hardening behaviour of the structure. On the contrary, a strain-softening occurs under compression, with decreasing values of the elasticity modulus as strain increases. This behaviour is qualitatively preserved when the length of the elastic elements or their pre-stress is changed by several orders of magnitude. Application of shear stresses leads to a more complex response, with initial stress-softening followed by stress-hardening.

This study illustrates how the response of the CSK of the living cell to applied stress might be a property of an integrated system and not a characteristic of individual components: even if each component of the tensegrity structure exhibits independently a linear elastic response with constant elastic modulus, the integrated response is non-linear with apparent elastic modulus that is strain-dependent. Such behaviour is related to the re-orientation of stresses which, even applied locally, are spatially distributed through the myriads of interconnected filaments composing the CSK.

In a recent work, Canadas et al. (Canadas et al., 2002) used the same tensegrity model to analyse the viscoelastic response of the CSK. Their tensegrity structure still includes 6 bars, but now the 24 cables are viscoelastic elements modelled as Voigt bodies (elastic element in parallel with viscous dashpot). This time dependence of cables rheology introduces into the above constitutive equation an additional global-damping matrix [C] which, assuming small displacements, modelled the proportionality of external forces to the rate of nodal displacements {u'}.

$$\{\mathbf{F}\} = [\mathbf{K}]\{\mathbf{u}\} + [\mathbf{C}]\{\mathbf{u'}\}$$

Some of the results obtained with the viscoelastic tensegrity structure are qualitatively similar, with an increase of the normalized apparent elasticity modulus E* with increasing global strain ε_i (strain hardening). On the other hand, the normalized viscosity modulus varies slowly with ε_i and remains almost independent of changes in T*, the normalized basal internal tension of the tensegrity structure. This is not the case for E*, which increases approximately with the square root of T*. Let us also mentioned the recent work of Coughlin and

Stamenovic (Coughlin and Stamenovic, 2003) which explore the simulated mechanical response of a tensegrity cell model probed by current cytomechanical methods such as twisting magnetocytometry (Laurent et al., 2002).

The so simulated stiffening response of cells is in agreement with biological experiments (Wang et al., 2002; Stamenovic et al., 2002, 2003; Laurent et al., 2002, 2003). However, extended validation is very difficult since the CSK is not a fixed scaffold, but a dynamic network that can be modulated by internal and external clues. Some of these dynamical aspects will be addressed in section 3.

2.4. Simulation of virtual cell models: the example of physically based computational models

Despite the virtues of the tensegrity model, it should be noted that it cannot explains certain dynamic features of the CSK such as its remodelling. We briefly reported here a recently developed computational approach that provides an alternative way for simulating the mechanical response of virtual objects modelling living cells (Promayon et al., 2003) and that could be more extensible. This approach is inspired by the physically based computational framework proposed for simulating the respiration movements of the human trunk (Promayon et al., 1996; 1997). Based on algorithms operating within an object oriented programming language, this approach is able to take into consideration dynamic changes of objects properties and shapes. In this modelling framework, cells are considered as three-dimensional elastic bodies submitted to internal cohesive forces as in the tensegrity approach. In addition, external attractive forces (gravity, chemo-attraction …) are also considered as possible control factors of the virtual cell dynamical features.

The virtual cell we have constructed within this framework is defined as a 3D incompressible object. From a computational point of view, this virtual cell is considered as an entity with its own properties (elasticity, contractility …) and history (interactions with other cells or ECM …). To simplify the calculations, a cell is defined by a 3D closed surface represented by a triangular mesh and its associated contour nodes.

Dynamical cell shape changes occur as a response to various forces (gravity, locally applied mechanical loads …) applied to each node of the mesh. The dynamic of the local cell response is then determined by the mass initially assigned to each node. Since mass-spring networks are known to be rather unstable systems, cell elasticity properties have been modelled by defining a local shape memory (Promayon et al., 1996). This means that the elastic property of a cell object is simply its ability to recover its original shape once deformed. This property is modelled by defining a local shape coordinate system in which each node of the structure is defined relatively to its neighbours by three parameters.

Some applications of this physically based computational approach are given below. First, as with tensegrity models, the mechanical response of specific cell architectures can be analysed. For example, one can simulate the effect of different intracellular organisation of the CSK which can mimic specific orientation of cell stress fibres. Figure 1 shows the simulated influence of transverse links within an elastic discrete envelope when the cell is submitted to uniaxial compression. The cell0 is defined as a strict elastic discrete envelop, with no internal links. In cell1, we considered a reinforcement of the cell architecture with horizontal elastic cross-links modelling CSK fibres. Finally, cell2's architecture includes internal diagonal elastic links connecting the apical and basal physical cell surfaces.

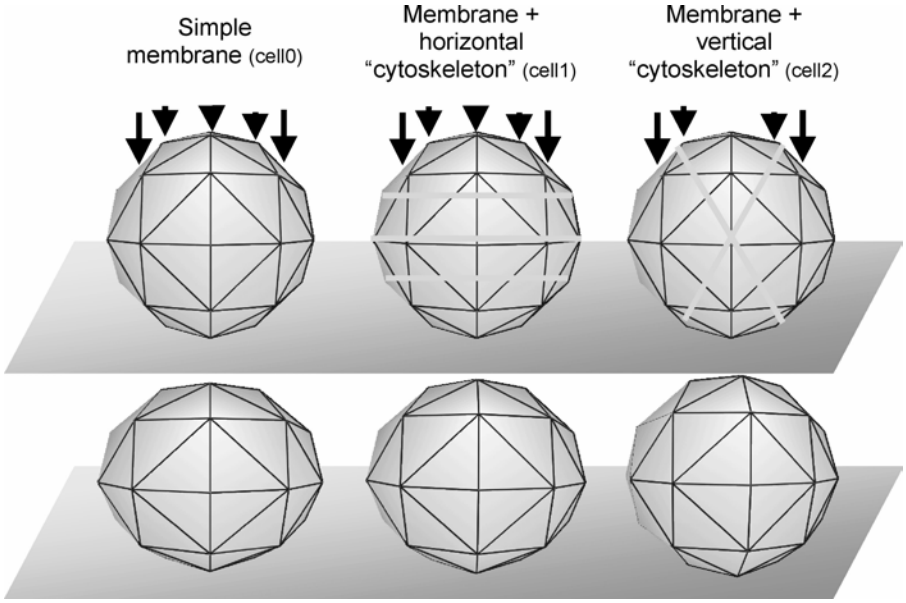


Figure 1: Influence of the "cytoskeleton" of the virtual on the mechanical response to vertical load. Three cell types are considered: no "cytoskeleton" (cell0), horizontal links (cell1), diagonal links (cell2). The first raw presents the initial 3D shape of each cell prototype. For a given fixed value k_{elas} of the elasticity modulus of each cell object, the second raw in the figure indicates the equilibrium state which is reached when a vertical loading force *Fs* is applied on the 5 (cell0 and cell1) or 4 (cell2) nodes marked with arrows.

To make a closer comparison with real cells, we consider a virtual cell devoid of internal elastic links. This provides a model of living cells like human erythrocytes, where the cell membrane is entirely responsible for the elastic deformation of the cell, the inner cytoplasm being only viscous. The relevance of this modelling approach with regard to real experiments is illustrated in figure 2 where optical tweezers experiments of Henon et al. (1999) to deform nearly spherical erythrocytes have been simulated. In the experiments, a force F is exerted on two silica microbeads which are stuck to the erythrocyte membrane in diametrical position. By slowly incrementing the distance between the two trapped beads, an increasing stress is applied to the cell membrane. To simulate this experiment, a force F_S has been locally exerted on two opposite nodes of the physical cell membrane, pulling them apart. The simulation parameters are the elasticity of the cell, i.e. k_{elas} and the modulus of F_S. Figure 2 shows that k_{elas} could be chosen in order to model the linear part of the erythrocyte response, but also the nonlinear stiffening of the cell for forces up to 15pN.

The microplates experiments of Thoumine et al. (1997 ; 1999 ; 2000) can also be simulated in a similar fashion (Sauvaget, 2001) (Fig. 3). Experimentally, the adhesion of a fibroblast is realised between two glass microplates, one of them being slightly flexible. The mechanical of the cell response to stretching is measured. To qualitatively simulate these experiments, a cell object was defined as a two-region object: virtual plasma and nucleus membranes, both of them being represented by triangulated surfaces.

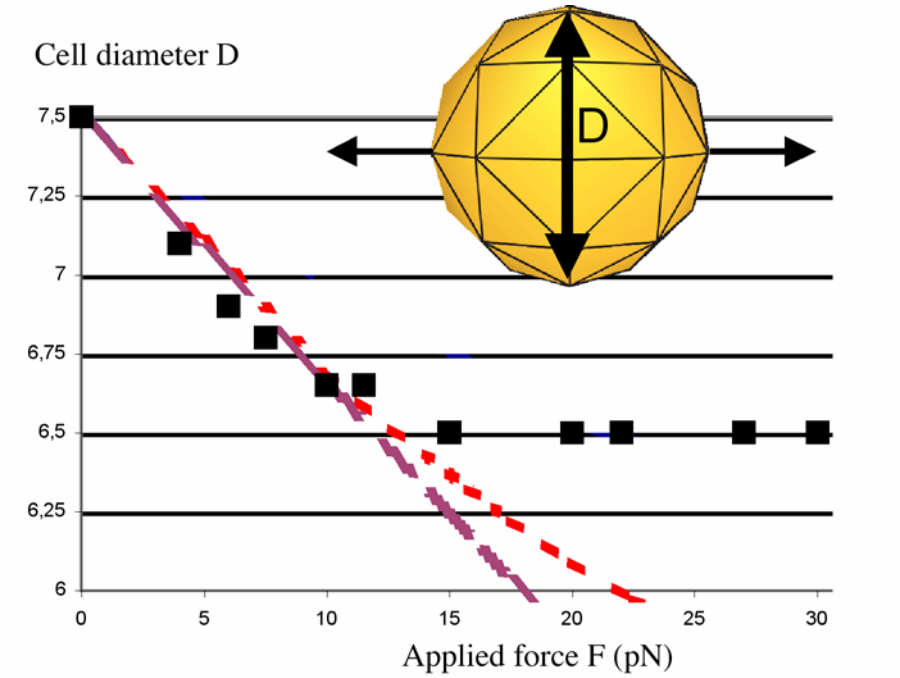
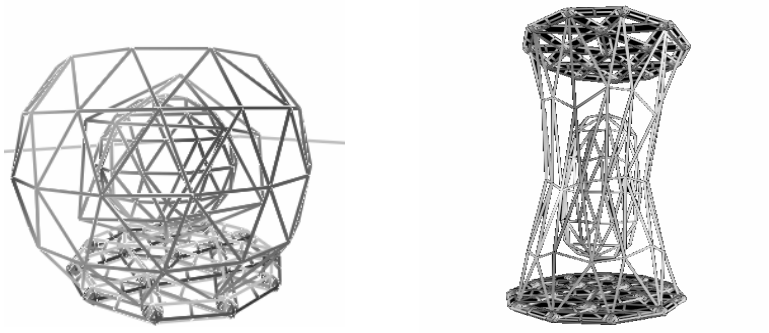


Figure 2: Simulated virtual spherical red blood cell (RBC) suspended in an hypotonic solution. Optical tweezers double trap is simulated by exerting locally a force F_S on two opposite nodes of the cell object contour (upper insert). The variation with load of the cell object diameter D(F_S) (in µm) in a plane perpendicular to the loading direction is simulated and compared to experimental data published by Henon et al. (1999) With appropriate scaling of the force, one can adjusted the parameter k_{elas} such that the experimental mechanical response of RBC can be nicely fitted in the linear elastic regime. Increasing the elasticity modulus k_{elas} induces a stiffer response which qualitatively reproduces the departure from the linear regime at larger traction forces.

Each node of the membrane is elastically linked to its neighbours as well as to the corresponding node in the other membrane. Microplates are modelled by two circular rigid objects attached to the apical and basal part of the external membrane, one of them being translated vertically. The corresponding shape of the virtual cell and nucleus at equilibrium is shown on the right of figure 3.

Figure 3: Simulated stretching of a virtual cell and its "nucleus" which mimics microplates experiments. Left: rest shape of the two embedded structures. Right: simulated cell and associated nucleus deformations.



For large amplitude deformations (hyperelastic behaviour), a non-linear mechanical response of the virtual cell could be observed. This property indicates that, as observed with the cellular model of tensegrity of Ingber and col , the global mechanical behaviour of this virtual cell is not the sum of its individual component responses. Further developments are needed to analyse the theoretical properties of such a virtual cell, especially with regard to mechanical properties exhibited by continuous finite elements models. However, the simulation of the microplates experiments reported here illustrates the capability of this approach to deal with multi-scale dynamical phenomena. For example, it is also possible to simulate cell population behaviours such as tissues (figure 4) or cell interactions during cell migration (Promayon et al., 2003)

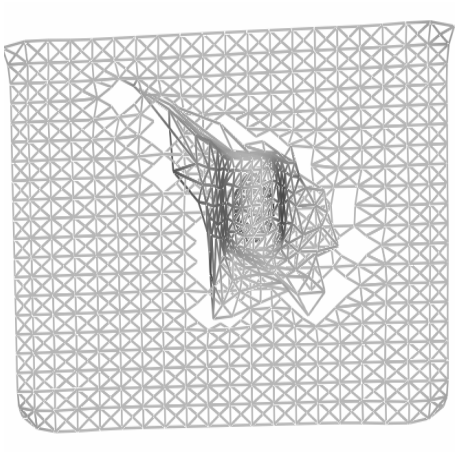


Figure 4: Simulated contraction of an elastic substratum by a virtual adherent fibroblast linked to this extracellular film by localised "focal contacts".

3. Oscillating cellular deformations and virtual cell models

The existence of a mechanical continuum within the cell means that oscillatory deformations could occur and affect gene expression. We therefore review some of the theoretical models proposed for analysing and modelling oscillations in cell shape. Each of them highlights a particular biophysical process as the central mechanism responsible for changes in cell shape. Nevertheless, the common theme is that all the interactions considered are integrated into a single response, namely an oscillation of cellular protrusions.

It has been known for a long time that living cells change their shape by extruding and remodelling different types of membrane protrusions. Filopodia are finger-like protrusions of the plasma membrane while lamellipodia are sheet-like protrusions associated with the filamentous actin (F-actin) network. It was once widely held that the cell membrane fluctuates without any particular direction in space and without any particular coherence in time. However, recent progress in cell imagery techniques and cell image analysis has revealed the organised dynamics of cell protrusions (Germain et al., 1999). For example, Killich et al. (1994) have reported the existence of different patterns of morphological changes in the amoebae *Dictyostelium discoideum*. Ehrengruber et al. (1995) show that neutrophils undergo periodic cytoskeletal rearrangements that lead to cycles of shape change with periods of 8-10s that are associated with sinusoidal oscillations of F-actin. Different hypotheses have been proposed to explain the formation of cellular protrusions. These hypotheses have considerable implications for our understanding of intracellular signalling and in generating testable predictions.

3.1. Cortical F-actin solation/gelation models.

In the early eighties, Oster and Perelson (1985) proposed a model of lamellipodial motion based on the physical chemistry of actomyosin gels. The model consists of a sheet of cytogel attached to the substratum by elastic tethers. The rhythmic activity of extending and retracting lamellipods is assumed to be driven by alternating phases of solation and gelation of the cortical actomyosin gel respectively. This phase transition is controlled by intracellular levels of calcium which is stored in intracellular compartments and which is released into the cytoplasm; this release is under the control of mechanisms that operate in a complex, non-linear (autocatalytic) fashion known as the calcium-induced calcium release mechanism (CICR). In Oster and Perelson's model, raising levels of free intracellular calcium concentration activates solation factors which disrupt the F-actin gel network. Factors such as gelsolin or severin can either break the actin chains themselves or break the cross-links between the chains or induce a depolymerisation of the chains. Such breakdown results in the gel swelling up to a point where the swelling pressure is balanced by elastic resistance of the network.

One important aspect of this non-linear process is that calcium also triggers the actomyosin contractile machinery (see also section 3.7). There is thus a defined range of calcium concentration where contraction occurs. In addition to the CICR mechanism quoted above, it is clear, regarding non-linear systems theory, that the cell has all the physical and biochemical ingredients needed to induce spontaneous self-sustained oscillations above some critical threshold. The trigger is assumed here to be an initial leak of calcium into the cell at the leading edge of the cell membrane. Leaking can be induced by the bindings of extracellular factors to membrane receptors or by a mechanical stimulus exerted at adhesion site and possibly accompanied by a modification of ion channels.

3.2. Intercalation of actin monomers: the Brownian ratchet mechanism

Peskin et al. (1993) formulated a theory to account for the force generated by the polymerisation process itself when the filaments are rigid. They proposed that the addition, below the cell membrane, of G-actin monomers at the end of F-actin growing filaments could exploit the Brownian motion of any diffusing object in the front of the filament. Thus, random fluctuations of the plasma membrane would create a free sub-membrane space where this intercalation could take place (Abraham et al., 1999 ; Borisy and Svitkina, 2000). This ratchet mechanism could explain the formation of thin cell protrusions when F-actin filaments are perpendicular to the membrane surface, but it cannot satisfactorily explain the formation of lamellipodia.

Molginer and Oster (1996) reconsidered this Brownian ratchet mechanism this model by assuming that the protrusion formation is more likely driven by the thermal bending undulations of semi-stiff F-actin filaments (Molginer and Oster, 2003a). This "elastic Brownian ratchet" mechanism has been recently reviewed (Molginer and Oster, 2003b) in order to account for the fact that actin filaments attach to the bacterium surface during actin-based propulsion of the bacterial pathogen *Listeria* (such bacterial cells being considered as simplified model systems for identifying essential factors of eukaryotic cell motility, see section 3.3 below). In an alternative work also considering the elastic polymerisation ratchet mechanism as the mechanical basis of cell protrusion, Molginer and Edelstein-Keshet (Molginer and Edelstein-Keshet, 2002) developped a reaction-diffusion-advection model which capture most of the biochemical (nucleation, growth, capping and depolymerisation of actin filaments) and geometrical aspects of actin dynamics in a cell protrusion. The prediction

of their model is that an optimal density of filaments barbed (fast-growing) ends is needed to insure a maximal protrusion velocity of the cell membrane. From available data, these authors derived an estimate protrusion velocity in the order of 0.4 μm/sec for a membrane resistance of 50 pN/μm, falling to 0.25 μm/sec when the membrane resistance is two times higher. They further speculate that the cell regulates its motility as well as its direction of motion by rapidly increasing the local density of uncapped barbed ends (through Arp2/3 activation, filaments uncapping, ...) and decreasing locally cell membrane resistance (disruption of cell cortex links to the membrane). However, a number of simplifications are likely to limit the range of validity of the results: one of them is the oscillatory feature of cell protrusions, which will be addressed in section 3.6.

3.3. Alternative to the ratchet model.

The pathogen bacterium *Listeria monocytogenes* uses actin polymerisation to propel itself through the cytoplasm and the membrane of infected cells (Theriot et al, 1992; Frischknecht and Way, 2001). Experimental data show that the cell motility results from cooperation between the bacterium and the host cytoplasm proteins. The bacterium surface protein ActA controls the activity of the complex Arp2/3 that initiates actin polymerisation (Welch et al., 1997). Actin dynamics is also controlled by an actin depolymerising factor (ADF/cofilin) and capping proteins, which are in the cytoplasm host. The last two factors maintain a high level of actin monomers in the cytoplasm to achieve filament growth at the bacterium surface. In vitro studies proved that movement was possible with a limited number of proteins, including Arp2/3, ADF/cofilin and a capping protein (Loisel et al., 1999). Biophysical investigations demonstrate that the bacterium and its actin-tail are tightly bound, which rules out the ratchet model approach for this system (Gerbal et al., 2000). In addition, the same group measured the actin-tail Young modulus at a value of 103-104 Pa, a value 10 times larger than the cytoplasm rigidity. Using the framework of elasticity theory, Gerbal et al. (2000) proposed that the mechanical stresses generated at the *Listeria* cell surface are relieved at the back of the bacterium pushing the cell forwards. Their model accounted satisfactorily for the cell speed (about 0.1 $\mu m.s^{-1}$) and was extended to explain the hopping motion observed in a *Listeria* mutant

3.4. Actin polymerisation, F-actin nucleation and reaction-diffusion models

3.4.1. Some experimental data

Actin dynamics plays a major role not only in cell movement (Condeelis, 1993) but also in cell adhesion or neuron plasticity (Colicos et al., 2001; Star et al., 2002). Characterisation of actin filaments growth proved the importance of the polymerisation/depolymerisation balance at the filament ends and the role of proteins in inducing actin polymerisation (e.g. Arp2/3), severing actin filaments (e.g. gelsolin, ADF/cofilin) or protecting the filaments ends by capping proteins (Pollard et al., 2000). Actin monomers associate to form filaments with a polarity (barbed *vs*. pointed ends). At the barbed end, subunits associate rapidly, with a low equilibrium actin monomer concentration ($C_{eq,B} = 0.08\mu M$). In contrast, the dynamics is much more slower at the pointed end but with a larger equilibrium monomer concentration ($C_{eq,P} = 0.5\mu M$, Carlier et al., 1997).

At steady-state, the actin monomer concentration is:

$$C_{eq} = \frac{k_B^+ C_{eq,B} + k_P^+ C_{eq,P}}{k_B^+ + k_P^+}$$

where k_B^+ and k_P^+ are the association rate of actin monomers to the barbed (B) or pointed (P) ends. At steady-state, the growth at the barbed end is exactly compensated by the disassembly at the pointed end, a dynamical state called *treadmilling*. However, the predicted treadmilling steady-state flux (i.e. $k_B^+(C_{eq} - C_{eq,B}) = -k_P^+(C_{eq} - C_{eq,P})$) is too slow ($0.2s^{-1}$) to account for the rapid turnover observed *in vivo*. This suggests that other cellular factors affect actin dynamics, including interactions with intracellular proteins (Carlier et al, 1997a, Pollard et al., 2000), intracellular signalling (Machesky et al., 1999; Mullins, 2000) and movement generation (Borisy et al., 2000).

3.4.2. Models for actin network formation

In view of the complexity of the dynamics of actin networks, models were first formulated to analyse the polymerisation and fragmentation of isolated actin filaments *in vitro* (Edelsetein-Keshet et al., 1998; Ermentrout et al., 1998). These models use the classical framework of kinetic differential equations without addressing the question of interactions between filaments and the formation of network. A further advance was made with the study of actin bundle formation using the kinetic approach (Edelstein-Keshet, 1998). This model was refined to explain the length distribution of actin filaments in a lamellipod (Edelstein-Keshet et al., 2001). Actin filament orientation was also studied using a Boltzmann-like equation (Geigant et al., 1998). However, these models are based on the kinetics of actin polymerisation or actin filament association in the absence of geometrical or mechanical constraints. A recent attempt to address the more complex question of the generation of actin networks was done by Maly and Borisy (2001) who developed a model for this generation as a process of self-organization. They were able to account for the preferential direction of the actin filament bundle observed *in vivo*. Finally, the analysis of actin gel formation on the surface of beads provided both experimental and theoretical insights into the regulation of actin networks in cells (Noireaux et al., 2000).

Models for spatio-temporal F-actin interactions *in vivo* were developed by LeGuyader and Hyver (1997), who analysed the oscillatory dynamics of the cortical actomyosin ring of human lymphoblasts by interpreting it in terms of a reaction-diffusion process. They proposed a three-variable model which takes into account free (Z(x,t) variable) and membrane-bound (Y(x,t) variable) F-actin as well as nucleation proteins (X(x,t) variable) along the antero-posterior axis which connects the two poles of the cell. The model equations are as follows:

$$\begin{cases} \frac{\partial X(x,t)}{\partial t} = K - \alpha XY + D_1 \frac{\partial^2 X}{\partial x^2} \\ \frac{\partial Y(x,t)}{\partial t} = \alpha XY - \beta YZ + D_2 \frac{\partial^2 Y}{\partial x^2} \\ \frac{\partial Z(x,t)}{\partial t} = \beta YZ - \gamma \frac{Z}{(\delta + Z)} + D_3 \frac{\partial^2 Z}{\partial x^2} \end{cases}$$

with zero-flux boundary conditions.

This model assumes the existence of non-linear reactions in which both the synthesis of F-actin and transformation of membrane-bound into free actin are autocatalytic. Furthermore, the diffusion coefficient of actin microfilaments bound to the membrane is very low, 4.10^3 times less than diffusion of nucleation proteins. Free actin microfilaments are partly integrated in the constriction ring, where they participate with myosin to the oscillatory

cell contraction. The above1D model generates oscillatory actin waves between the two poles of the cells (LeGuyader and Hyver, 1997). A still more representative aspect of this oscillatory dynamics is obtained by considering a 2D representation of the unfolded 3D cell cortex, projected onto a 2D surface. We simulated the behaviour the above model in such a 2D space, with additional scaling of the cortex deformation in the y direction by the F-actin concentration of the constriction ring (Figure 5). We thus obtained a back and forth motion of the constriction ring along the cell.

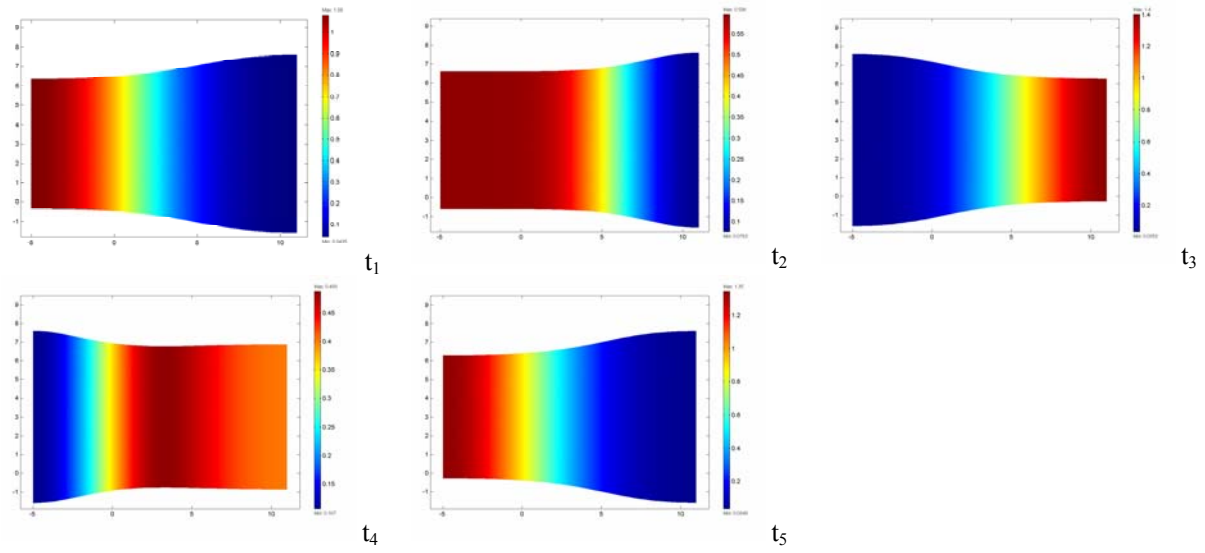


Figure 5: Simulated evolution of the back and forth propagation of the F-actin concentration Z(x,y,t) between the two cell poles. The period is about 10 time units and five consecutive times are presented. Constriction of the projected 2D cell cortex is obtained by rescaling the y dimension with the Z values along the x axis (see plate 25).

This theoretical behaviour is in agreement with the experimental work of Bornens et al. (1989) and with more recent work showing that the disruption of the microtubule network by nocodazole induces cortical oscillations (Pletjushkina et al., 2001). Bornens et al. (1989) suggested that oscillating concentrations of nucleation proteins between the two poles of the cell would indeed create a polymerisation/depolymerisation wave of actin travelling through the cell. Such behaviour was also reported during the extension of pseudopods in *Dictyostelium discoïdeum* (Vicker et al., 2000; Vicker, 2002).

3.5. Protrusive dynamics due to the modulation of stress-strain relationships within the actomyosin cytogel.

The cytomechanical model of Lewis and Murray (1992) extends the solation/gelation model by considering the stress-strain relationships within the actomyosin cytogel. The cytogel is modelled as a viscoelastic continuum submitted to active stress and osmotic pressure. In addition, the sol/gel transition is controlled by the resulting strain level within the cytogel. At high strain, the gelation rate is increased, while at low strain, the solation rate increases. The model dynamics is controlled by the non-linear stress-strain relationship defining qualitatively the contractile actomyosin stress. The inhomogeneous spatial solutions generated by this model have been specifically discussed with regard to the patterned formation of microvilli at the cell surface (Murray, 2003).

3.6. Coupling of actin-dynamics with cell cortex curvature.

In a series of papers, Alt and col. (1995; 1999) proposed a modelling approach in which cell protrusions dynamics are due to the biophysical properties (viscoelasticity, contractility,...) of the cortical network of actin and myosin filaments underlying the cell membrane. This more or less dense network is able (i) to disassemble at locations where it becomes too condensed, (ii) to reassemble in cell protrusions like lamellipodia. Thus, cell protrusions mainly result from a mechanical balance between stresses acting on the cell cortex (mechanical forces generated by the actomyosin complex, tension forces due to the local membrane curvature (Raucher and Sheetz, 2000)) and associated F-actin polymerisation/ depolymerisation induced by intracellular network -free space variation in each moving protrusion.

A minimal three-variable model has therefore been developed to describe the spontaneous, self-organized, dynamics of cell deformations. This model has been used to study the spatio-temporal deformations of keratinocytes (Alt et al., 1995) as well the morphological changes in L929 fibroblasts (Stephanou et al., 2003, Fig. 6). The model takes into account (i) the dynamics of F-actin polymerisation/depolymerisation in the cell cortex, (ii) the contractile activity generated by the actin/myosin interactions, (iii) the F-actin convection. The local amount of F-actin also determines the intensity of the resistive stress applied on the membrane as the result of CSK-cell cortex attachments. This resistive stress plus the stress induced by the cell cortex local curvature is assumed to balance the intracellular hydrostatic pressure.

The analysis of the morphological changes of adherent cells is a three-dimensional free-boundary problem. However, for cells cultures on a rigid substrate, a simpler one-dimensional problem can be considered by considering the cell surface projection on the substrate and by further assuming that the F-actin density as well as its convective tangential velocity is constant in the radial direction. In a cylindrical coordinate system (r,θ), the remaining variables in the cytomechanical model are then: (i) the F-actin concentration in the cortex $a(\theta,t)$, (ii) the F-actin tangential velocity $v(\theta,t)$, (iii) the cell membrane position or the width of cell cortex annulus $L(\theta,t)$ measured from an virtual cell body delimited by an inner circle with radius R_0, with $L(\theta,t) << R_0$ (Alt and Tranquillo, 1995).

The spatio-temporal evolution of these three variables is given by a system of three partial differential equations which define respectively:

- variations of cortical F-actin concentrations, where the net rate η of actin polymerisation/depolymerisation depends on the local value of F-actin concentration relatively to the chemical equilibrium value a*.

$$\frac{\partial(L.a)}{\partial t} + \frac{\partial(L.a.v)}{\partial \theta} = \eta.L.(a^* - a)$$

- the balance of forces applied on the cell cortex in the radial direction. The model takes into account a viscous friction of the cell protruding over the rigid substratum, with coefficient ϕ_1, the intracellular hydrostatic pressure β_1, the resistive elastic stress of the cell CSK controlled by the elasticity coefficient $\gamma_1$and a curvature-dependent stress due to the surface tension of the cell cortex modulated by the coefficient τ_1.

$$a.\phi_1 \frac{\partial L}{\partial t} = \beta_1 - \gamma_1.L.a + \frac{\partial}{\partial \theta}\left(\tau_1.a.\frac{\partial L}{\partial \theta}\right)$$

- the balance of forces in the tangential direction. It includes the frictional drag of the actin cortex moving in the viscous cytosol, with magnitude controlled by the drag coefficient

ϕ_0, a viscous stress with viscosity coefficient μ and the membrane curvature induced stress with coefficient τ_0.

$$a.\phi_0.v = \frac{\partial}{\partial\theta}\left[\mu_0.a.\frac{\partial v}{\partial\theta} + \sigma_0(a,a_{sat}) - \frac{\partial}{\partial\theta}\left(\tau_0.a.\frac{\partial L}{\partial\theta}\right)\right]$$

In addition, the contractile stress of the actomyosin network is modelled by the non-linear function σ_0 (a(θ,t), a_{sat}). Two mechanical states can be distinguished according to the value of the network F-actin concentration a(θ,t)). At low concentration values (a(θ,t) < a_{sat}), the contractile stress increases, while above the saturation threshold a_{sat}, the contractile stress decreases exponentially as a consequence of the network swelling. The non-linear function σ_0 (a(θ,t), a_{sat}) proposed by Alt and Tranquillo (1995) is the following:

$$\sigma_0(a,a_{sat}) = \psi_0.\, a^2.\, \exp(-a/a_{sat})$$

where the coefficient ψ_0 controls the magnitude of the contractile stress.

The existence and properties of protruding and retracting cell membrane protrusions are related to oscillatory solutions of the cytomechanical model. This theoretical analysis is performed in a standard way by looking for critical values of the model's parameters above which small random perturbations are amplified (Hopf bifurcation) until a coherent spatio-temporal pattern emerges with typical unstable modes or wave length (Fig. 6). As it may be expected intuitively, high values of the cell cortex surface tension reduce the number of cell protrusions whilst high values for the contractile efficiency of the actomyosin network contractility increase the number of oscillatory cell protrusions by favouring the destabilisation of higher unstable modes. In an extension of this model (Stephanou and Tracqui, 2002), the influence of extra-cellular factors on protrusion dynamics and cell migration has been analysed.

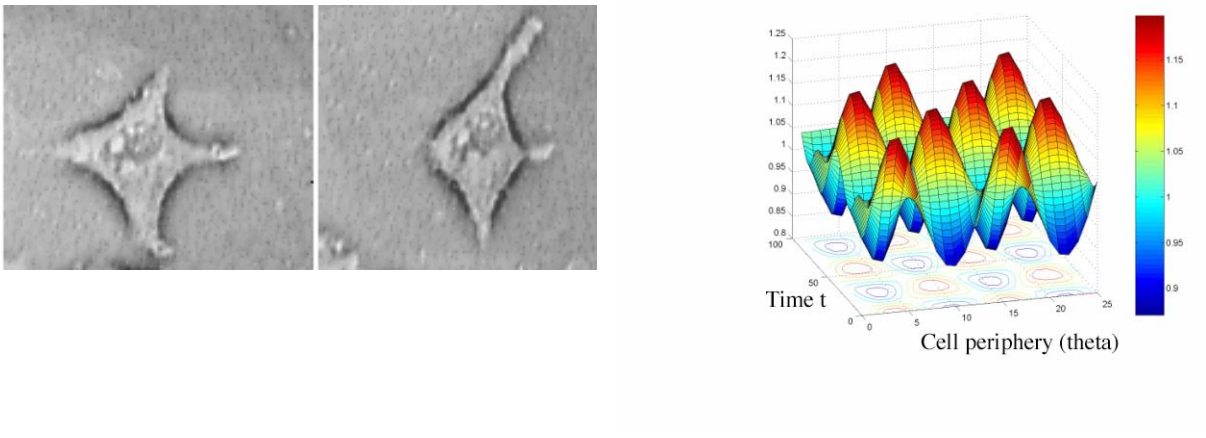


Figure 6: Spontaneous oscillatory protrusions of a L929 fibroblast observed at two consecutive times (left) and simulated (right) evolution with time t_s of the length L(θ, t_s) of the cell protrusion along the cell periphery (angular position θ) for an unstable spatial mode m=4 corresponding to oscillatory changes of cell shape with two protrusions occurring alternatively at perpendicular directions (see plate 26).

3.7. Modelling neuron actin-driven deformation

Neuron plasticity presents different cases of cell deformation coupled to neuron development, such as axon guidance and spine formation during neuron potentiation (Engert and Bonheoffer, 1999, Da Silva and Dotti, 2002, Colicos et al., 2001, Matus, 2000). Basically, neuron actin-driven deformation is under the control of two intracellular signalling pathways mediated, respectively, by RhoA and Rac1/Cdc42 (see e.g. Ridley, 2001, Dickson, 2001). Although both can influence the actin polymerisation/depolymerisation process, they have opposite effects on cell modification. In addition, actin dynamics regulation by proteins such as ADF/Cofilin is also involved in the progression of diseases (Alzheimer); some evidences suggest that altered production and/or localization of these proteins leads to cognitive impairment (Bamburg and Wiggan, 2002).

In a recent study, N. Huc et al. (2002) considered a model for cell shape modification accompanying neuron potentiation or development (Huc et al., 2002). This model is based on a set of ODE for the signalling routes from the calcium influx to actin polymerisation activation (signalling dynamics). In a second step, we considered the mechanical properties of a dendrite.

Signalling dynamics. Basically, we use a simplified version of the model by Zhabotinsky (2000) for auto-phosphorylation of the CaMKII in response to calcium influx. In the original model, the CaMKII is a decameric enzyme and 11 states are possible, depending on the degree of subunit phosphorylation. In the present work, we simplify the equations so that the kinase is represented by a low, medium and highly phosphorylated state, denoted C_0, C_1 and C_2. The phosphatase PP1 and its inhibitor I1 are also considered. Calcium induces the dephosphorylation of the active form of the inhibitor through calcineurin activation. In contrast, the kinase A (PKA) phosphorylate the inactive form of the inhibitor.

$$
\begin{aligned}
\frac{dC_0}{dt} &= -K_1\left(\left[Ca^{2+}\right],\{C_i\}\right)C_0 + K_2\left(\left[Ca^{2+}\right],\{C_i\},P\right)C_1 \\
\frac{dC_1}{dt} &= K_1\left(\left[Ca^{2+}\right],\{C_i\}\right)C_0 - K_2\left(\left[Ca^{2+}\right],\{C_i\},P\right)C_1 \\
&\quad - K_3\left(\left[Ca^{2+}\right],\{C_i\}\right)C_1 + K_4\left(\left[Ca^{2+}\right],\{C_i\},P\right)C_2 \\
\frac{dC_2}{dt} &= K_3\left(\left[Ca^{2+}\right],\{C_i\}\right)C_1 - K_4\left(\left[Ca^{2+}\right],\{C_i\},P\right)C_2 \\
\frac{dP}{dt} &= -k_i PI + k_{-i}\overline{P} \\
\frac{dI}{dt} &= -k_i PI + k_{-i}\overline{P} + k_{PKA}\bar{I} - k_{CaN}\left(\left[Ca^{2+}\right]\right)I
\end{aligned}
$$

C_0, C_1 and C_2 refer to low, medium and high phosphorylation states of the multimeric CaMKII (Zhabotinsky, 2000); P is the phosphatase PP1 which is inhibited by I1 (ki and k-i are the association and dissociation rates.) CaMKII autophosphorylation is triggered by calcium (via the calmodulin.) Dephosphorylation depends on the phosphatase (PP1.) Kinetic rates *Ki* are non-linear functions of calcium, the different forms of the CaMKII and the phosphatase *P*; k_{PKA} and k_{caN} are the PKA-dependent phosphorylation and CaN-depend dephosphorylation rates of the inhibitor.

Mechanics. We assume that the material in the dendrite can be described in the framework of the elasticity equations. The force equilibrium inside the dendrite section reads:

$$\text{in the dendrite} \quad : \quad \rho \frac{d^2\mathbf{u}}{dt^2} = \nabla.\boldsymbol{\sigma}(\mathbf{u}, E(C_2))$$
$$\text{on the membrane} \quad : \quad \boldsymbol{\sigma}(u, E(C_2))\mathbf{n} = p(C_2, t)\mathbf{n}$$

$\boldsymbol{u}$ is the displacement of the material; $\boldsymbol{\sigma}$ is the stress tensor; $\mathbf{n}$ is the outward normal vector; E is the Young modulus of the material. We assume that the active form of the CaMKII (C_2) is responsible for a pressure term exerted on the material at the boundary. To connect the biochemical activation of CaMKII and the mechanics of spine protrusion, we assume that the kinase has a dual role. First, as suggested by experimental models, actin is directly involved in membane protrusion (Borisy and Svitkina, 2000.) Thus, we suppose that CaMKII activity controls a pressure p which is given by the first differential equation. In addition, there is a wealth of experimental evidence that cell filaments are severed or capped by proteins whose activity is Ca^{2+} or kinase dependent (Kwiatkowski, 1999.) Therefore, we suppose that CaMKII may control the material stiffness according to the second equation below.

$$\tau \frac{dp(C_2,t)}{dt} = P_0 f_2(C_2) - p(C_2,t)$$
$$E(C_2) = E_0 - (E_0 - E_1) f_1(C_2)$$

In the above equations, τ is a relaxation time; P_0 is the maximal pressure exerted on the membrane; f_2 is a Hill function of degree 6; E_0 and E_1 are, respectively, the minimum and maximum Young modulus of the material; f_1 is a Hill function of degree 2. The stress tensor $\boldsymbol{\sigma}$ is related to the displacement $\boldsymbol{u}$ and to the Young modulus E by the classical constitutive relation of the elasticity theory (Landau and Lifchitz, 1967):

$$\boldsymbol{\sigma}(\mathbf{u}) = \lambda Tr(\boldsymbol{\varepsilon})\mathbf{I} + 2\mu\boldsymbol{\varepsilon}$$

where $\mathbf{I}$ is the identity matrix; $\mathrm{Tr}(\boldsymbol{\varepsilon})$ is the trace of the strain tensor $\boldsymbol{\varepsilon}$; λ and μ are the Lamé coefficients:

$$\lambda = \frac{E(C_2)\nu}{(1-2\nu)(1+\nu)} \quad ; \quad \mu = \frac{E(C_2)}{2(1+\nu)}$$

where ν is the Poisson coefficient of the material. The strain tensor $\boldsymbol{\varepsilon}$ is the symmetric part of the gradient of the displacement, $\boldsymbol{u}$, which reads:

$$\boldsymbol{\varepsilon} = \frac{1}{2}\left(\nabla\mathbf{u} + {}^T\nabla\mathbf{u}\right) = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$$

Illustration. We consider a disk representing a dendrite section; the calcium influx is at the top of the disk. As shown in figure 7, the simulation of the above equations gives rise to a membrane protrusion (figure 7, left panel). Assuming that CaMKII controls the cytoskeleton mechanical properties (cf. the above section on mechanics), the dendrite protrusion is larger (figure 7, right panel).

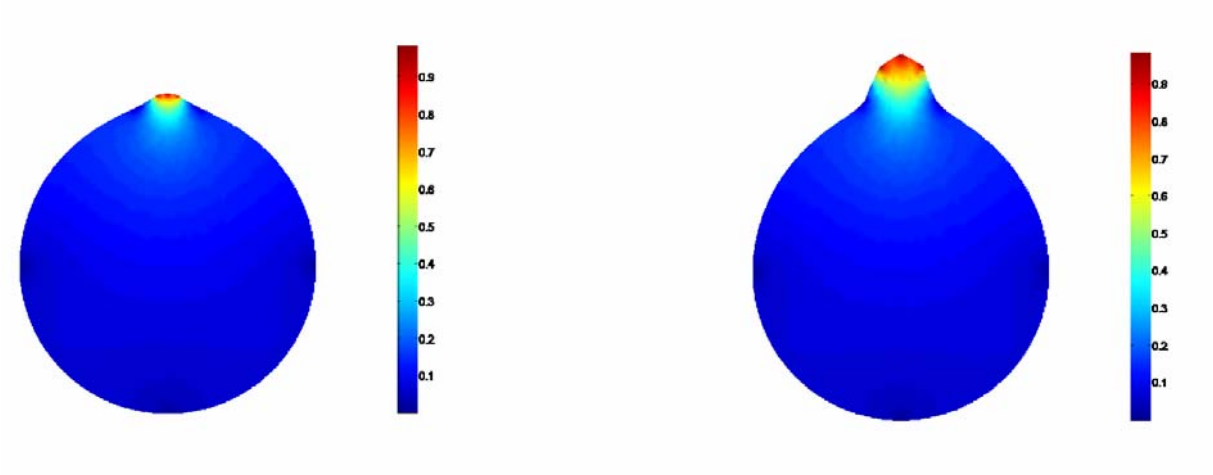


Figure 7: Force generation and membrane deformation. CaMKII-dependent induction of a stable spine protrusion in a dendrite. Active CaMKII generates a 5.5 pN force at the top of the dendrite section. The colormap code for the active CaMKII fraction. A noticeable protrusion is observed when the Young modulus of the cell decreases with CaMKII activity (compare right vs. left) deformation (see plate 27).

3.8. Simulated growth of branched actin filaments

To study the assembly, the movements and the dissociation of large amount of molecules in a virtual cell a simulation program has been developed. The simulator is driven by the description of the model written in a language we have also developed. With this language, the simulation program is not only dedicated to a particular model. We will present here a simulation of the growth of branched actin filaments in a prokaryote simulated cell.

The simulation program implements a three dimensional space bounded by a spherical membrane representing the virtual cell. This cell is initially filled with a population of molecules of various types. When the simulation begins, these molecules diffuse and interact according to the reaction rules described in the model. Periodically, the simulator shows the content of the cell with a 3D *OpenGL* user interface. During the simulation, the user can rotate the cell, focus to and zoom a particular area.

3.8.1. Simulator description

The simulator is a stochastic automaton driven by reaction rules between molecules. Each molecule is represented by a record including its type, its position, a list of links to some other molecules and other internal data. The simulator keeps track of each assembly in real time from the computer point of view.
A step of simulation is the done by examining each molecule and applying the following process: (the molecule *S* is chosen randomly in order to avoid artefacts)

- Check if close enough to *S* in a location randomly chosen *L* there is another molecule *T*.
- If so, and if a reaction rule is given between a molecule of the type of *S* and a molecule of the type of *T*, this rule is applied, according to a probability representing the reaction kinetics.
- Else, molecule *S* may move to the empty location *L*, according to a probability representing the diffusion speed.

When all the molecules involved in the cell have been processed, the current step is completed and new one can begin. The step simulated time slice is set to 100 micro-seconds, which correspond to the average time for a protein to move to a distance of 10 nanometres (approximately its diameter).

3.8.1.1. The rules

The simulator implements four kinds of interaction rules between two molecules: the source *S* and the target *T*:

- **Reaction**: *S* reacts with *T* to produce two other types of molecules *S′* and *T′*.
- **Association**: *S* get bound to *T* to produce the complex *S′*-*T′*. Of course *S′* (resp. *T′*) can keep the type of *S* (resp. *T*).
- **Dissociation**: the complex *S*-*T* can break and leave individual molecules *S′* and *T′*
- **Catalyse**: the complex *S*-*T* can be transformed to *S′*-*T′*.

Each rule is given a probability of execution, which, on the long run correspond to a reaction kinetics. For the association rule, a maximum number of links can be specified.

3.8.1.2. Configuration

The simulator use a configuration file to describe the model the user wants to simulate. This file contains four sections. The first one describes the molecules involved in the model. The second specifies the diffusion rate of each molecule. The third section describes all the reaction rules which will be applied during the simulation. The last section describes the initial population and location for each kind of molecules. Here is an example of the types of molecules involved in the simulation of the growth of actin filaments:

```
molecule
      BR (0, 200, 200),      // branching protein.
      P (200, 0, 0),         // filamentous actin 'plus' end.
      M (0, 200, 0),         // filamentous actin 'minus' end.
      AF (200, 200, 0),      // actin inside the filament.
      AG (200, 100, 0);      // phosphorylated globular actin.
```

With these definitions, the *plus* end (**P**) will be displayed in red, the *minus* end (**M**) in green, the filament itself (**AF**) in yellow and the free globular actin (**AG**) in orange. The branching protein (**BR)** will be displayed in cyan.
In this example, only the free G-actin molecules and the branching protein can diffuse. The filaments themselves are frozen:

```
      speed (AF) = 0.0;      // diffusion speed is zero
      speed (P) = 0.0;       // for the filaments.
      speed (M) = 0.0;
      speed (AG) = 1.0;      // high diffusion speed.
      speed (BR) = 1.0;
```

The following rules show the formation of the polarised dimers from two free phosphorylated free G-actin molecules:

```
      AG  +  AG  ->   M(1) * P(1)      [0.05];
```

```
M   *  P   ->   AG   + AG        [0.5];
```

The right part of the first specify that a minus end **M** can be bound to only one plus end **P**, and conversely, a plus end **P** can be bound to only one minus end **M**. The second rule is the reverse reaction, the depolymerisation of the complex, giving back two free G-actin molecules.
The next rules show the growth of a filament from the plus end:

```
AG  +  P  ->   P(1)  =  AF(1)    [0.8];
P   *  AF ->   AG    +  P        [0.001];
```

The first rule shows how a free G-actin molecule can be bound to the plus end of an already existing filament. The equal sign in the right part of the rule specify that the link must be aligned with the filament. The second rule shows the reverse reaction, the depolymerisation from the plus end giving back one free G-actin molecule.

The next two rules show the growth of a filament from the minus end. One can notice that the polymerisation kinetic is lower than the previous one.

```
AG  +  M   ->   M(1)  =  AF(1)   [0.3];
M   *  AF  ->   AG    +  M       [0.005];
```

This last rule shows how the branching protein makes grow a filament from another one: the branching protein is bound to a filamentous F-actin molecule with an angle of 70° (this is the meaning of the slash in the right part of the rule) and become the plus end of the new filament:

```
B  +  AF  ->   P(1) / AF(1)      [0.5];
```

3.8.1.3. Initialisation

The simulation is initialised by the statement:

```
cube (0, 0, 0, 12, AG);
cube (10, 0, 0, 5, BR);
```

A cube of length *12* located at the centre of the cell *(0, 0, 0)* is filled with 12^3 = 1728 molecules of free globular actin. Another cube of length *5* located at 10 along the X axis is filled with 125 branching proteins. When the simulator is started, these molecules will diffuse and react according to the rules. Polarised dimers first assemble and then the filaments grow until one end touches the membrane or an equilibrium state is reached (see Fig. 8).

3.8.2. Simulation results

First, we have observed that the number of primary filaments depends only on the average number of dimers which in turn depends on the concentration of G-actin and on the dimerisation kinetics. Second, the lengths of the filaments are distributed in two groups: a small number of very long filaments and a gaussian distribution of the length of the other ones. Quite all the filaments are stopped by the membrane of the cell. As the polymerisation speed is greater for the plus end, this end is statistically the one which touches the membrane.

The other end is growing and shrinking and reaches a stationary state (depending again on the concentration of G-actin and on the polymerisation kinetic).

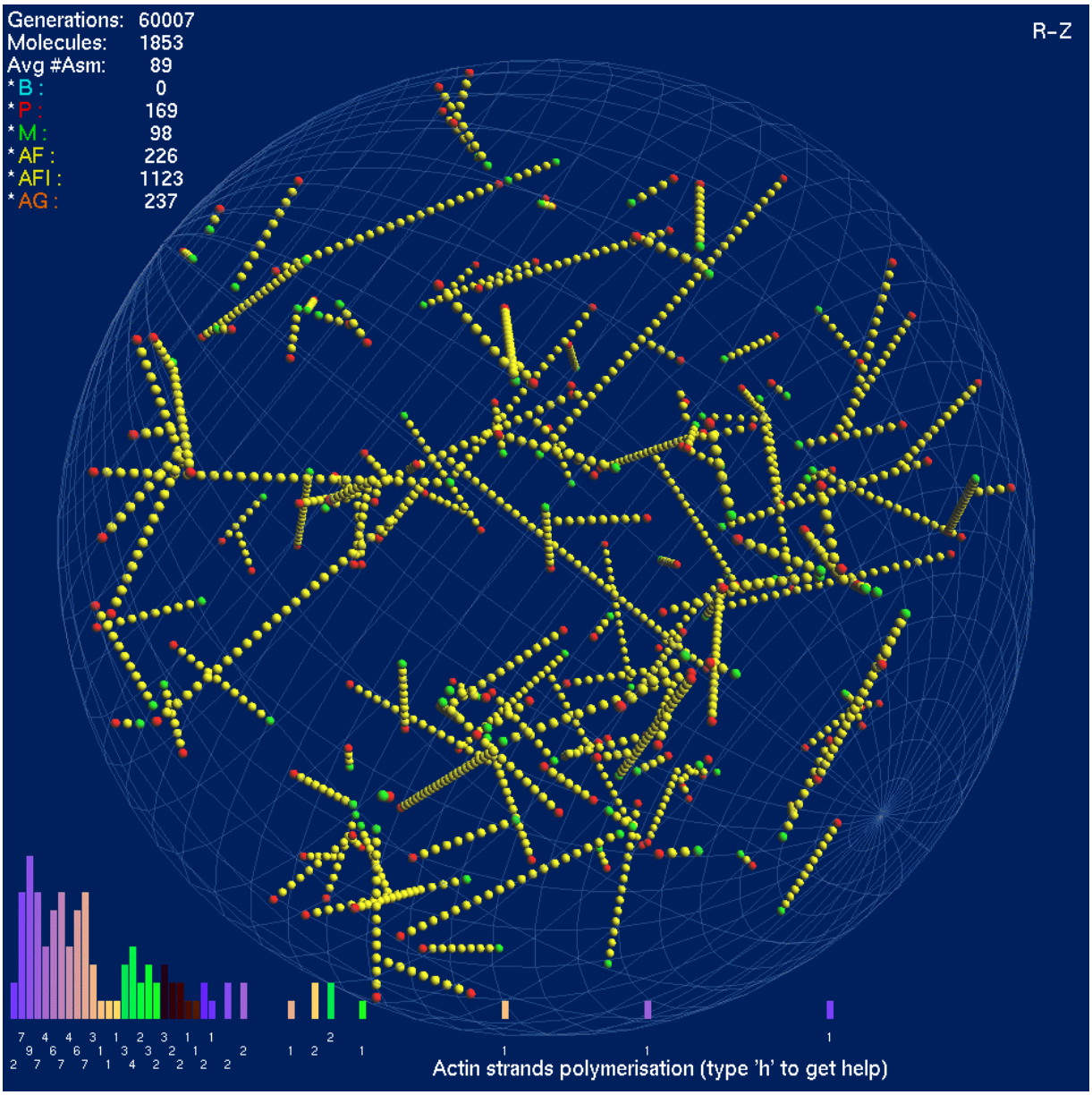


Figure 8: A view of the virtual cell filled with growing actin filaments. On the bottom left of the screenshot there is a histogram of the lengths of the filaments (see plate 28).

The long filaments are those which pass near the centre of the cell, some of them extends to the membrane by their two ends. The small ones are close to the membrane and nearly parallel to it. In this particular simulation, the branching protein is removed just after initiating a new branch of actin. As a consequence, the number of branches is at most equal to the number of branching proteins. It may have fewer branches because the depolymerisation process can crunch a whole branch. Because of the presence of the membrane, the concentration of branching protein is a bit higher near the membrane than in the centre of the cell. This is the reason why the filaments have branches more often near the membrane than in the centre of the cell.

4. Artificial tensegrity

Another approach to studying the coupling between mechanical forces and cytoskeletal dynamics is to construct an *in silico* system in which populations of artificial cells containing different proteins with cytoskeletal properties are subjected to selection for resistance to hydrostatic pressure. By allowing mutations to alter the properties of the proteins and by selecting the surviving cells, it might be expected that one or more types of CSK would evolve. In essence, the idea is to explore the parameters underlying the formation of tensegrity structures by feeding artificial cells and selecting for those that evolve the best structures. In our project, the initial cell consists of a lipid membrane in the form of a monolayer and several types of proteins, the membrane is under pressure and membrane units can diffuse. The cell is fed by the random insertion of proteins and lipids.

The following rules are applied:

1. Turgor pressure results from the difference in concentration of molecules between the outside and the inside of the cell.
2. A cell is maintained until it lyses where lysing is defined as having a breach in membrane integrity that cannot be repaired within a certain period.
3. Two cells are compared and the one retained is either the one that lasts the longer or the one that maintains the higher turgor pressure.
4. Mutations are made by introducing new types of proteins.
5. Components that are used are less likely to be discarded than those that are not used (use is defined as forming part of a structure under tension or compression)
6. Tension in polymers can rupture links or increase lateral interactions (by 'deforming' constituent proteins so as to strengthen the polymer)

Components may include:

- Membrane proteins that respond to curvature and to which other proteins can bind
- Proteins that cause filaments to branch at different angles
- Proteins that nucleate radial structures
- Proteins that can act as anchors in the membrane
- Proteins that cross-link filaments
- Proteases
- Proteins that change affinities for other proteins under stress
- Lipids of two types (cone and inverted cone) that form a monolayer.
- Calcium (in the form of a gradient that is higher outside the cell)

Variables include:

- Binding affinities that may depend on the tension in the system or on activation by another protein (equivalent of post-translational modification). These affinities should allow for the possibility of dynamic assembly and disassembly of polymers.
- Growth by insertion of new proteins and lipids into 'neutral spaces' in different ratios and at different relative rates.

Mechanical aspects:

Turgor is calculated from the density of molecules within the cell. Individual molecules produce more turgor than those that are in the form of polymers or aggregates (molecules with no free spaces around them generate no turgor). This turgor then acts at the membrane. The membrane can deform by movement of lipids normal to its plane.

At this stage, artificial tensegrity is simply a gedanken or thought experiment. Nevertheless, several predictions can be made that might be tested by selection over many generations. Firstly, large cells should have fewer problems with turgor pressure than small cells. Secondly, calcium should have an important role in strengthening the CSK to resist lysis (hence a small leak is self-repairing). Thirdly, cells that do not grow (or move) should have a CSK parallel to the membrane whilst those that can grow should have one that is perpendicular. Fourthly, there should be a reserve pool of lipids and cytoskeletal components near the membrane in readiness for incorporation in the membrane in time of need. Finally, division should occur spontaneously between two segregated cytoskeletal assemblies.

5. Discussion

As recently pointed out by T. D. Pollard (Pollard, 2003) in the context of cellular motility, the complexity of the mechanisms driving cell dynamics will force cell biologists to depend increasingly on mathematical models to test their hypotheses. This brief review emphasizes the advantages and limitations of continuous versus discrete modelling approaches to cell behaviour. Continuous models can account for a large variety of cellular dynamics including the protrusive activity which is coupled to modifications of continuous mechanical properties such as membrane tension, cell cortex viscoelasticity or mechanical stresses developed by the F-actin network. However, a more refined description of CSK organisation, including the orientation of filaments or the formation of stress fibres, is hardly compatible with a continuous formulation, although recent models bridge the gap between mesoscopic mechanical properties of the actin cortex and a description at the molecular level (Maly et al, 2001). Thus, discrete tensegrity models seem to provide a more adequate description of the cell as a physical object. However, the tensegrity paradigm is still a matter of active controversy, as illustrated by recent papers (Ingber et al., 2000; Wang et al., 2001). For Ingber, the intransigence of the remaining critics seem to "largely result from an overly strict definition of what tensegrity is and how it can be applied" (Ingber et al., 2000). As quoted above, the tensegrity model states that: (i) cells and tissues exhibit integrated mechanical behaviour through use of specific structural principles, namely the discontinuous-compression/continuous-tension construction submitted to a pre-existing tension or prestress (Pourati et al., 1998), and (ii) the cell has an elastic submembranous skeleton with its associated lipid bilayer which can be linked to the internal CSK depending on type of cell adhesion.

Can experiments be designed to discriminate between the models? It is worth reporting here the different interpretations of apparently similar experiments such as those based on the induction of cell deformation through a direct manipulation of transmembrane receptors. According to Ingber and col., on the one hand, application of mechanical stresses to integrins using surface-bound micropipettes pre-coated with fibronectin induces CSK reorganisation, nucleus elongation along the tension lines as well as reorganisation within nucleoli, i.e. deep inside the nucleus (Maniotis et al., 1997). On the other hand, and as expected, application of similar stresses to membrane receptors that are not linked to the F-actin CSK did not result in such reorganisation. Opposite conclusions, however, were drawn by Heideman and col. (1999; cited in Ingber et al., 2000) based on the application of similar mechanical stresses to integrin membrane receptors with glass needles treated with laminin, an ECM adhesion protein. Formation of an actin spot was observed on the cytoplasmic side of the membrane, inducing a locally high deformation of the membrane rather than a global change in cell shape. Heideman and colleagues thus concluded that the elastic cortical CSK is not connected to the internal microtubule cytoskelton, which is in complete disagreement with the fundamental tenet of tensegrity.

Analysing the argument in detail is beyond the scope of this review but we should mention that Ingber's reply is that experiments showing a lack of action at a distance when pulling on the cell via integrins *before* the formation of focal adhesion complexes are not valid as proofs of the failure of the tensegrity model. It is clear that the tensegrity paradigm is stimulating both theoretical and experimental work greatly and is leading to the development of new physical methods of quantification. For example, it has led to experiments to determine the significance of the compression of microtubules compression for cell mechanics. Ingber (Ingber et al., 2000) reported that microtubules counterbalanced approximately one-third of the total cellular prestress within a cell stimulated by histamine, a chemical constrictor, whilst noting that cell attachment to a rigid substratum would decrease this prestress level. Prestress within the cell can be estimated from the displacement of small fluorescent microbeads within a deformable polyacrylamide gel which can be used as a cell culture substratum, the search for other quantification methods is now on.

From a theoretical point of view, a clear advantage of the tensegrity model paradigm is to provide an alternative view to cell engineering models that can only describe the mechanical behaviour of cells by ad hoc "data fitting" models based for example on a combination of rheological elements. Through the cell tensegrity model, more specific questions can be addressed regarding specific cell behaviours such as strain-hardening or CSK stiffness (Volokh et al., 2000, 2002 ; Volokh, 2003 ; Wendling et al., 2000 ; Canadas et al., 2002) or the process of mechanotransduction. In the latter case, tensegrity-based predictions can be compared to theoretical predictions inferred from other cell models like the percolation model of Shafrir et al. (2000). This possibility of exploring various mechanistic hypotheses is a real advantage of the tensegrity model when compared to other cell simulation models such as Electronic Cell (http://e-cell.org/) or Virtual cell (http:// www.nrcam.uchc.edu/ vcell_development/ vcell_dev.html) where cytomechanical parameters are absent from the theoretical framework. That said, the dynamic remodelling of the cell is not yet taken into account in the current tensegrity models. This is one of the requirements for future work on modelling in which the integration of mechanical and biochemical properties into a single model may genomic data to be interpreted.

Acknowledgements

We thankfully acknowledge Thomas Sauvaget for his contribution to the design of computational "cellular" phymulob objects. Part of this work was supported by the ACI "Bio-informatique" of the CNRS.

References

Abraham V. C., V. Krishnamurthi, D. L. Taylor and F. Lanni, The actin-based nanomachine at the leading edge of migrating cells, *Biophys. J.*, 77:1721-1732, 1999.

Alt W., O. Brosteanu, B. Hinz and H. W. Kaiser, Patterns of spontaneous motility in videomicrographs of human epidermal keratinocytes. *Bioch. Cell. Biol.* **73**, 441-459, 1995.

Alt W. and R. T. Tranquillo, Basic morphogenetic system modeling shape changes of migrating cells: How to explain fluctuating lamellipodial dynamics, *J.Biol.Syst*. 3 : 905-916,1995.

Alt W. and M. Dembo, Cytoplasm dynamics and cell motion: two-phase flow models, *Math Biosci.*, 156(1-2):207-28, 1999.

Balaban N. Q., U.S. Schwarz, D. Riveline, P. Goichberg, G. Tzur, I. Sabanay, D. Mahalu, S. Safran, A. Bershadsky , L. Addadi and B. Geiger, Force and focal adhesion assembly: a close relationship studied using elastic micropatterned substrates, *Nat Cell Biol.* **3**(5): 466-472, 2001.

Bamburg J.R. and Wiggan P. ADF/cofilin and actin dynamics in disease. *TRENDS in cell biology,* 12:598-605, 2002.

Borisy, G.G. and T.M. Svitkina. Actin machinery:push-pull the envelope, *Curr. Opin. Cell Biol.* **12**:104-112, 2000.

Bornens M., M. Paintraud and C. Celati, The cortical microfilament system of lymphoblastes display a périodique oscillatory activity in the absence of microtubules : implications for cell polarity, *J. Cell Biol.,* **109** :1071-1083, 1989.

Caille N, O. Thoumine, Y. Tardy Y and J. J. Meister. Contribution of the nucleus to the mechanical properties of endothelial cells. *J Biomech* **35**(2):177-87.

Canadas P., V. M. Laurent, C. Oddou, D. Isabey and S. Wendling, A cellular tensegrity model to analyse the structural viscoelasticity of the cytoskeleton, *J. Theor. Biol.* **218**:155-173, 2002.

Carlier MF. and D. Pantaloni. Control of actin dynamics in cell motility. *J. Mol. Biol.* **269**:459-467, 1997.

Chen J, B. Fabry , E.L. Schiffrin EL and N.Wang N. Twisting integrin receptors increases endothelin-1 gene expression in endothelial cells. Am J Physiol Cell Physiol 280(6):C1475-1484, 2001.

Colicos M., Collins BE., Sailor M., and Y. Goda. Remodeling of synaptic actin induced by photoconductive stimulation. *Cell,* **107**:605-616, 2001.

Condeelis, J. Life at the leading edge: the formation of cell protrusions; *Annu. Rev. Cell. Biol*. **9**:411-444, 1993.

Coughlin M.F. and D. Stamenovic, A prestressed cable network model of the adherent cell cytoskeleton, *Biophys. J.,* 84:1328-1336, 2003.

Da Silva JS and Dotti C. Breaking the neuronal sphere: regulation of the actin cytoskeleton in neuritogenesis. *Nature Neuroscience*, 3:694-704, 2002.

Dedhar S, Hannigan G E. Integrin cytoplasmic interactions and bidirectionnal transmembrane signalling *Curr Opin Cell Biol* **8** : 657-669, 1996.

Dembo M., T. Oliver, A. Ishihara and K. Jacobson, Imaging the traction stresses exerted by locomoting cells with the elastic substratum method, *Biophys. J.,* **70**:2008-2022, 1996.

Dembo M. and Y. L. Wang, Stresses at the cell-to-substrate interface during locomotion of fibroblasts, *Biophys. J.,* **76**:2307-2316, 1999.

Dickson B. Rho GTPase in growth cone guidance. *Curr. Opin. Neurobiol.* 11:103-110, 2001.

Eastwood M, Porter R, Khan U, Mcgrouther G, Brown R. Quantitative analysis of collagen gel contractile forces generated by dermal fibroblasts and the relationship to cell morphology. J Cell Phys 166 : 33-42, 1996.

Edelstein-Keshet L. A mathematical approach to cytoskeletal assembly. *Eur. Biophys. J.* **27**:521-531, 1998.

Edelstein-Keshet L. and GB. Ermentrout. Models for the length distributions of actin filaments: I. simple polymerization and fragmentation. *Bull. Math. Biol.* **60:**449-475, 1998.

Edelstein-Keshet L. and GB. Ermentrout. A model for actin-filament length distribution in a lamellipod. *J. Math. Biol.* **43**:325-355, 2001.

Edwards Y.S., L.M. Sutherland, J.H. Power, T.E. Nicholas and A.W. Murray, Cyclic stretch induces both apoptosis and secretion in rat alveolar type II cells, *FEBS lett.* **448**(1) :127-130, 1999.

Ehrengruber M. U., T. D. Coates and D. A. Deranleau, Shape oscillations: a fundamental response of human neutrophils stimulated by chemotactic peptides?, *FEBS Lett.,* **359**:229-232, 1995.

Engert F., Bonhoeffer T. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature*, 399:66-70, 1999.

Ermentrout GB., and L. Edelstein-Keshet L. Models for the length distributions of actin filaments: II. Polymerisation and fragmentation by gelsolin actin together. *Bull. Math. Biol.* **60:**477-503, 1998.

Frischknecht F. and M. Way. Surfing pathogens and the lessons learned for actin polymerization. *Trends Cell. Biol.* **11**:30-38, 2001.

Fujisawa T., T. Hattori, K. Takahashi, T. Kuboki, A. Yamashita and M. Takigawa, Cyclic mechanical stress induces extracellular matrix degradation in cultured chondrocytes via gene expression of matrix metalloproteinases and interleukin-1, *J. Biochem.,* **125**(5) : 966-975, 1999.

Geigant E., Ladizhansky K. and A. Mogilner. An integrodifferential model for orientational distribution of F-actin in cells. *SIAM J. Appl. Math.* **59**:787-809, 1998.

Gerbal F., Chaitkin P., Rabin Y. and J. Prost. An elastic analysis of Listeria monocytogenes propulsion. *Biophys. J.* **79**:2259-2275, 2000.

Gerbal F., Laurent V., Ott A., Carlier MF, Chaitkin P. and J. Prost. Measurement of the elasticity of the actin tail of Listeria monocytogenes. *Eur. Biophys. J.* **29**:134-140, 2000.
Germain F., A. Doisy, X. Ronot and P. Tracqui. Characterization of cell deformation and migration using a parametric estimation of image motion, *IEEE Trans. Biomed. Eng*., **46**(5), p.584-600, 1999.

Guilak F., M. Sato, C. M. Stanford and R. A. Brand, Cell mechanics, *J. Biomech*. **33**:1-2, 2000.

Heidemann S.R., S. Kaech, R.E. Buxbaum and A. Matus, Direct observation of the mechanical behaviors of the cytoskeleton in living fibroblasts, *J. Cell Biol.,* **145** :109-122, 1999.

Hénon S., G. Lenormand, A. Richert and F. Gallet, A new determination of the shear modulus of the human erythrocyte membrane using optical tweezers. *Biophys J.* **76**:1145-51, 1999.

Hodgkin A.L. and Huxley A.F. A quantitative of membrane current and its application to conduction ans excitation in nerve. J. Physiol. (Lond.) 117:500-544, 1952.

Huang S. and D.E. Ingber, The structural and mechanical complexity of cell growth control, *Nature Cell Biol.*, **1**:E131-E138, 1999.

Huc N., Nemri A., Goldberg Y. and Martiel J.L. Actin-based spine protrusion in potentiated neurons. FENS Forum, Paris, 2002.

Ingber D. E. and J. D. Jamieson, Cells as tensegrity structures : architectural regulation of histodifferentition by physical forces transduced over basement membrane, in Gene expression during normal and malignant differentition, eds., Academic Press, pp.13-32, 1985.

Ingber D. E., Cellular tensegrity : defining new rules of biological design that govern the cytoskeleton,*J. Cell Sci.* **104**: 613-627, 1993.

Ingber D.E., Tensegrity : the architectural basis of cellular mechanotransduction, *Annu Rev Physio* **59** : 575-79, 1997.

Ingber D. E., S. R. Heideman, P. Lamoureux and R. E. Buxbaum, Opposing views on tensegrity as a structural framework for understanding cell mechanics, *J. Appl. Physiol.,* **89**:1663-1678, 2000.

Janmey P. A., The cytoskeleton and cell signaling: component localization and mechanical coupling, *Physiol. Rev.,* **78**(3):763-781, 1998.

Kaspar D., W. Seidl, C. Neidlinger-Wilke, A. Ignatius and L. Claes, Dynamic cell stretching increases human osteoblast proliferation and CICP synthesis but decreases osteocalcin synthesis and alkaline phosphatase activity, *J. Biomech*. **33**:45-51, 2000.

Killich, T., Plath, P. J., Hass, E. C., Wei, X., Bultmann, H., Rensing, L. and Vicker, M.G. Cell movement and shape are non-random and determined by intracellular, oscillatory rotating waves in dictyostelium amoebae. *Biosystems* **33** :75-87, 1994.

Kolodney M. and R.T. Wysolmersky, Isometric contraction by fibroblasts and endothelial cells in tissues cultures: a quantitative study. *J Cell Biol* **117** :73-82, 1992.
Kwiatkowski D. Functions of gelsolin: motility, signalling, apoptosis, cancer. *Curr. Op. Cell Biol* 11:103-108, 1999.

Landau L., and Lifchitz. Théorie de l'élasticité. Editions Mir Moscou, 1967.

Laurent V.M., S. Henon, E. Planus, R. Fodil, M. Balland, D. Isabey and F. Gallet, Assessment of mechanical properties of adherent living cells by bead micromanipulation: comparison of magnetic twisting cytometry vs optical tweezers. *J Biomech. Eng.* **124**(4):408-421, 2002.

Laurent V.M., E. Planus, R. Fodil and D. Isabey, Mechanical assessment by magnetocytometry of the cytosolic and cortical cytoskeletal compartments in adherent epithelial cells. *Biorheology* **40**(1-3):235-240, 2003.

Le Guyader, H. and C. Hyver, Periodic activity of the cortical cytoskeleton of the lymphoblast: modelling by a reaction-diffusion system. *C. R. Acad. Sci. Paris*, Ser III, **320**:59-65, 1997.

Lelievre S., V. M. Weaver and M. J. Bissell. Extracellular matrix signalling from the cellular membrane skeleton to the nuclear skeleton : a model of gene regulation*, Recent Prog. Horm. Res.* 51:417-432, 1996.

Lewis M. A. and J. D. Murray, Analysis of dynamic and stationary pattern formation in the cell cortex, *J. Math. Biol.,* **31**:25-71, 1992.

Loisel TP, Boujemaa R., Pantaloni D. and MF Carlier. Reconstitution of actin-based motility of Listeria and Shigella using pure proteins. *Nature*: **401**, 613-616, 1999.

Machesky LM. And RH. Insall. Signalling to actin dynamics. *J. Cell. Biol.* **146**:267-272, 1999.

Maly IV. And GG. Borisy. Self-organization of a propulsive actin network as an evolutionary process. *Proc. Nat. Acad. Sci. USA*. **98**:11324-11329, 2001.

Maniotis A.J., K. Bojanowski and D.E. Ingber, Mechanical continuity and reversible chromosome disassembly within intact genomes removed from living cells, *J. Cell Biochem.,* **65** :114-130, 1997.

Matus A. Actin-based plasticity in dendritic spines. *Science.* 290:754-758., 2000.

Meyer C. J., F. J. Alenghat, P. Rim, J. H. Fong, B. Fabry and D. E. Ingber, Mechanical control of cyclic AMP signalling and gene transcription through integrins, *Nature Cell Biol*., **2**:666-668, 2000.

Mogilner A. and L. Edelstein-Keshet, Regulation of actin dynamics in rapidly moving cells: a quantitative analysis, *Biophys. J.* 83:1237-1258, 2002.

Mogilner A. and G. Oster, The physics of lamellipodial protrusion, *Eur. Biophys. J.* **25**:47-53, 1996.

Mogilner A. and G. Oster, G., Polymer motors: Pushing out the front and pulling up the back, *Current Biology,* in press, 2003a.

Mogilner A. and G. Oster, Force generation by actin polymerisation II: the elastic ratchet and tethered filaments, *Biophys. J.*, 84:1591-1605, 2003b.

Murray J. D., Mathematical Biology, 3rd edition, Vol. I-II, Springer Verlag, 2003.

Noireaux V., Golsteyn RM., Friederich E., Prost J., Antony C., Louvard D. and C. Sykes. Growing an actin gel on spherical surfaces. *Biophys. J.* **78**:1643-1654, 2000.

Oster, G. F. and A. S. Perelson, Cell spreading and motility: a model lamellipod, *J. Math Biol*, **21**:383-388, 1985.

Peskin, C. S., G. M. Odell and G. F. Oster (1993). Cellular motions and thermal fluctuations: the brownian ratchet, *Biophys. J.,* **65** :316-324, 1993.

Pollard TD., Blanchoin L. and RD. Mullins. Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. *Annu. Rev. Biophys. Biomol. Struct*. **29**:545-576, 2000.

Pollard T.D., The cytoskeleton, cellular motility and the reductionist agenda, *Nature*, **422**:741-745, 2003.

Pourati J., A. Maniotis, D. Spiegel, J. L. Schaffer, J. P. Butler, J. F. Fredberg, D. E. Ingber, D. Stamenovic and N. Wang, Is cytoskeletal tension a major determinant of cell deformability in adherent endothelial cells ?, *Am. J. Physiol.,* **274**:C1283-1289, 1998.

Pletjushkina O. J., Z. Rajfur, P. Pomorski, T. N. Oliver, J. M. Vasiliev and K. A. Jacobson, Induction of cortical oscillations in spreading cells by depolymerization of microtubules, *Cell Motil Cytoskeleton*, **48**:235-44, 2001.

Promayon E., Baconnier P., Puech C. Physically-based deformations constrained in displacements and volume. In Computer Graphic Forum*, vol.* 15(3), pp.155-164, 1996.

Promayon E., Baconnier P., Puech C. Physically-based model for simulating the human trunk respiration movements. Lecture notes in Computer Science, vol. 1205, pp 379-388. Springer Verlag, 1997.

Promayon E., J. L. Martiel and P. Tracqui, Physical object oriented 3D simulations of cell migration and déformations, in "*Polymer and Cell Dynamics - Multiscale Modeling and Numerical Simulations*", Birkhäuser Basel, pp 125-138, 2003.

Raucher, D. and M.P. Sheetz (2000). Cell spreading and lamellipodial extension rate is regulated by membrane tension, *J. Cell Biol.,* **148**:127-136, 2000.

Ridley A. Rho family proteins: coordinating cell responses. TRENDS in cell Biology, 11, 471-477, 2001.

Sauvaget T., Mémoire de DEA, Univ. J. Fourier, Grenoble, 2001.

Schmidt C, H. Pommerenke , F. Durr , B. Nebe and J. Rychly, Mechanical stressing of integrin receptors induces enhanced tyrosine phosphorylation of cytoskeletally anchored proteins, Mechanical stressing of integrin receptors induces enhanced tyrosine phosphorylation of cytoskeletally anchored proteins, J. Biol. Chem., 273(9): 5081-5085, 1998.

Stamenovic D, Fredberg J J, Wang N, Butler J P, Ingber D E. A microstructural approach to cytoskeletal mechanics based on tensegrity , *J Theor Biol* **181** : 125-136, 1996.

Stamenovic D., Z. Liang, J. Chen and N. Wang, Effect of the cytoskeletal prestress on the mechanical impedance of cultured airway smooth muscle cells, *J Appl Physiol.* **92**(4):1443-1450, 2002.

Stamenovic D. , S. M. Mijailovich, I.M. Tolic-Norrelykke and N. Wang, Experimental tests of the cellular tensegrity hypothesis. *Biorheology*. **40**(1-3):221-225, 2003.

Shafrir Y., Ben-Avraham D. and G. Forgacs, Trafficking and signalling through the cytoskeleton: a specific mechanism, *J. Cell Sci.,* **113**:2747-2757, 2000.

Star E., Kwiatkowski DJ., and VN. Murthy. Rapid turnover of actin in dendritic spines and its regulation by activity. *Nature neurosci*. **5**:239-246, 2002.

Stephanou A., X. Ronot and P. Tracqui, Analysis of cell motility combining cytomechanical model simulations and an optical flow method, in "*Polymer and Cell Dynamics - Multiscale Modeling and Numerical Simulations*", Birkhäuser Basel, pp 91-112, 2003.

Stephanou A. and P. Tracqui, Cytomechanics of cell deformations and migration: from models to experiments, *C.R. Biologies,* **325**:295-308, 2002.

Tabony J., N. Glade , C. Papaseit and J. Demongeot, Gravity dependence of microtubule preparations. *J. Gravitational Physiology* **9**:245- 248, 2002.

Tan J.L., J. Tien, D.M. Pirone, D.S. Gray, K. Bhadriraju and C.S. Chen, Cells lying on a bed of microneedles: an approach to isolate mechanical force, Proc Natl Acad Sci U S A. **100**(4):1484-1489, 2003.

Theriot JA, Mitchison TJ, Tilney LG and DA Portnoy. The rate of actin-based motility of intracellular L. monocytogenes equals the rate of actin polymerization. *Nature*, **357**:257-260, 1992.

Thoumine O. and A. Ott, Time scale dependent viscoelastic and contractile regimes in fibroblasts probed by microplate manipulation, *J. Cell Sci.,* **110**:2109-2116, 1997.

Thoumine O., A. Ott, O. Cardoso and J.J. Meister, Microplates: a new tool for manipulation and mechanical perturbation of individual cells, *J Biochem Biophys Methods.* **39**:47-62, 1999.

Thoumine O. and J. J. Meister, Dynamics of adhesive rupture between fibroblasts and fibronectin: microplate manipulation and deterministic model, *Eur Biophys J.,* **29**:409-419, 2000.

Trinkhaus J. P., Cells into organs. The forces that shape the embryo, 2nd edition, Prentice-Hall, Englewood, 1984.

Vicker M. G., Reaction-diffusion waves of actin filament polymerisation/depolymerization in Dictyostelium pseudopodium extension and cell locomotion, *Biophys. Chem.,* **84**:87-98, 2000.

Vicker M.G., Eukaryotic cell locomotion depends on the propagation of self-organized reaction-diffusion waves and oscillations of actin filament assembly, ,*Exp Cell Res.* **275**(1):54-66, 2002.

Volokh K. Y., O. Vilnay and M. Belsky, Tensegrity architecture explains linear stiffening and predicts softening of living cells*, J. Biomech.* **33**:1543-1549, 2000.

Volokh K. Y., O. Vilnay and M. Belsky, Cell cytoskeleton and tensegrity, *Biorheology* **39**:63-67, 2002.

Volokh K. Y, Cytoskeletal architecture and mechanical behavior of living cells, *Biorheology* **40**:213-220, 2003.

Wang J. H. C., E. S. Grood, J. Florer and R. Wenstrup, Alignment and proliferation of MC3T3-E1 osteoblats in microgrooved silicone substrata sujected to cyclic stretching, *J. Biomech*. **33**:729-735, 2000.

Wang N, Butler J P, Ingber D E. Mechanotransduction across the cell surface and through the cytoskeleton , *Science* **260** : 1124-1127, 1993.

Wang N, K. Naruse, D. Stamenovic, J. J. Fredberg, S. M. Mijailovich, I. M. Tolic-Norrelykke, T. Polte, R. Mannix and D. E. Ingber, Mechanical behavior in living cells consistent with the tensegrity model, *Proc. Natl. Acad. Sci. USA*, **98**:7765-7770, 2001.

Welch MD, Iwamatsu A. and TJ Mitchison. Actin polymerization is induced by Arp2/3 protein complex at the surface of Listeria monocytogenes. *Nature* **385**: 265-269, 1997.

Wendling S., C. Oddou and D. Isabey, Stiffening response of a cellular tensegrity model, *J. Theor. Biol*., **196** :309-329, 1999.

Wendling S., E. Plannus, V. Laurent, L. Barbe, A. Mary, C. Oddou and D. Isabey, Role of cellular tone and microenvirnmental conditions on cytosquelette stiffness assessed by tensegrity model, *Eur. Phys. J. Appl. Phys.,* **9** :51-52, 2000.

Zhabotinsky. A Bistability in the Ca2+/calmodulin-dependent kinase-phosphatase system. Biophys J., 79:2211-2221, 2000.

Computational Models for Integrative and Developmental Biology

Jean-Louis Giavitto[1], Christophe Godin[2],
Olivier Michel[1] and Przemyslaw Prusinkiewicz[3]

[1]Laboratoire de Méthodes Informatiques, Université d'Évry Val d'Essonne & CNRS UMR 8042, Genopole, Tour Évry 2, 523 Place des terrasses de l'agora, 91000 Évry, France
[2]AMAP, UMR CIRAD-CNRS-INRA-Université Montpellier II, TA 40/PS2 34398 Montpellier Cedex 5, France.
[3]Department of Computer Science, University of Calgary 2500 University Drive N.W. Calgary, Alberta, Canada T2N 1N4.

giavitto,michel@lami.univ-evry.fr
godin@cirad.fr
pwp@cpsc.ucalgary.ca

1 Introduction

The relation between biology and computation has a long history reviewed by Langton [LIL89]. In this paper, we classify the interactions between computer science and biology in three areas :

1. *Bioinformatics* develops the automated management and analysis of biological data.
2. *Computational Biology* looks at biological entities as information processing systems with the final goal of a better understanding of nature using computer science notions.
3. *Biological Computing* goes in the reverse direction and studies how biological techniques can help out with computational problems.

Bioinformatics consists of developing software tools to support and help the biologist in the analysis and comprehension of biological systems. A good example is the development of data-bases supporting the genome project [Kan00].

Biological Computing imports some *biological metaphors* [Pat94] to develop new way of computing and to design new algorithms. From the beginning of computer sciences, biological processes have been abstracted to produce new computational models : formal neural networks inspired by natural neurons, evolutionary algorithm inspired by Darwinian evolution (see the "Parallel Problem Solving from Nature" (PPSN) conference series), parallel computer architecture (e.g. cellular automata) inspired by biological tissues (see for example the "Information Processing in Cells and Tissues" (IPCAT) conference series), DNA computing abstracted from biochemistry [Pau98a], cooperative distributed algorithm (e.g. multi-agents) motivated by ethological behaviors or social interactions, ...

Computational Biology. Here we are mainly interested in computer modeling and simulation of biological processes. The computer simulation of a biological process implies the definition of a model sufficiently rigorous to lead to a program. With such a formal model, it is possible to systematically explore the system's behavior and sometimes to make predictions. This kind of study is part of the more general idea of *simulated experiments* (also called *in silico* experiments by biologists and numerical experiments by physicists). These experiments are required when in-vivo or in-vitro experiments are out of reach for economical, practical or ethical reasons. Note however that the simulation of a computer model is only one of its possible use : because it is formal, it is possible to reason about it and for example to infer some properties (existence of steady state, stability, phase changes, etc.) that can be checked against the natural phenomena.

More generally, formal models can have a pedagogical, normative, constructive or ideological role :

- pedagogical and heuristic : the model is used to share knowledge about a given system or to illustrate a set of complex relationships involved in a biological process.
- normative : the model is used as a reference between scientists or to compare several systems.
- constructive : the model is used as a blueprint in the design of a new biological entity. Biology has reached the point where in addition to the study of already existing natural entities, it has to design new biologicals artifacts (drug design, metabolic pathways, genetically modified organisms, ...).
- ideological : a model illustrates some biological paradigm and constraints furthermore the investigated schemes. Biology has imported a number of notions developed in computer science, for instance the notion of programs, memory, information, control, etc. [Ste88, Kel95], that have then structured biological theories.

The transfer of concepts and tools between biology and computer science is not a one-way process and often, a computing model inspired initially by a biological phenomena, leads to a formalism used later in simulation of some (other) biological processes. A good example is given by the history of cellular automata (CA) : initially developed by J. Von Neuman [VN66], they abstract the idea of a tissue of cells, to investigate the notion of self-reproducing programs. The CA formalism then has been largely used in biological simulation, for example to model the growth of tumor (Eden's models) or in ecology (it has been also successful in numerous other application domains, like in physics).

The contributions of Computational biology in the area of molecular dynamics or ecological modeling, are now well established. They are largely centered around the notion of *dynamical systems*. What appears now is that this kind of computational models can make connections between molecular mechanisms and the physiological properties of a cell. The theme

$$\text{gene expression} \longrightarrow \textit{system dynamics} \longrightarrow \text{cell physiology}$$

is an emerging paradigm [JTN00] that becomes increasingly more important as we try to integrate the exponential knowledge of all the cells components in a true understanding of the cell. However, this formalization from biology to dynamical system and back to biology, has long been advocated in the more general domain of the development [Smi99, Kau95].

2 Dynamical systems

2.1 Basic definitions

Many natural phenomena can be modeled as *dynamical systems*. At any point in time, a dynamical system is characterized by its *state*. A state is represented by a set of *state variables*. For example, in the description of planetary motions around the sun, the set of state variables may represent positions and velocities of the planets. Changes of the state over time are described by a *transition function*, which determines the next state of the system (over some time increment) as a function of its previous state and, possibly, the values of external variables (input to the system). This progression of states forms a *trajectory* of the system in its *phase space* (the set of all possible states of the system).

Mathematical objects with diverse properties can be considered dynamical systems. For instance, state variables may take values from a continuous or discrete domain. Likewise, time may advance continuously or in discrete steps. Examples of dynamical systems characterized by different combinations of these features are listed in Table 1.

In simple cases, trajectories of dynamical systems may be expressed using mathematical formulas. For example, the ODE (ordinary differential equation) describing the motion of a mass on

TAB. 1 – Some formalisms used to specify dynamical systems according to the discrete or continuous nature of time and state variables.

C : continuous, D : discrete.	ODE	Iterated Mappings	Finite Automata
Time	C	D	D
State	C	C	D

a spring has an analytical solution expressed by a sine function (linear spring, in the absence of friction and damping). In more complex cases, analytic formulas representing trajectories of the system may not exist, and the behavior of the system is best studied using computer simulations.

By their nature, simulations operate in discrete time. Models initially formulated in terms of continuous time must therefore be discretized. Strategies for discretizing time in a manner leading to efficient simulations have extensively been studied in the scope of simulation theory, *e.g.* [Kre86].

Dynamical systems with apparently simple specifications may have very complex trajectories. This phenomenon is called *chaotic behavior*, *c.f.* [PJS92], and is relevant to biological systems, for example populations models [May75, May76].

2.2 Structured dynamical systems

Many biological systems are structured, which means that they can be decomposed into parts. The advancement of the state of the whole system is then viewed as the result of the advancement of the state of its parts. For example, the operation of a gene regulation network can be described in terms of the activities of individual genes.

Formally, we use the term *structured dynamical system* to denote a dynamical system divided into component subsystems (units). The set of state variables of the whole system is the Cartesian product of the sets of state variables of the component subsystems. Accordingly, the state transition function of the whole system can be described as the product of the state transition functions of these subsystems. Similarly to non-structured systems, structured dynamical systems can be defined assuming continuous or discrete state variables and time. In addition, the components can be arranged in a continuous or discrete manner in space. Some of the formalisms resulting from different combinations of these features are listed in Table 2.

TAB. 2 – Some formalisms used to specify structured dynamical systems according to the continuous or discrete nature of space, time, and state variables of the components. The heading "Numerical Solutions" refers to explicit numerical solutions of partial differential equations and systems of coupled ordinary differential equations.

C : continuous, D : discrete.	PDE	Coupled ODE	Numerical Solutions	Cellular Automata
Space	C	D	D	D
Time	C	C	D	D
States	C	C	C	D

Time management is an important issue in the modeling and simulation of structured systems [Lyn96]. For example, state transitions may occur *synchronously* (simultaneously in all components) or *asynchronously* (in one component at a time). Furthermore, efficient simulation techniques may assume different rates of time progression in different components [Jef85].

In many cases, the transition function of each subsystem depends only on a (small) subset of the state variables of the whole system. If the components of the system are discrete (i.e., excluding partial differential equations, or PDEs), these dependencies can be depicted as a *directed graph*, with the nodes representing the subsystems and the arrows indicating the inputs to each subsystem. We say that this graph defines the *topology* of the structured dynamical system, and call *neighbors* the pairs of subsystems (directly) connected by arrows.

The topology of a structured dynamical system may reflect its *spatial organization*, in the sense that only physically close subsystems are connected. A dynamical system with this property is said to be *locally* defined. Locality is an important feature of systems that model physical reality, because physical means of information exchange ultimately have a local character (e.g., transport of signaling molecules between neighboring cells). On the other hand, physically-based models need not to be rigorously local. For example, when modeling plants, it may be convenient to assume that higher branches cast shadow on lower branches without simulating the local mechanism of light propagation through space.

When the number of components in a structured dynamical systems is large, the exhaustive listing of all connections between the components becomes impractical or infeasible. This limitation can be overcome in several ways. For example, if the components are arranged in a regular pattern, the neighbors of each component need not to be listed explicitly. This is the case of *cellular automata* (*e.g.* [TM87], in which cells are arranged in a square grid). *Group-based fields* [GM01b] are a generalization of this idea, allowing for a wider range of connection patterns. Large structures can also be defined by *simulated development*, discussed next.

2.3 Dynamical systems with a dynamic structure

A developing multicellular organism can be viewed as a dynamical system in which not only the values of state variables, but also the *set* of state variables and the state transition function change over time. These phenomena can be captured using an extension of structured dynamic systems, in which the set of subsystems and/or the topology of their connections may dynamically change. We call these systems *dynamical systems with a dynamic structure* [GM01b], or $(\mathrm{DS})^2$-systems in short.

For example, let us consider a model of a multicellular organism, defined at the level of individual cells. When a cell divides, the subsystem that represents it is replaced by two subsystems that represent the daughter cells. Furthermore, the topology of the whole system is adjusted to :

- remove connections (neighborhood relations) between the mother cell and the rest of the organism,
- create connections between the daughter cells,
- insert connections between the daughter cells and the rest of the system.

These operations make it possible to gradually create a large network of interconnected cells.

2.4 A Taxonomy of Formalisms

From a computer science (or a mathematical) point of view, the problem raised by the simulation of dynamical systems with a dynamical structure is that of the programming paradigm (or the modeling language) well fitted to the specification of such systems. For instance, the PDE formalism is not a relevant solution because it prescribes an *a priori* given set of relations between an *a priori* given set of variables. Consequently, these two sets, which embed implicitly the structural

interaction between the entities or the system parts, cannot evolve jointly with the running state of the system [Mic96, pp 6, 85], [GM01b, chapter 1].

However, there exist several formalisms that can be used. The criteria used to classify the DS formalism in section 2.1 and 2.2 are still valid and the representation of time and state can be discrete or continuous for $(DS)^2$ as for standard DS. Here we propose an additional criterion to distinguish between the topological nature of the system structure. Table 3 presents some formalisms for the discrete time case.

TAB. 3 – Some formalisms used for the modeling of $(DS)^2$, according to the underlying topology of the state.

Topology	Multiset	Sequence	Uniform	Combinatorial
Formalism	multiset rewriting	L-systems	GBF	map L-systems, Graph-grammars, MTG, MGS

In this table, the first line gives the type of the topology used to connect the subcomponents of a system. In a *multiset*, all elements are considered to be connected to each other. In a *sequence*, elements are ordered linearly ; this case includes lists and extends also to tree-like structures. *Uniform* structures represents a regular neighborhood : for example, in a rectangular lattice (Von Neumann neighborhood), each element has exactly four neighbors. *Combinatorial* structures are used to define arbitrary connections between the components.

Considering solely the type of the topology underlying the structure of a state is only a partial caracterization that does not emphasize other several important points. Let us mention some of them.

- The relationship between the components can take place in an *a priori* structure. This approach is also known as the Newtonian conception of space where phenomena take place in a predefined scene. The other approach, which has been promoted by Leibniz, considers the topology as the result of the connection between the existing entities. In this point of view, the topology results from the dynamic connection between the system elements. This distinction is found in biology with the notions of *space oriented* or *structure oriented* models. For instance, accretive growth (growth on the boundaries) is an example of a space oriented process and intercalary growth (growth from the inside) is an example of a structure oriented process.
- There are several degrees in the dynamic of the structure. In the simplest case, the type of the topology remains the same during the evolutions of the system. An example is the growth of *Anabaena* filaments (Cf. section 4.2) where the system is always described as a sequence of cells. In addition, once a cell is connected with two neighbors, these connections remain the same. On the other hand, during the development of an embryo, several domains of cells change dramatically their shapes. For instance, the neural tube is formed dorsally in the embryonic development of Vertebrates by the joining of the 2 upturned neural folds formed by the edges of the ectodermal neural plate, giving rise to the brain and spinal nerve cord. In this process, which implies cell migration, the connections of a cell change over time and the global shape changes from a sheet to a tube.
- We have assumed that the interaction between the system parts can be described by a graph. Implicitly, this implies that elements interact two by two, which is not always the case. More elaborated interaction may imply more participants (e.g. a chemical reaction between two chemicals that requires also a catalyst ; or the many-to-one relation between a subsystem

and its decomposition). An interaction between n participants can be modeled by an n-edge in an *hypergraph*. An alternative representation is to use a more convienient mathematical notion developped in combinatorial algebraic topology : a n-simplex embedded in an *abstract simplicial complex* [GV01]. An abstract simplicial complex is a standard way to build a space by gluing elementary cells called simplices. From this point of view, graphs are the special case of one-dimensional spaces build by gluing vertices (simplices of dimension 0) and edges (simplices of dimension 1). In our case, the dimension of a simplex is directly linked with the number of participants in an elementary interaction. The complex represents the totality of the interactions.

– The notion of dimension also appears in the interactions between components in the following way. Often, the components of a system have a physical nature and the logical neighborhood established by the component interaction is the same as the spatial neighborhood implied by the physical structure of the system. For example, the topology implied by the representation of the cell sub-structures is tridimensional (compartments), bidimensional (membranes) and zero-dimensional (molecules). Obviously, the interactions that must be described depend of the dimension of the invoked entities : for instance, a flow of molecules can be conceived only through a membrane boundary between two compartments, not between a filament and another molecule ; conservation laws depend on the topological nature of the entities, etc. From this point of view, multiset corresponds to a trivial topology (two points are always neighbors), L-systems corresponds to one-dimensional topologies and a GBF described by n fundamental generators (cf. below, section 5) describe n-dimensional topologies.

2.5 Outline

Following table 3, the next sections and chapters presents some formalisms usable for $(\mathrm{DS})^2$ Modeling :

– Section 3 reviews the use of multisets to model biological state and multiset rewriting to specify the evolution function.
– Section 4 sketches the L-system formalism. This formalism is an effective approach for the modeling of linear and branching structure. For instance, it as largely been applied in the field of plant growing.
– Section 5 presents a general framework, instantiated in a programming language, that is able to unify several approaches by using a topological point of view.

3 Multiset Rewriting and the Modeling of Biological Systems

3.1 Basic Concepts

Consider a simple chemical system of two molecules types A and B. We suppose that only deterministic second-order catalytic reactions are allowed, that is : a collision of two molecules will catalyze the formation of a specific third molecule and the two colliding molecules are regarded as catalysts. The possible reaction rules are given explicitly as follows :

$$\begin{aligned} r_1 : &\quad A + A \longrightarrow A + A + B \\ r_2 : &\quad A + B \longrightarrow A + B + B \\ r_3 : &\quad B + B \longrightarrow B + B + A \end{aligned}$$

A simulation in which every molecule is explicitly stored and every single collision is explicitly performed can easily be implemented if the chemical reactor is abstracted as a *multiset*. Unlike a set, an element can occur several times in a multiset. In the following, we denote a multiset using

braces : $\{A, C, A, D, B, C\}$ is a multiset m with elements A and C occurring twice, and elements B and D occurring only one time. To simulate the chemical reaction, we simply interpret each rule as a transformation of the multiset. For instance, the rule r_1 specifies that two molecules A taken in the multiset have to be replaced by the three molecules A, A and B. For example, if reaction r_1 occurs in m at a given time step t_0, then m is transformed in $\{A, C, A, D, B, C, B\}$ (one additional B is produced). See figure 1.

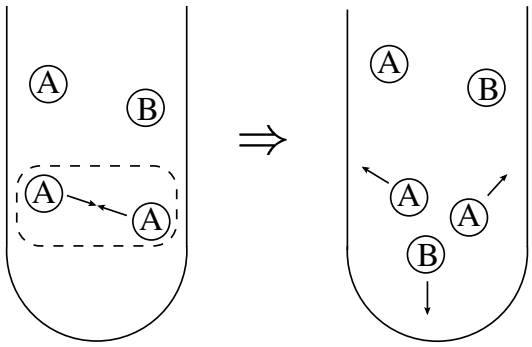


FIG. 1 – Illustration of one occurence of a reaction r_1 occuring in a test tube considered as a multiset of molecules.

Because several chemical reactions can occur *in parallel* (which means that several reactions involving different elements occur in the same time step), the strategy is to apply in parallel as many transformations as possible to the multiset. Such transformations are iterated to model the evolution of the state of the reactor. However, several competing rules may apply at the same time step : for instance consider a chemical reactor described by $\{A, A, B\}$ at time t_0 and subject to the two reactions r_1 and r_2. If r_1 occurs, then there is no longer A at t_0 to proceed with r_2 and vice-versa. The two reactions cannot occur together because there are not enough resources. In this case, we consider that one of the two rules is chosen in a non-deterministic manner. No assumption is made on the order on which the reactions occur.

The "+" sign that appears in the left and right hand side of the rules means that the linked molecules are present together in the chemical reactor. Thus, the left hand side of rule r_2 can also be equivalently written $B + A$. From a mathematical point of view, it is very convenient to consider $+$ as a formal commutative-associative operator used to construct multisets : a multiset $\{A, C, A, D, B, C\}$ is simply a formal sum $A + C + A + D + B + C$. The associativity and the commutativity properties are simply the expression that the elements of this last sum can be rearranged in any order. Then, rules like the r_i rules can be interpreted as rules for rewriting such formal expression. Abstractly, we can say that a chemical reaction can be modeled as a multiset rewriting system.

This modeling paradigm can be extended from this chemical example to other situations and its biological relevance is advocated in several recent papers [Man01, FMP00]. To quote[1] Fisher *et al.* [FMP00] : "A biological system is represented as a term of the form $t_1 + t_2 + \cdots + t_n$ where each term t_i represents either an entity or a message [or signal, command, information, action, etc.] addressed to an entity. [The simulation of the physical evolution of the biosystem] is achieved through term rewriting, where the left hand side of a rule typically matches an entity and a message addressed to it, and where the right hand side specifies the entity's updated state, and possibly other messages addressed to other entities. The operator $+$ that joins entities and messages is associative and commutative, achieving an ' associative commutative soup ', where entities swim around looking for messages addressed to them."

[1] with adaptations in the terminology, brackets are our comments

3.2 Division, Growth and Diffusion Processes

To illustrate this paradigm in a biological situation, we consider the multiplication of a mono-cellular organism in a test tube. A cell exists in one of two forms A or B. Type A and B can be used to characterize a phase of the life cycle of the cell, or as a cell polarity, etc. The division of a cell of type A produces one cell of type A and one of type B. In contrast, a cell of type B does not divide but evolves to give a cell of type A. This can be summarized by the two rules :

$$\begin{array}{lrcl} r_1: & A & \longrightarrow & A+B \\ r_2: & B & \longrightarrow & A \end{array}$$

Starting from a test tube with three initial cells, abstracted as a multiset $m_0 = \{A, B, B\}$, the first three evolutions are :

$$m_0 \to \{A, B, A, A\} \to \{A, B, A, B, A, B, A\} \to \{A, B, A, B, A, B, A, B, A, A, A\} \to \dots$$

There exists several software environments that support multiset rewriting (see next paragraph). So the previous two rules *directly turn to a computer program* that simulates the growing and division processes of this hypothetic mono-cellular organism. In fact, these rules fit well the development of Anabaena, which is described more in details in the next section, if we neglect the sequential organization of the cells. However, this model admit also other interpretations. For example, Fibonacci studied (in the year 1202) about how fast rabbits could breed under some ideal circumstances. Suppose a newly-born pair of rabbits, one male, one female, are put in a field. Rabbits are able to mate after one month so that at the end of its second month a female can produce another pair of rabbits. We simplify the model assuming that rabbits never die and that a female always produces one new pair (one male, one female) every month from the second month on. We model by symbol B a newly-born pair of rabbits and by symbol A a mature pair of rabbits. Then the rule r_1 expresses that a mature pair produces a newly-born pair and survive and rule r_2 specifies the maturation of a new pair.

The simulation of this process can be used to determine, for example, the relative ratio of A and B types in a population after some time. However, as mentioned in the introduction, the use of a formal model is not restricted to simulation and can be used to prove formal properties of the system without looking at the results of the simulation (e.g. : Fibonacci was able to prove that the ratio between B and A converges to the golden section as the time goes).

In the previous examples, each entity (a molecule, a cell or a pair of rabbits) is represented as an element of a multiset. In addition, the multiset structure allows objects to interact in a rather unstructured way, in the sense that an interaction between two objects is enabled simply by virtue of both being present in the multiset. In other word, there is no *localization* of the entities. Here is an example of another approach, where multiset rewriting is used in another way to take into account a geometric information. The problem is to model the diffusion of a set of particles on a line. The line is discretized as a sequence of small boxes, indexed by a natural integer, each containing zero or many particles. At each time step, a particle can choose to stay in the same box, or to jump to a neighboring box, with the same probability. See figure 2. The state of a particle is the index of the box where it resides. The entire state of the system is represented as a multiset of indices. The evolution of the system is then specified as three rules :

$$\begin{array}{lrcl} r_1: & n & \longrightarrow & n \\ r_2: & n & \longrightarrow & n-1 \\ r_3: & n & \longrightarrow & n+1 \end{array}$$

where n is an integer and the operations "$+$" and "$-$" that appear in the right hand side are the usual arithmetic operators. Rule r_1 specifies the behavior of a particle that stay in the same box ;

rule r_2 corresponds to a particle that jumps to the box at the left ; and rule r_3 defines a particle jumping to the right. Another solution is to factorize the three rules into one :

$$r: \quad n \longrightarrow n + \text{Random}(-1, 0, 1)$$

where the function Random$(\dots)$ returns randomly one of its arguments. In the case of three competing rules, we must assume that there is some fairness in the choice of the rules r_1 to r_3 to be applied, i.e., they have the same probability of being chosen. If there is more chance to stay in a box than to leave it, then the underlying formalism must be able to express some finer control over the rule application. As a matter of fact, specifying an application strategy of the rules that respect the symmetries of the system can be very difficult.

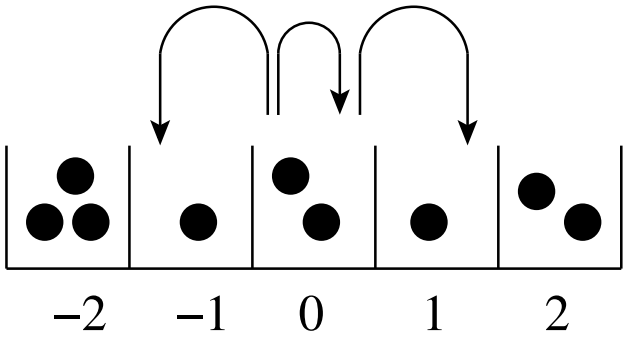


FIG. 2 – Diffusion of a particle along a line

3.3 Applications, Theories and Tools for Multiset Rewriting

Multiset rewriting has inspired several applications leading to the emergence of a new field : *Artificial Chemistry*. The home page [Dit00] and reference [DZB00] are a good introduction to this new area. There is a growing body of applications in artificial life, chemical and biological modeling, information processing and optimization. More specifically, Artificial Chemistry has been advocated as a productive framework for the study of pre-biotic and bio-chemical evolution, and for the study of the evolution of organization in general.

Multiset rewriting has also been used to extend other formalisms. For example, a multiset of L-systems is used to model an ecosystem (a multiset) of individual plants (modeled using L-system), see [LP02].

From the computer science point of view, the use of the chemical metaphor as a *computing model* has been investigated by Gamma [BM86, BCM87] in the middle of the eighties. A good review of the research done about Gamma can be found in [BFM01]. The previous examples of rules can be seen as simple Gamma programs. The CHemical Abstract Machine (CHAM) formalism extends these ideas with a focus on the expression of semantic of non deterministic processes [BB90]. The CHAM is an elaboration on the original Gamma formalism introducing the notion of sub-solution enclosed in a membrane. It is shown that models of algebraic process calculi can be defined in a very natural way using a CHAM : the fact that concurrency (between rule application) is a primitive built-in notion makes proof far easier than in the usual process semantics. The motivations of Gamma and the CHAM are the development of a formalism to support the specification and the programming of parallel and non deterministic programs. Multiset rewriting lies at the core of the formalism.

From the point of view of *term rewriting* [DJ90], multiset rewriting is the special case where the operators considered are both associative and commutative. In this domain, the perspective is

more logical and directed towards the concepts of rewriting calculus and rewriting logic. The applications considered are the design of theorem provers, logic programming languages, constraint solvers and decision procedures. Several frameworks provide efficient and expressive environments to apply rewrite rules following dedicated strategies. It is worth mentioning ELAN [ela02] and MAUDE [mau02].

At last but not least, in the domain of formal language theory and computational complexity, P systems [Pau98b, Pau00] are a new distributed parallel computing model based on the notion of a membrane structure. This paradigm extends standard multiset rewriting introducing the notion of membrane. A membrane structure is a nesting of compartments represented, e.g, by a Venn diagram without intersection and with a unique superset : the skin. Objects are placed in the regions defined by the membranes and evolve following various transformations : an object can evolve into another object, can pass through a membrane or dissolve its containing membrane. In the initial definition of the P systems, each region defined by a membrane corresponds to a multiset of atomic objects which can evolve following evolution rules very similar to Gamma's (the right hand side of each rule is augmented to specify the destination of the results of the reaction). The membrane structure enables the specification of some localization of the processes. For an example, see section 5. Several alternatives have been devised and a region can be equipped with various computational mechanisms : string rewriting, splicing systems (DNA computing), etc. From the calculability point of view, several variants of such computing devices can compute all recursively enumerable sets of natural numbers. When an enhanced parallelism is provided, by means of membrane division (and, in certain variants where one works with string-objects, by means of object replication), NP-complete problems[2] can be solved in linear time (of course, making use of an exponential space).

4 L-systems

4.1 Basic notions

L-systems were introduced in 1968 in the landmark paper by A. Lindenmayer, *Mathematical models for cellular interaction in development* [Lin68]. They provide a well developed and flexible

[2] A short introduction for biologists to the notion of NP problems : Computer scientists classify the complexity of solving a problem by an algorithm. This classification obviously depends of the kind of device that runs the algorithm. All our actual digital computers behave in the same way w.r.t to this classification and are abstracted by a formal device called a *Turing Machine* (TM).

A decision problem is a problem that is solved by a *yes* or a *no* answer. There are many known decision problems for which no-one has been able to solve with a polynomial time algorithm on a TM, that is, they cannot be solved in a number of steps that is polynomial with the size of the input. Such problem are difficult to solve on a current computer and we are limited to handle only very small problem size.

For some such problems it has been possible to show that a non-deterministic Turing machine (NDTM from now on) could solve them in polynomial time. We can understand this intuitively as follows. If we think of decision problems as all being questions of the type, "does there exist an object with property Y within the set of objects X" then a non-deterministic algorithm is one that, given a guess of a candidate object $x \in X$, can answer correctly "yes" or "no" to the question of whether or not x has property Y. Such an algorithm can only answer "yes" in polynomial time if we happen to correctly guess an $x \in X$ with property Y. A non-deterministic Turing machine can be intuitively understood to always correctly guess an $x \in X$ with property Y, if such an x exists. The class NP (*Non-deterministic polynomial time*) is the class of decision problems that can be solved in polynomial time by a NDTM. A problem is NP-hard if solving it in polynomial time would make it possible to solve all problems in class NP in polynomial time. A NP-complete problem is NP problem which is also NP-hard : thus a solution for one NP-complete problem would solve all problems in NP. Many (but not all) naturally arising problems in class NP are in fact NP-complete. One of the first problem ever shown to be NP-complete was the Hamilton's problem in graph theory posed by William Hamilton : given a graph, is there a path through the graph which visits each vertex precisely once (a "Hamiltonian path") ?

A big question in theoretical computer science is to establish or to infirm the equation $P = NP$ stating the equality of the class NP and the class P of problems that can be solved in polynomial time on a classical deterministic TM.

tool for modeling and simulating a restricted but biologically important class of dynamic systems with a dynamic structure : linear and branching structures.

Originally, Lindenmayer described his formalism in terms of cellular automata, in which — in contrast to the standard definition — the cells could divide. Subsequently he observed that L-systems can be formulated in a simpler and more elegant manner in terms of formal language theory [Lin71]. That theory was originally proposed by Chomsky [Cho56, Cho57] to describe the syntax of natural languages. Its fundamental notion is that of a (generative) *grammar*, which consists of *productions* or *rewriting rules*. In general, a production replaces a symbol by zero, one, or several new symbols. They may represent words in a sentence, as in the original interpretation by Chomsky, but they also may represent cells or other components of a living organism, as was proposed by Lindenmayer. The use of related formalisms in the description of such apparently distant notions as languages and biological structures may seem surprising at first. In fact, it reflect the common dynamic nature of sentences under construction and developing organisms.

Applications of L-systems to modeling have an extensive literature, last reviewed in [Pru98] and [Pru99]. Below we outline one variant, called *parametric L-systems* [Han92, PH90, PL90] Within this formalism, the individual subsystems are called *modules*. Each module is represented by a symbol (letter) with optional parameters. This letter and parameters jointly characterize the module's state. For instance, the letter may represent a cell type, while the parameters may represent quantitative attributes of the cell, such as its dimensions and concentrations of chemicals that it contains.

The assumption that the organism forms a filament makes it possible to represent it at any moment of time as a string of modules, called a *parametric word*. For example, the string

$$A(2.5)B(3.14, 0.2)CA(1.3) \tag{1}$$

may represent an organism that consists of four cells. The first cell has type A and is characterized by one parameter, the value of which is equal to 2.5. The remaining symbols have an analogous interpretation.

An L-system model describes the development of the entire structure by operating on individual modules. A production specifies the fate of a unit over a given time interval as a function of its current state and, optionally, the states of its neighbors. For example, the production

$$A(x) < B(y, z) > C \rightarrow CB(x + y, z/2) \tag{2}$$

operates on a module B that appears in the *context* of a module A to its left and module C to its right. The left and right contexts are separated from the *strict predecessor* B by the metasymbols (i.e., the symbols that do not represent modules) $<$ and $>$, respectively. In this example, module B divides into a module C and a new module B. The arithmetic expressions in the production's successor determine new parameter values. Hence, when applied to string (1), production (2) will yield the string

$$A(2.5)CB(5.64, 0.1)CA(1.3). \tag{3}$$

Simultaneous application of productions to all modules advances the state of the whole structure. If the set of module types is finite, the corresponding finite set of productions provides a mechanism for advancing the state of the entire structure independently of its size (the number of modules).

4.2 A sample model

We will illustrate the notion of genetic L-systems by constructing a model of heterocyst differentiation in a growing filament of the cyanobacterium *Anabaena*. The following description is adapted from [HP96].

The cells of *Anabaena* are organized into filaments which consist of sequences of *vegetative cells* separated by *heterocysts*. The vegetative cells divide into two cells of unequal length and, in some cases, differentiate into heterocysts which do not further divide. The organism maintains an approximately constant spacing between heterocysts : whenever the distance between two heterocysts becomes too large due to the division and elongation of vegetative cells, a new heterocyst emerges.

What mechanisms is responsible for the differentiation of heterocysts and the maintenance of the approximately constant spacing between them ? Baker and Herman [BH70, BH72] (see also [dL87, HR75, Lin74] proposed the following simulation model. The heterocysts produce a substance that diffuses along the filament and is used by the vegetative cells. This substance inhibits the differentiation of vegetative cells into heterocysts. When its level in a cell drops below a threshold value, the cell detects that it is no longer inhibited and differentiates into a heterocyst.

Although the model of Baker and Herman is capable of reproducing the observed pattern of heterocyst spacing, it is very sensitive to parameter values. Small changes in these values easily result in filaments with pairs of heterocysts appearing almost simultaneously, close to each other. This is not surprising, considering the operation of the model. The gradient of the concentration of the inhibitor may be too small near the middle of a sequence of vegetative cells to precisely define the point in which a new heterocyst should differentiate. Consequently, the threshold value may be reached almost simultaneously by several neighboring cells, resulting in the differentiation of two or more heterocysts close to each other.

The above model can be improved assuming that the prospective heterocysts compete until one "wins" and suppresses the differentiation of its neighbors. This "interactive" model was originally proposed by Wilcox *et al* [WMS73]. It can be formalized using the framework of the *activator-inhibitor* class of reaction-diffusion models [Mei82]. In addition to the substance that inhibits the differentiation, the cells are assumed to carry a substance called the activator. The concentration of the activator is the criterion that distinguishes the vegetative cells (low concentration) from the heterocysts (high concentration). The activator and inhibitor are antagonistic substances : the production of the activator is suppressed by the inhibitor unless the concentration of the inhibitor is low. In that case, production of the activator drastically increases through an autocatalytic process (an increased concentration of the activator promotes its own further production). High concentration of the activator also promotes the production of the inhibitor, which diffuses to the neighboring cells. This establishes a ground for competition in which activator-producing cells attempt to suppress production of the activator in the neighboring cells. For proper values of parameters that control this process, only individual, widely spaced cells are able to maintain the high-activation state.

An L-system implementation of these mechanisms (a variant of the L-system from [HP96]) is given below.

$$
\begin{array}{ll}
\omega : & M(0.5, 0.1, 200, \text{right})M(0.5, 0.1, 100, \text{right})M(0.5, 0.1, 100, \text{right}) \\
p_1 : & M(s_l, a_l, h_l, p_l) < M(s, a, h, p) > M(s_r, a_r, h_r, p_r) : \\
 & \quad s < s_{max} \;\&\; a < a_{th} \rightarrow M(s', a', h', p) \\
p_2 : & M(s_l, a_l, h_l, p_l) < M(s, a, h, p) > M(s_r, a_r, h_r, p_r) : \\
 & \quad s \geq s_{max} \;\&\; a < a_{th} \;\&\; p = \text{left} \rightarrow \\
 & \quad M(ks', a', h', \text{left})M((1-k)s', a', h', \text{right}) \\
p_3 : & M(s_l, a_l, h_l, p_l) < M(s, a, h, p) > M(s_r, a_r, h_r, p_r) : \\
 & \quad s \geq s_{max} \;\&\; a < a_{th} \;\&\; p = \text{right} \rightarrow \\
 & \quad M((1-k)s', a', h', \text{left})M(ks', a', h', \text{right}) \\
p_4 : & M(s_l, a_l, h_l, p_l) < M(s, a, h, p) > M(s_r, a_r, h_r, p_r) : \\
 & \quad a \geq a_{th} \rightarrow M(s, a', h', p)
\end{array}
$$

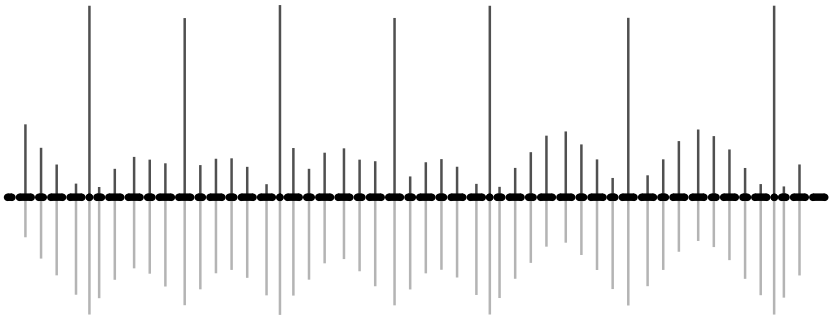


FIG. 3 – Fragment of a simulated filament of *Anabaena*. Vertical lines indicate the concentrations of the activator and inhibitor (above and below the cells, respectively). Notice the sharp peaks of the activator concentration that define the heterocysts, and high levels of the inhibitor concentration in the neighboring vegetative, which prevent their differentiation. The parameters used in the simulation were : $\rho = 3$, $\kappa = 0.001$, $a_0 = 0.01$, $\mu = 0.1$, $h_0 = 0.001$, $\nu = 0.45$, $D_h = 0.004$, $a_{th} = 1$, $k = 0.38196$, $s_{max} = 1$, $r = 0.002$, and $w = 0.001$.

where

$$\begin{aligned} s' &= s(1 + r\Delta t), \\ a' &= a + \left(\frac{\rho}{h}\left(\frac{a^2}{1+\kappa a^2} + a_0\right) - \mu a\right)\Delta t, \\ h' &= h + \left(\rho\left(\frac{a^2}{1+\kappa a^2} + h_0\right) - \nu h + D_h \frac{h_l + h_r - h}{sw}\right)\Delta t. \end{aligned}$$

The cells are specified as modules M, where parameter s stands for cell length, a is the concentration of the activator, h is the concentration of the inhibitor, and p denotes polarity, which plays a role during cell division. All productions are context-sensitive to capture diffusion of the activator and inhibitor. It is assumed that the main barrier for the diffusion are cell walls of width w. Production p_1 characterizes growth of vegetative cells ($a < a_{th}$), controlled by the growth rate r. A cell that reaches the maximum length of s_{max} divides into two unequal daughter cells, with the lengths controlled by constant $k < 0.5$. The respective positions of the longer and shorter cells depends on the polarity p of the mother cell, as described by productions p_2 and p_3. Increase of the concentration of the activator a to or above the threshold value a_{th} indicates the emergence of a heterocyst. According to production p_4, a heterocyst does not further elongate or divide. The equations for s', a', and h' govern the exponential elongation of the cells and the activator-inhibitor interactions [Mei82].

The operation of the model is illustrated in Figure 3. The vertical lines indicate the concentrations of the activator (above the filament) and inhibitor (below the filament) associated with each cell.

It is interesting from the historical perspective that the interactive model of Wilcox *et al.* [WMS73] and its subsequent L-system implementation [HP96] predicted the essential structure of the gene regulation network that controls the development of *Anabaena* filaments in nature [Ada00]. The activator corresponds to the protein HetR, which plays a key role in the maintenance of the heterocyst state, whereas the inhibitor corresponds to the protein PatS (or a fragment of it), which diffuses across the filament and maintains the spacing between the heterocysts. The character of interactions captured by the simulation model is consistent with the postulated structure of the gene regulation network, in which HetR upregulates its own production as well as the production of PatS, whereas PatS downregulates production of HetR.

We believe that models of similar nature, integrating the action of genes into developmental models of multicellular structures, will become more widely used in the future, offering insights

into developmental processes that are difficult to obtain through observations and qualitative reasoning alone.

5 The MGS Approach

5.1 Motivations and Background

The previous examples of formalisms do not fully address issues of structural interactions between entities or system parts because of the *lack of topological organization*. The need to represent more structured organizations (than sequence or multiset) of entities and their interactions has been already stressed [FMP00] and motivates several extensions of rewriting (see for one example amongst others [BH00]). However, a general drawback with these extensions is that they work with a fixed topology of entities, and it is not obvious at all how to extend this to systems where the relationships between entities are drastically changing. This is precisely one of the main motivations of the MGS research project[3].

MGS is aimed at the representation and manipulation of local transformations of entities structured by *abstract topologies* [GM01b, GM02]. A set of entities organized by an abstract topology is called a *topological collection*. Topological means here that each collection type defines a neighborhood relation specifying both the notion of *locality* and the notion of *sub-collection*. The collection types can range in MGS from totally unstructured with sets and multisets to more structured with sequences and GBFs [GMS95, Mic96, GM01a] (other topologies are currently under development and include Voronoï partitions and arbitrary combinatorial neighborhoods).

The *global transformation* of a topological collection C consists in the parallel application of a set of *local transformations*. A local transformation is specified by a rewriting rule r that specifies the change of a sub-collection. A rewrite rule r :

1. selects a sub-collection A in C,
2. computes a new collection B as a function f of A and its neighbors,
3. and specifies the insertion of B in place of A into C.

These steps are summarized in figures 4 and 5. The topology of B depends on f and can be different from the topology of A. For example, a set in a sequence can be replaced by a sequence. Moreover, the topological structure of C can be changed through the application of transformations. These features enables the modeling of $(\text{DS})^2$: states of a DS are represented by collections and transformations are used to model transition functions on these structured states.

As a programming language based on topological concepts, MGS integrates the idea of topological collections and their transformations into a general high-level functional programming language : topological collections are just new kinds of values and transformations are functions acting on collections. The approach is purely declarative : operators acting on values combine values to give new values, they do not act by side-effect.

5.2 Biological Examples in MGS

In this subsection, we sketch several examples in various domains to exemplify the versatility of the MGS formalism.

[3] MGS is the acronym of "*(encore) un Modèle Géneral de Simulation (de système dynamique)*" (yet another General Model for the Simulation of dynamical systems). The MGS home page is located at url www.lami.univ-evry.fr/mgs where additional informations are available.

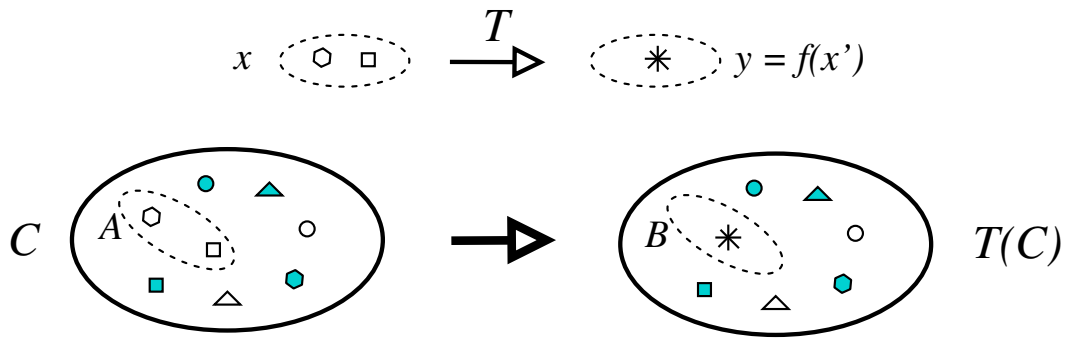


FIG. 4 – A basic transformation of a topological collection. Collection C is of some kind (set, sequence, array, cyclic grid, tree, term, etc). A rule T specifies that a sub-collection A of C has to be substituted by a collection B computed from A. The right hand side of the rule is computed from the sub-collection matched by the left hand side x and its possible neighbors x' in the collection C.

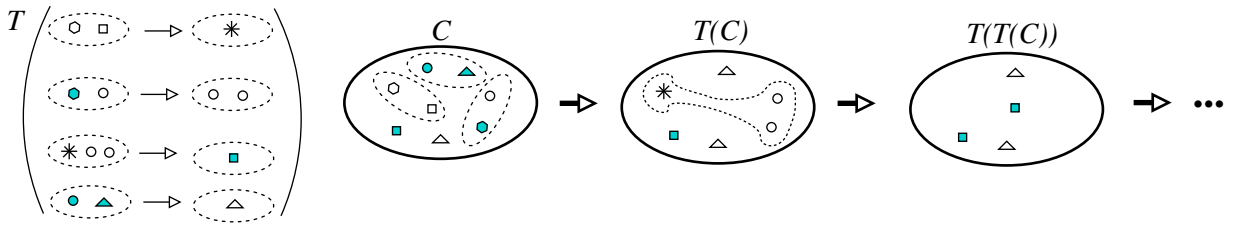


FIG. 5 – Transformation and iteration of a transformation. A transformation T is a set of basic transformations applied synchronously to make one evolution step. The basic transformations do not interact together. A transformation is then iterated to build the successive states of the system.

The Eden Model

We start with a simple model of growth sometimes called the Eden model (specifically, a type B Eden model) [Ede58]. The model has been used since the 1960's as a model for such things as tumor growth and growth of cities. In this model, a 2D space is partitioned in empty or occupied cells (we use the white-space character and the `C` letter). We start with only one occupied cell. At each step, occupied cells with an empty neighbor are selected, and the corresponding empty cell is made occupied.

The corresponding `MGS` model starts by defining the 2D partition using a *group based field* (GBF in short). A GBF is an extension of the notion of array, where the elements are indexed by the elements of a group, called the *shape* of the GBF [GMS95, GM01a]. This kind of collection can be used to describe uniform and regular topologies. For example :

```
gbf Grid2 = < north, east >
```

defines a shape called *Grid2*, corresponding to the Von Neuman neighborhood in a classical array (a cell above, below, left or right – not diagonal). The two names `north` and `east` refer to the directions that can be followed to reach the neighbors of an element. These directions are the *generators* of the underlying group structure. The list of the generators can be completed by giving equations that constraint the displacement in the shape :

```
gbf Hexagon = < east, north, northeast;
                east + north = northeast >
```

defines an hexagonal lattice that tiles the plane, see. figure 6. Each cell has six neighbors (following

the three generators and their inverses). The equation `east + north = northeast` specifies that a move following `northeast` is the same as a move to `east` followed by a move to `north`.

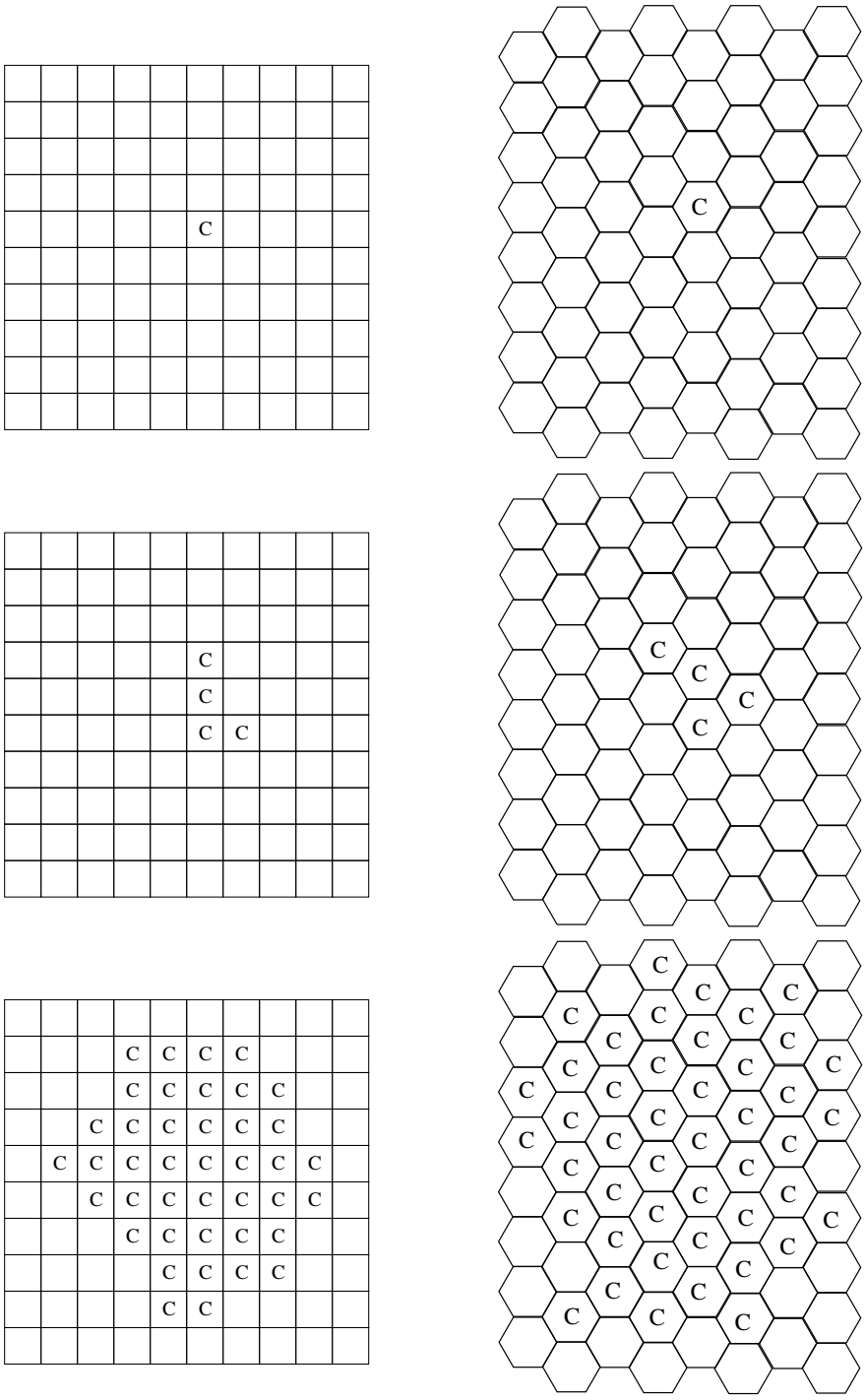


FIG. 6 – Eden's model on a grid and on an hexagonal mesh (initial state, and states after the 3 and the 7 time steps). The *same* transformation is used for both cases.

The Eden's aggregation process is simply described as the following transformation :

```
trans Eden = {
    x,y / (x = "C") & (y = " ") ⇒ x,"C";
}
```

the keyword `trans` introduce the rules of a transformation. A rule takes the following form :

$$pattern \Rightarrow expression$$

where *pattern* in the left hand side of the rule matches a sub-collection A of the collection C on which the transformation is applied. The sub-collection A is substituted in C by the collection B computed by the *expression* in the right hand side of the rule. Here, the pattern "x`,`y" filters an element y neighbor of an element x such that the value of x is occupied and the value of y is empty. The conditions on the elements matched are given by the expression after the "`/`" operator and the comma operator "`,`" means that x and y must be neighbors. The right hand side specifies that the couple x`,`y matched by the left hand side must be replaced by a couple x`,"C"`.

The transformation *Eden* defines a function that can then be applied to compute the evolution of some initial state. One of the advantages of the `MGS` approach, is that this transformation can apply indifferently on grid or hexagonal lattices (or *any* other collection kind). The meaning of the neighborhood operator "`,`" in the pattern of a rule depends on the collection on which the transformation is applied.

It is interesting to compare transformations on GBFs with the genuine cellular automata (CA) formalism (see the corresponding chapter). There are several differences. The notion of GBF extends the usual square grid of CA to more general Cayley graphs. The pattern in a rule may match arbitrary domain, not only one cell as it is usually the case for CA. Moreover, the value of a cell can be arbitrary complex (even another GBF) and is not restricted to take a value in a finite set.

Restriction Enzymes

This example shows the ability to nest different topologies to achieve the modeling of a biological organization. We want to represent the action of a set of restriction enzymes on the DNA. The DNA structure is simplified as a sequence of letters `A`, `C`, `T` and `G`. The DNA strings are collected in a multiset. Thus we have to manipulate a multiset of sequences (this kind of nested structures has been proved useful in other areas, e.g. [LP02]).

A restriction enzyme is represented as a rule that splits the DNA strings ; for instance a rule like :

```
EcoRI =  x+ as X,
         (cut+ as CUT / CUT = "G","A","A","T","T","C"),
         y+ as Y
      ⇒ (X, "G") : :("A","A","T","T","C", Y) : :seq :()
```

corresponds to the *EcoRI* restriction enzyme with recognition sequence `GˆAATTC` (the point of cleavage is marked with `ˆ`). The x+ pattern filters the part of the DNA string before the recognition sequence and the result is named X (the + operator denotes repetition of neighbors). Identically, Y names the part of the string after the recognition sequence. The right hand side of the rule constructs the two resulting parts as a sequence of two sequences (the `: :` operator indicates the construction of a nested sequence).

We assume that all restrictions enzyme rules are collected into one transformation. We need an additional rule, called `Void` for specifying that a DNA string without recognition sequence must be inserted as such :

```
trans Restriction = {
    EcoRI = ...;
    ...;
    Void = x+ as X ={flat=false}=> X
}
```

The attribute "`flat=false`" in the body of the arrow of rule `Void` indicates that the X (which is a sequence) must be inserted in the resulting multiset as one single entity. This contrasts with the rule `EcoRi` whose right hand side computes a sequence of elements to be inserted in the enclosing multiset.

The transformation *Restriction* can then be applied to the DNA strings floating in a multiset using the simple transformation :

```
trans Apply = { dna ⇒ Restriction(dna) }
```

A Localized Signaling Network

At last but not least, we want to sketch the modeling of a spatially distributed biochemical network in `MGS`. We rely on a model proposed by A. E. Bugrim [Bug00]. The example focuses on a small signaling network that consists of cAMP and calcium signaling. See figure 7 for a more complete description.

The corresponding topological structure mimics the spatial organization of the cell using nested multisets, see figure 8. The `MGS` declarations :

```
collection Volume = bag;
collection Membrane = bag;

collection Environment = Volume;
collection Plasma = Membrane;
collection Cytosol = Volume;
collection EndoRetic = Membrane;
```

are used to introduce some new kinds of multisets (the `bag` keyword). This kinds are used here mainly do describe the hierarchy of localization and compartments and can be used, if necessary, to discriminate between multisets.

The main part of the corresponding `MGS` program consists in defining the ontology of this application domain : there exists several molecules, each have a name ; some exists in two state : active or inactive ; some are characterized as receptors ; etc. Such ontology is described in `MGS` using *subtyping*. These subtypes are then used in pattern-matching to select entities with or without some properties. For example, a molecule is described as a record having or not some fields. Record type in `MGS` may specify the presence or the absence of a field, or the value of a specific field. For instance :

```
state Molecule = {name};
state Activity = {activation};
state Activated = {activation = 1};
state Inactivated = {activation = 0};
state ATP = Molecule + {name = "atp"};
```

define five record types. The record type declaration is introduced by the keyword `state`. *Molecule* is the type of any record having at least a field named `name`. *Activated* is the type of a record having at least a field named `activation` and with value `1`. This type is a subtype of *Activity* which only requires the presence of the field `activation`. The type *ATP* corresponds to a molecule named `"atp"`.

Three kinds of transformations are used to define the processes of the Bugrim's model. The first class corresponds to some ancillary transformations. For example

```
trans ActivateReceptor = { r :Receptor → r + {activation=1} }
```

is a rule that updates to `1` the field `activation` of an entity r of type *Receptor*. This kind of transformations is triggered by a rule of the sole transformation of the second class. This transformation summarize all the rule corresponding of the description of the biochemistry (they are about 10 reactions in this pathway) :

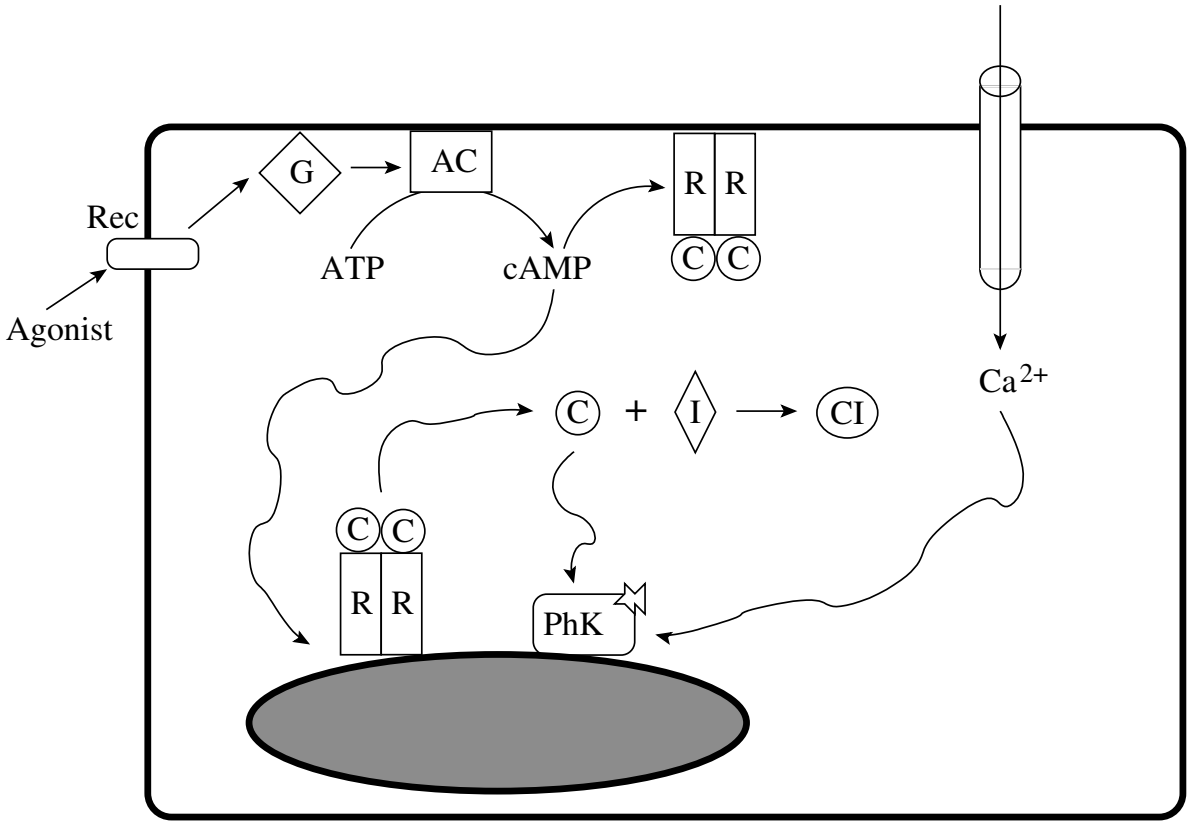


FIG. 7 – *cAMP and calcium signaling pathways* (this schema is reprinted from [Bug00]). The different components of the two pathways are localized at various places within the cell.
The first steps of the cAMP pathway occur at the plasma membrane, starting with the activation of adrenegric receptors. Then, the cAMP molecules bind to a regulatory sub-unit of the protein kinase A, with the effect of dissociating a catalytic sub-unit C. The localization of PKA depends of a family of anchoring proteins AKAPs that target this kinase to different compartments. In this example, two localizations are considered : the plasma membrane and an internal compartment (e.g., nucleus or ER).
The calcium pathway starts by the activation of a channel in the plasma membrane. The fraction of PhK associated to the internal compartment is the target of both pathways. A possible inhibitor I of PKA is also considered.

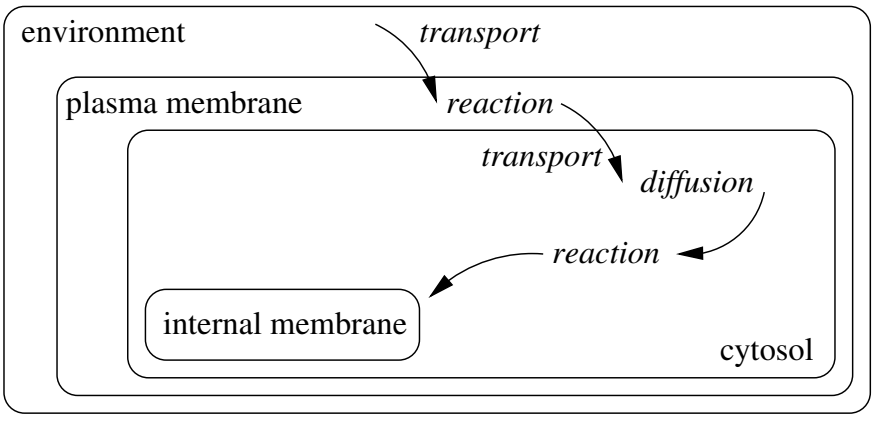


FIG. 8 – The reaction, diffusion and transport processes described in figure 7 are modeled as multiset transformations taking place in a nest of multisets. This is reminiscent of the P system approach, see section 3.

```
trans Biochemistry = {
    R1 = a :ActiveAgonist, p :Plasma
       ⇒ a+{activation=0},ActivateReceptor(p);
    ...
}
```

For example, rule `R1` specifies that an active agonist and a plasma membrane interact to inactivate the agonist and to transform the plasma with transformation *ActivateReceptor* (this transformation turn on all the activation fields of the receptors anchored in the plasma membrane).

There is also only one transformation in the last class of transformations. It is used to thread the biochemistry rules amongst the nested multisets :

```
fun Run(x) = Thread(Biochemistry(x));
trans Thread = {
    p :Membrane ⇒ Run(p);
    c :Volume ⇒ Run(c);
}
```

The transformation *Thread* applies the function *Run* to each entity of type *Membrane* or *Volume* found in the collection argument. The function *Run* consists in running the biochemistry transformation and then iterating the threading.

The complete `MGS` program is approximatively 150 line long, including the building of the initial system state. It describes 40 molecules in diverse states, uses of 5 auxiliary transformation to define 10 chemical interactions.

6 Multiscale graphs

The previous formalisms have been used to model the changes of structure that arise throughout time. However, biological structures may change also due to a change in the scale of observations.

On the one hand, plants appear as complex structures due to the intrication of many substructures at various levels of detail. On the other hand, plants are essentially spatially and temporally periodic structures which gives an overall impression of simplicity. In such a paradoxical situation, the question arises : what mathematical formalisms and what tools are necessary to model plants at several scales ?

In this chapter, we analyse how biological systems, such as plants, can be formally represented with combinatorial formalisms (see section 2). We particularly analyze how this formalism must be designed in order to account for a new dimension, namely the scale dimension. We then briefly describe the types of mathematical and computational tools that must be developed in this context.

6.1 Plants as modular organisms

The growth of a plant can be depicted as the result of two growth processes. This *apical growth process* gives the plant the ability to develop in one direction. During their activity, shoot meristems can give birth to distinct embryogenic cellular areas (always associated with corresponding leaves), called axillary or lateral meristems. This defines the *branching process*. Plants make branching structures if the meristems located at leaf axils enter an apical growth process. Using the branching process, plants can develop shoots in more than one direction. The overall *growth process* is thus the combination of both the apical growth process and the branching process. Growth is a fundamentally repetitive process which creates various forms of patterns repeated as ”modules” throughout the plant structure ([HRW86], [Bar91]). Figure 9 illustrates different types of modules that can be observed on plants.

For a given type of module, the plant can be split-up into a set of modules of this type. This defines a particular plant modularity. A plant modularity, is caracterised by the type of modules

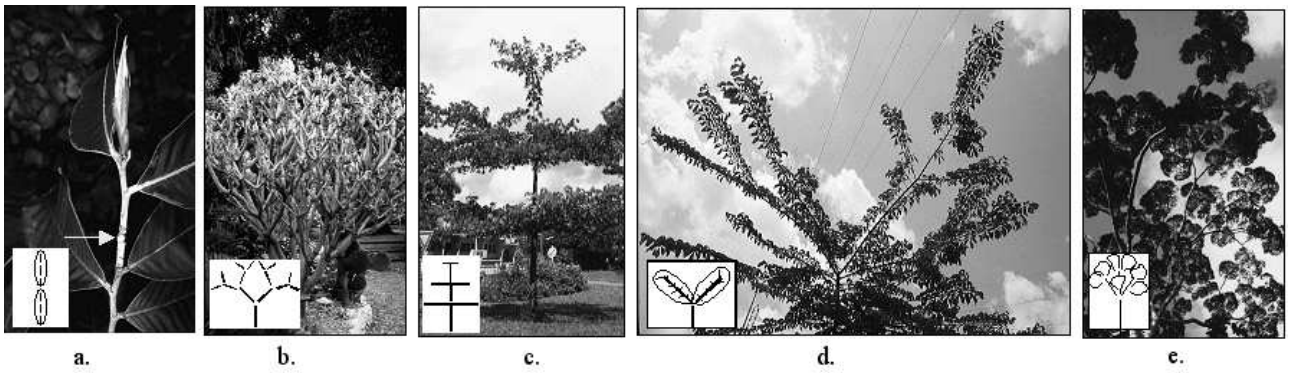


FIG. 9 – Different types of modularity in plants. a. nodes b. axes c. whorls d. branching systems e. crownlets

considered and their adjacency within the plant. This information can be represented by a *directed graph*.

A directed graph is defined by a set of objects, called vertices, and a binary relation between these vertices. The binary relation defines a set of ordered pair of vertices, called edges. In plant representations, vertices represent botanical entities and edges adjacency between these entities. Edges are always directed from oldest entities to youngest ones. Given an edge (a, b), we say that a is a *father* of b and b is a *son* of a. Directed graphs representing plants have tree-like structures : every vertex, except one, called the root, has exactly one father vertex. Morevover, in order to identify the different axes of a given plant, two types of connections are distinguished : an entity can either precede (type '$<$') or bear (type '+') another entity (Figure 10). In order to describe different characteristics of plant entities, vertices can have attributes, e.g. length, diameter, spatial location, leaf area, number of flowers, type of branched entities, *etc*.

6.2 Multiscale representations

Many modularities can exist on a single individual. Several types of modularity, stemming from either natural or artificial decomposition of the plant into modules, can exist within a single individual at the same time. For above-ground systems, at least the nodal (the plant is a set of leaves) and the axial modularity (the plant is a set of axes) coexist. If, in addition, the plant reiterates, a modularity by reiteration is superimposed on the previous ones. Thus, there always exist two or three types of modularities expressed in a plant simultaneously. There can be more, depending on the number of regular fluctuations that characterize the plant growth. This is the case, for example, for plants containing growth unit or annual shoot modules. These types of module can exist simultaneously in a plant, such as in apricot tree, evergreen oak or Aleppo pine. For a single plant, there is thus the theoretical possibility of finding numerous types of modularity, each one corresponding to a particular topological interpretation of the plant.

The existence of several modularities on the same plant can be illustrated by *Vochysia guyanensis* [San92]. For this plant, the number of modularities stemming from natural decomposition is relatively high. The highest scale corresponds to the description of the topological structure in terms of internodes. At a lower scale, the rhythmic elongation of stems produces an alternate sequence of cataphylls and developed leaves which enables the observer to define growth unit modules (11.a). The final stopping of stem elongation, due to the death of their apical meristem, makes it possible to group growth units into axes (11.a). The architectural unit of the young tree consists of a stack of such axes (11.b). The plant continues its development by reiterating its architectural unit. The resulting topological structure is described in terms of reiterated complexes. Eventually, at the lowest scale, the crown of the adult tree is a set of crownlets, each of them made of reiterated

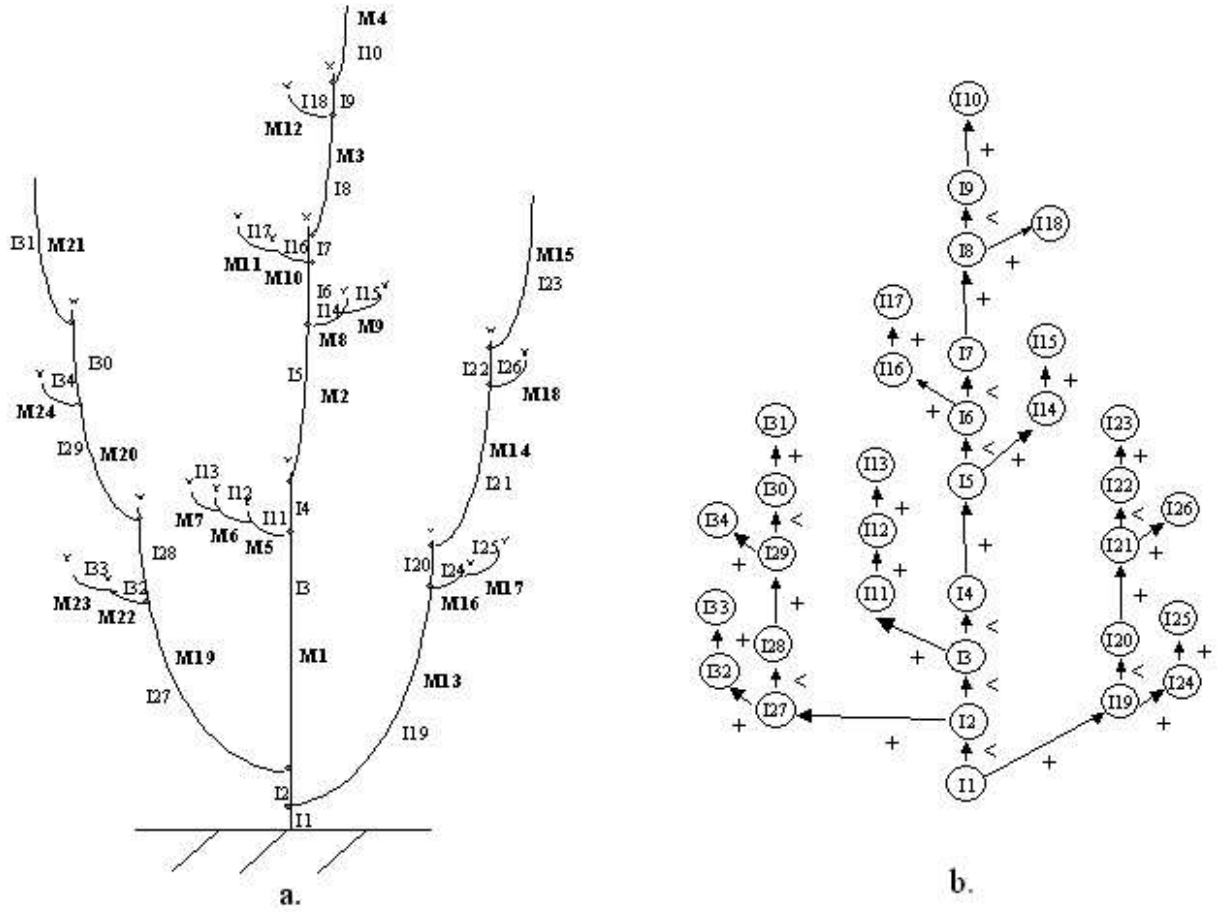


FIG. 10 – a. A tree b. The tree graph representation of its topology (at node scale)

complexes (11.c). The plant can thus be represented by a specific topological structure for each possible scale. The set of these topological structures defined at every scale and their relations characterizes the overall topological structure of the plant, *i.e. multiscale topological structure* of the plant.

To formally represent the multi-modular structure of plants, extension od directed graphs, called multiscale tree graphs (MTGs) [GC98], are used. The MTG formalism has been designed in order to enable users to express both the modularity and the multiscale nature of plant structures. Each scale of analysis corresponds to a modular structure which can be formally represented by a tree graph. Entities at one scale are decomposed into entities at finer scales. For instance, internodes of Figure 10.a can be grouped into growth units, leading to a more macroscopic description of the plant topology (Figure 12).

A MTG integrates in a homogeneous framework the different tree graphs corresponding to plant descriptions at different scales (Figure 13.a). Vertices at one scale are composed of vertices at a higher scale. If an entity a is composed of n entities $x_1, x_2, ..., x_n$, for every $i \in [1, n]$, a is called the *complex* of x_i, and x_i is a *component* of a. The complex of any entity x_i is denoted $\pi(x_i)$. If the scale of a is defined by the integer s, then for every $i \in [1, n]$, the scale of x_i is $s + 1$. The most macroscopic scale s_0 consists of a single vertex, representing the entire plant, and by convention has value 0. In order to maintain coherence between the different tree graph representations of a same individual, MTGs must respect the following consistency constraint : if there exists an edge (x, y) in the tree graph representing the plant structure at scale $s+1$, and if the complexes of x and y are different, then there necessarily exists a corresponding edge $(\pi(x), \pi(y))$ between these complexes in the tree graph representing the plant at scale s (Figure 13.b) This

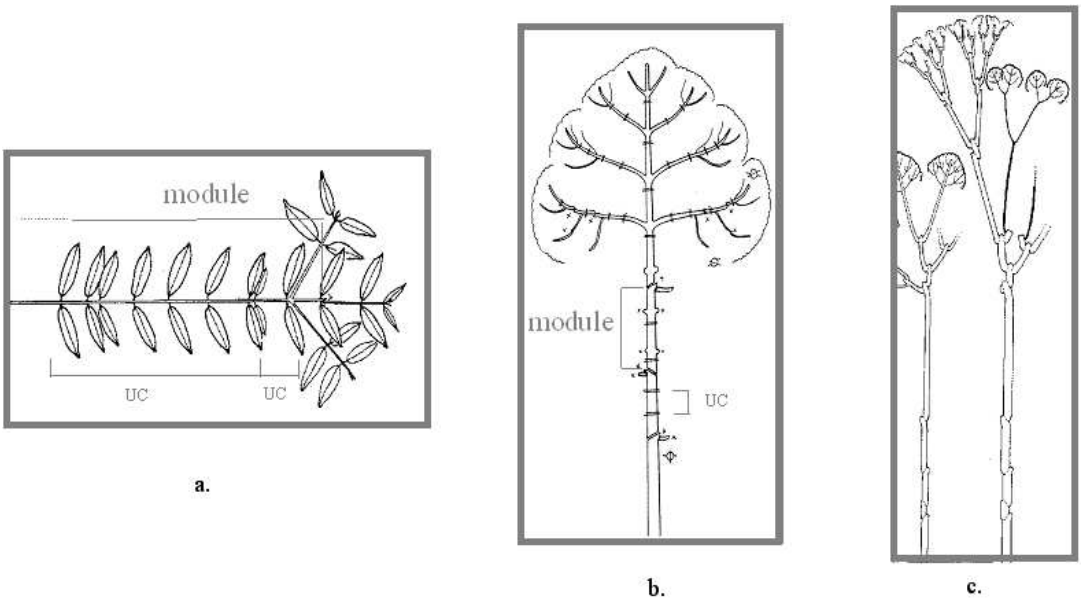


FIG. 11 – Nested modularities : a. nodes, growth units and axes. b. Architectural unit c. crowlets.

expresses that the connection between two macroentities results from the connection between two of their components.

6.3 Space of modularities

From a structural point of view, the relative position of two modularities in a plant can be of two types.

- Firstly, one modularity is a refinement of the other (Figure 14.a). For example, a topological structure represented in terms of growth units can be refined by considering the plant decomposition in terms of internodes. Each growth unit is considered as a set of internodes. Similarly, the axis structure of a plant can be interpreted as a refinement of the plant description in terms of branching systems, since each branching system can be decomposed into a set of axes. Hence, one modularity is a refinement of another if each module of the second can be decomposed into a set of modules of the first and, reciprocally, each module of the first modularity is a part of a module of the second. These modularities correspond to two topological structures representing the plant at two different *scales*. The highest scale corresponds to the finest modularity, while the lowest scale corresponds to the coarsest modularity. Within a plant representation, the scale of internodes is higher than the scale of growth units which is itself higher than the scale of axes.
- Secondly, the two modularities are not a refinement of eachother : they are overlapping (Figure 14.b). This is the case if at least one module of one modularity shares a common part with one module of the second modularity, whereas there is no inclusion of one into the other. Let us consider for example the topological structure of an apple tree in terms of both annual shoots and axes (14.b). At the beginning of the vegetative period, the apical meristem of some branches produces short shoots terminated by a flower, called ”bourse”[CL95]. During a second phase of the vegetative period, a vegetative shoot may develop on some bourses. These are called ”bourse shoots”. A bourse shoot is part of the same annual shoot as the bourse, since it is created during the same vegetative period. Therefore, some annual shoots are made of a bourse bearing a bourse shoot. Such an annual shoot is thus straddling two axes : on one side the axis terminated by the bourse and on the other side, the axis which

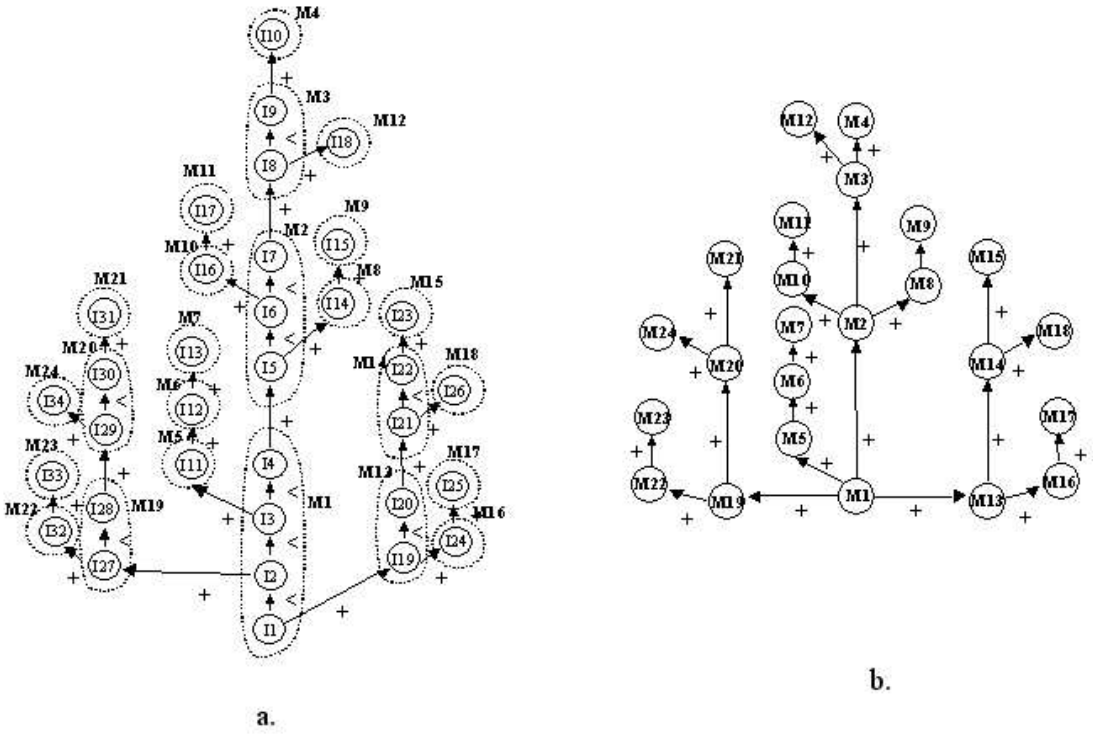


FIG. 12 – a. Partitionning graph of Figure 10 into growth units (M). b. Topology of the plant at scale M.

begins with the bourse shoot. Reciprocally, each axis is straddling two annual shoots. The modularities corresponding respectively to axes and annual shoots determine two topological interpretations of the plant which are not a refinement of eachother.

The different types of modularities that can be identified within a given plant define different topological structures. These modularities are *comparable* if they are refinements of each other. The refinement relation expresses the existence of a decomposition relation between the modules of the coarsest modularity and those of the finest. In the opposite case, modularities are *incomparable*, i.e. none of them is a refinement of the other. No decomposition relation exists between the modules of both modularities since they overlap.

Now, if we consider a graph g and different partitionning of the vertices of this graph, representing different modularities (Figure 15.a). Let us assume that each modularity is represented by a square element (Figure 15.b), and an edge is drawn from modulatities A to modularity B whenever A is a refinement of B. The graph obtained from this process is a lattice :

Let g be a tree graph. Let $L(g)$ be the set of all partitions on g, such that the induced macroscopic graph (quotient graph) is a tree graph.

$L(g)$ *is a lattice*

This proposition characterizes the space of all modularities that can be potentially defined on a given individual by a remarkable algebraic property : it is a sublattice of the partition lattice (the set of all subsets of a set). A multiscale graph is associated with only a subset of this sublattice. This subset corresponds to the set of modularities that are actually taken into consideration by the observer in the plant description. Multiscale graphs are thus a model of the observer's subjective interpretation of the plant.

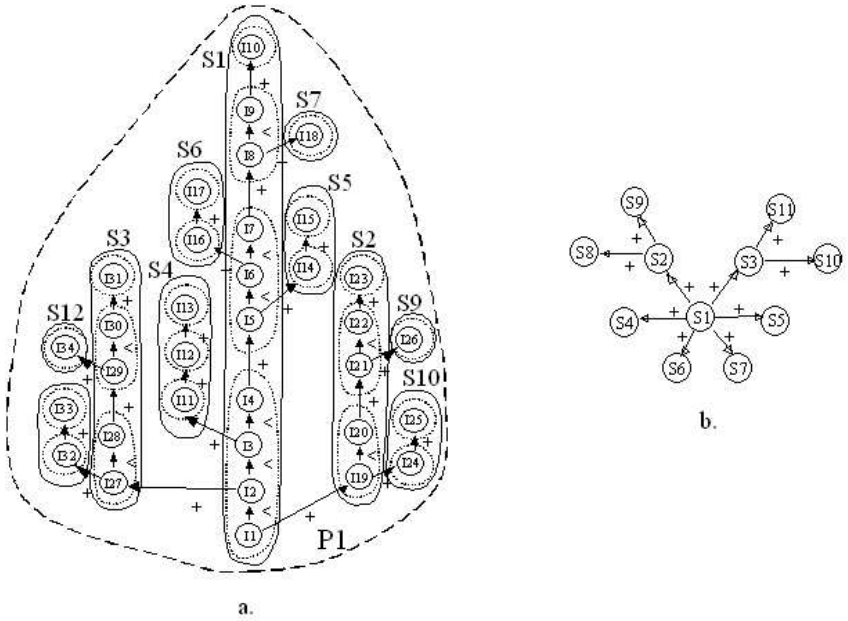


FIG. 13 – a. Multiscale graph corresponding to tree of Figure 10. b. corresponding topology at S module scale.

6.4 Growing multiscale structures

From a temporal point of view, the analysis of the relations between the different types of modularities is a delicate issue. Indeed, whereas the growth of a topological structure at a given scale seems to be a relatively clear phenomenon, the simultaneous growth of different topological structures representing a given individual, at different scales, raises the problem of understanding how these growth processes are linked to each other [GC98]. Figure 16 illustrates such a problem.

Consider an adult tree bearing a well hierarchized crown (16, date t_1). At a subsequent date t_2, a possible development of the crown may preserve the original hierarchy of branches. Another possible development is that one of the branches starts to compete with the trunk, yielding a reiterated complex (16 dates t_1 and t_2). This phenomenon can be interpreted in terms of MTGs (lower part of 16) if we assume that a component can belong to different complex entities throughout time.

The growth of a multiscale structure illustrates an important aspect of the model : rather than an objective plant topological structure, defined once and for all, a time-varying multiscale graph actually represents the plant topological structure as a subjective object depending on the observer's goals, knowledge and means of observation.

6.5 Handling plant architecture databases

Multiscale tree graphs are currently used as the backbone of a general methodology for measuring and analyzing plant topological structures, implemented in the AMAPmod software [GGC99]. Real plants are encoded by the observer using a specific coding language designed for this purpose. The multiscale plant topological structure can then be loaded into the computer. A set of dedicated tools, gathered in the AMAPmod software, enable the user to access these virtual plants and to explore them. They provide users with a methodology and corresponding tools to measure plants, create plant databases, analyse information extracted from these databases. This methodology can be depicted as follows (Figure 17).

Multiscale representation of plant architectures are described from either field observations or plant growth simulation programs, using a dedicated encoding language. The resulting database can then be analysed with various statistical analysis tools (e.g. [GBCC01]). Plants can be graphi-

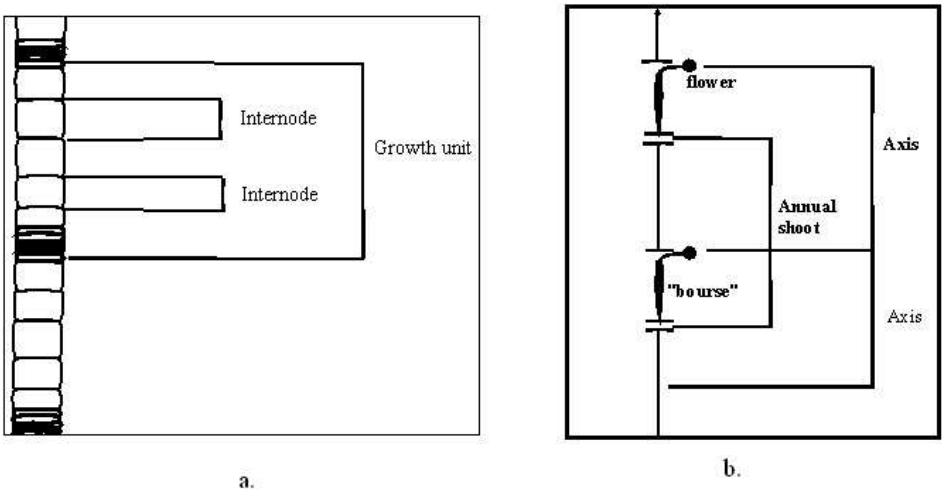


FIG. 14 – a. nested modularities. b. overlapping modularities

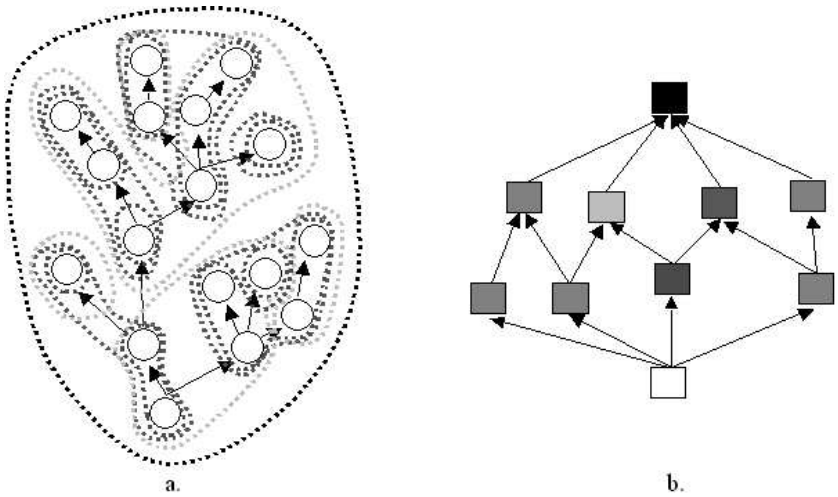


FIG. 15 – a. a general MTG. b. its corresponding modularity graph.

cally reconstructed at different scales and vizualised in 3 dimensions. Various types of data can be extracted and analysed with different viewpoints. Different families of probabilistic or stochastic models are provided in the system. These models are intended to be used as advanced statistical analysis tools for exploring in greater depth the information contained in the database. All these tools are available through a querying language called AML (AMAPmod Modelling Language) which enables the user to work on various objects, i.e. multiscale representation of plants, samples of data or models. AML provides the user with a homogeneous language-based interface to load, display, save, analyse or transform each type of object.

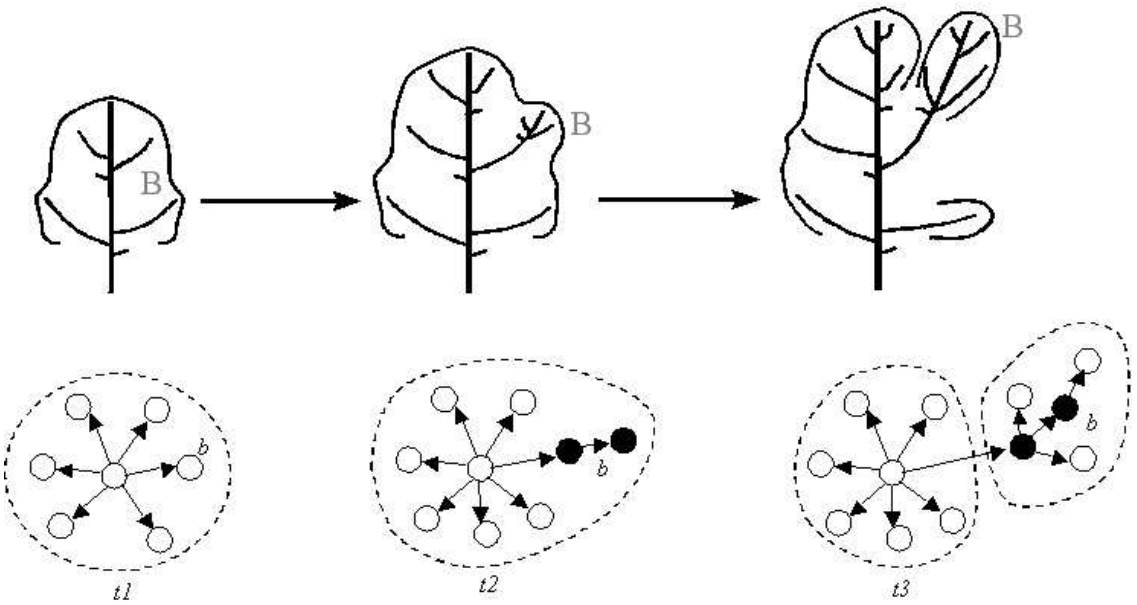


FIG. 16 – a. (upper part) reiterated complex is produced througout time. b. (lower part) Corresponding MTG interpretation

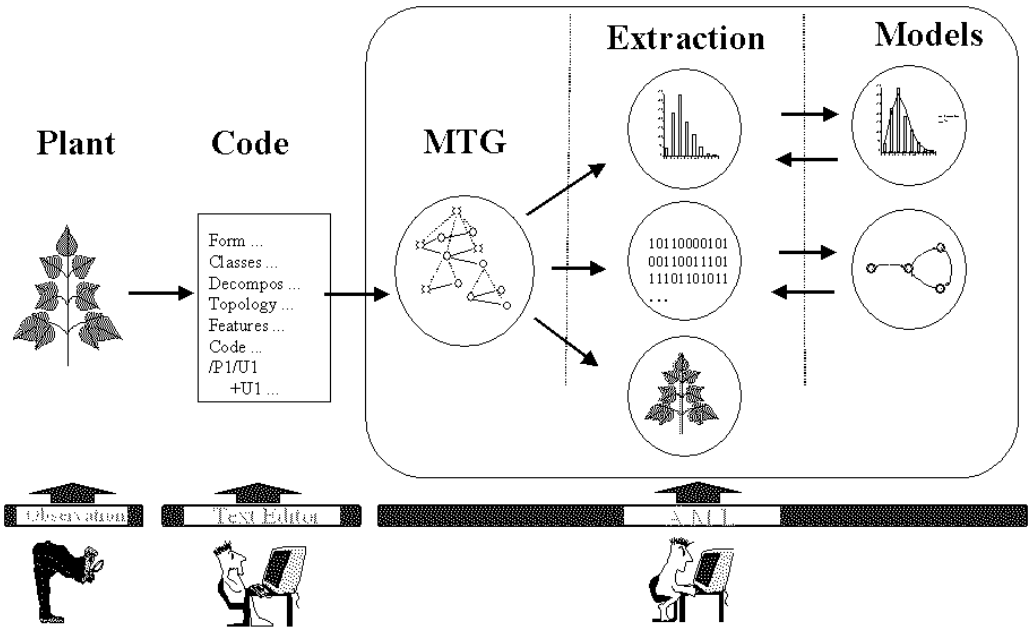


FIG. 17 – Synopsis of the AMAPmod system.

References

[Ada00] D. G. Adams. Heterocyst formation in cyanobacteria. *Current Opinoin in Microbiology*, 3 :618–624, 2000.

[Bar91] D. Barthélémy. Levels of organization and repetition phenomena in seed plants. *Acta Biotheoretica*, 39 :309–323, 1991.

[BB90] G. Berry and G. Boudol. The chemical abstract machine. In *Conf. Record 17th ACM Symp. on Principles of Programmming Languages, POPL'90, San Francisco, CA, USA, 17–19 Jan. 1990*, pages 81–94. ACM Press, New York, 1990.

[BCM87] J. P. Banatre, A. Coutant, and Daniel Le Metayer. Parallel machines for multiset transformation and their programming style. Technical Report RR-0759, Inria, 1987.

[BFM01] Jean-Pierre Banâtre, Pascal Fradet, and Daniel Le Métayer. Gamma and the chemical reaction model : Fifteen years after. *Lecture Notes in Computer Science*, 2235 :17– ? ?, 2001.

[BH70] R. Baker and G. T. Herman. CELIA — a cellular linear iterative array simulator. In *Proceedings of the Fourth Conference on Applications of Simulation* (9–11 December 1970), pages 64–73, 1970.

[BH72] R. Baker and G. T. Herman. Simulation of organisms using a developmental model, parts I and II. *International Journal of Bio-Medical Computing*, 3 :201–215 and 251–267, 1972.

[BH00] Ronald Brown and Anne Heyworth. Using rewriting systems to compute left kan extensions and induced actions of categories. *Journal of Symbolic Computation*, 29(1) :5–31, January 2000.

[BM86] J. P. Banatre and Daniel Le Metayer. A new computational model and its discipline of programming. Technical Report RR-0566, Inria, 1986.

[Bug00] A. Bugrim. A logic-based approach for computational analysis of spatially distributed biochemical networks. In *ISMB 2000*, San Diego California, August 2000.

[Cho56] N. Chomsky. Three models for the description of language. *IRE Trans. on Information Theory*, 2(3) :113–124, 1956.

[Cho57] N. Chomsky, editor. *Syntactic structures*. Mouton & Co., The Hague, 1957.

[CL95] E. Costes and P. L. Lauri. Processus de croissance en relation avec la ramification sylleptique et la floraison chez pommier. In J. Bouchon, editor, *Architecture des Arbres Fruitiers et Forestiers*, volume 74, pages 41–50, Montpellier, France, 1995. INRA Editions.

[Dit00] P. Dittrich. Artificial chemistry page, 2000.
http ://ls11-www.cs.uni-dortmund.de/achem.

[DJ90] N. Dershowitz and J.-P. Jouannaud. *Handbook of Theoretical Computer Science*, volume B, chapter Rewrite systems, pages 244–320. Elsevier Science, 1990.

[dL87] C. G. de Koster and A. Lindenmayer. Discrete and continuous models for heterocyst differentiation in growing filaments of blue-green bacteria. *Acta Biotheoretica*, 36 :249–273, 1987.

[DZB00] P. Dittrich, Jens Ziegler, and Wolfgang Banzhaf. Artificial chemistries - a review. *Artificial Life*, 2000. (to be submitted, available from the authors).

[Ede58] M. Eden. In H. P. Yockey, editor, *Symposium on Information Theory in Biology*, page 359, New York, 1958. Pergamon Press.

[ela02] Elan home page, 2002. http ://www.loria.fr/equipes/protheo/SOFTWARES/ELAN/.

[FMP00] Michael Fisher, Grant Malcolm, and Raymond Paton. Spatio-logical processes in intracellular signalling. *BioSystems*, 55 :83–92, 2000.

[GBCC01] Y. Guédon, D. Barthélémy, Y. Caraglio, and E. Costes. Pattern analysis in branching and axillary flowering sequences. *Journal of Theoretical Biology*, 212 :481–520, 2001.

[GC98] C. Godin and Y. Caraglio. A multiscale model of plant topological structures. *Journal of Theoretical Biology*, 191 :1–46, 1998.

[GGC99] C. Godin, Y. Guédon, and E. Costes. Exploration of plant architecture databases with the AMAPmod software illustrated on an apple-tree hybrid family. *Agronomie*, 19(03-avr) :163–184, 1999.

[GM01a] J.-L. Giavitto and O. Michel. Declarative definition of group indexed data structures and approximation of their domains. In *Proceedings of the 3nd International ACM SIGPLAN Conference on Principles and Practice of Declarative Programming (PPDP-01)*. ACM Press, September 2001.

[GM01b] J.-L. Giavitto and O. Michel. MGS : A programming language for the transformation of topological collections. Research Report 61-2001, CNRS - Université d'Evry Val d'Esonne, Evry, France, 2001.

[GM02] J.-L. Giavitto and O. Michel. The topological structures of membrane computing. *Fundamenta Informaticae*, 49 :107–129, 2002.

[GMS95] J.-L. Giavitto, O. Michel, and J.-P. Sansonnet. Group based fields. In I. Takayasu, R. H. Jr. Halstead, and C. Queinnec, editors, *Parallel Symbolic Languages and Systems (International Workshop PSLS'95)*, volume 1068 of *Lecture Notes in Computer Sciences*, pages 209–215, Beaune (France), 2–4 October 1995. Springer-Verlag.

[GV01] J.-L. Giavitto and E. Valencia. *Diagrammatic Representation and Reasonning*, chapter A Topological Framework for Modeling Diagrammatic Reasoning Tasks. Springer-Verlag, 2001.

[Han92] J. S. Hanan. *Parametric L-systems and their application to the modelling and visualization of plants*. PhD thesis, University of Regina, June 1992.

[HP96] M. Hammel and P. Prusinkiewicz. Visualization of developmental processes by extrusion in space-time. In *Proceedings of Graphics Interface '96*, pages 246–258, 1996.

[HR75] G. T. Herman and G. Rozenberg. *Developmental systems and languages.* North-Holland, Amsterdam, 1975.

[HRW86] J. L. Harper, B. R. Rosen, and J. White. *The growth and form of modular organisms*. The Royal Society, "London, UK", 1986.

[Jef85] D. Jefferson. Virtual time. *ACM Transactions on Programming Languages and Systems*, 7(3) :404–425, July 1985.

[JTN00] K. Chen J.J. Tyson, M.T. Borisuk and B. Novak. *Computational Modeling of Genetic and Biochemical Networks*, chapter Analysis of Complex Dynamics in Cell Cycle Regulation, pages 287–306. MIT Press, 2000.

[Kan00] Minoru Kanehisa. *Post-genome informatics*. Oxford University Press, 2000. ISBN 0-19-850326-1.

[Kau95] S Kaufman. *The Origins of Order : Self-Organization and Selection in Evolution*. Oxford University Press, 1995.

[Kel95] Evelyn Fox Kelle. *Refiguring Life : Metaphors of Twentieth-century Biology*. Columbia University Press, 1995.

[Kre86] W. Kreutzer. *System simulation : Programming styles and languages*. Addison-Wesley, Sydney, 1986.

[LIL89] C. Langton, L. In, and C. Langton. Artificial life, 1989.

[Lin68] A. Lindenmayer. Mathematical models for cellular interaction in development, Parts I and II. *Journal of Theoretical Biology*, 18 :280–315, 1968.

[Lin71] A. Lindenmayer. Developmental systems without cellular interaction, their languages and grammars. *Journal of Theoretical Biology*, 30 :455–484, 1971.

[Lin74] A. Lindenmayer. Adding continuous components to L-systems. In G. Rozenberg and A. Salomaa, editors, *L Systems*, Lecture Notes in Computer Science 15, pages 53–68. Springer-Verlag, Berlin, 1974.

[LP02] Brendav Lane and Przemek Prusinkiewicz. Specifying spatial distributions for multi-level models of plant communities. In *proc. of Graphics Interface 2002*, 2002.

[Lyn96] N. A. Lynch. *Distributed algorithms*. Morgan Kauffman, Los Altos, CA, 1996.

[Man01] Vincenzo Manca. Logical string rewriting. *Theoretical Computer Science*, 264 :25–51, 2001.

[mau02] Maude home page, 2002. http ://maude.csl.sri.com/.

[May75] R. M. May. Biological population models obeying difference equations : Stable points, stalbe cycles, and chaos. *Journal of Theoretical Biology*, 51 :511–524, 1975.

[May76] R. M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261 :459–467, 1976.

[Mei82] H. Meinhardt. *Models of biological pattern formation*. Academic Press, New York, 1982.

[Mic96] O. Michel. *Représentations dynamiques de l'espace dans un langage déclaratif de simulation*. PhD thesis, Université de Paris-Sud, centre d'Orsay, December 1996. N°4596, (in french).

[Pat94] Ray Paton, editor. *Computing With Biological Metaphors.* Chapman & Hall, 1994.

[Pau98a] Gheorge Paun, editor. *Computing with Bio-Molecules : Theory and Experiments*. Springer, 1998.

[Pau98b] Gheorghe Paun. Computing with membranes. Technical Report TUCS-TR-208, TUCS - Turku Centre for Computer Science, November 11 1998.

[Pau00] G. Paun. From cells to computers : Computing with membranes (p systems). In *Workshop on Grammar Systems*, Bad Ischl, austria, July 2000.

[PH90] P. Prusinkiewicz and J. Hanan. Visualization of botanical structures and processes using parametric L-systems. In D. Thalmann, editor, *Scientific visualization and graphics simulation*, pages 183–201. J. Wiley & Sons, Chichester, 1990.

[PJS92] H.-O. Peitgen, H. J urgens, and D. Saupe, editors. *Chaos and fractals. New frontiers of science*. Springer-Verlag, New York, 1992.

[PL90] P. Prusinkiewicz and A. Lindenmayer. *The algorithmic beauty of plants*. Springer-Verlag, New York, 1990. With J. S. Hanan, F. D. Fracchia, D. R. Fowler, M. J. M. de Boer, and L. Mercer.

[Pru98] P. Prusinkiewicz. Modeling of spatial structure and development of plants : a review. *Scientia Horticulturae*, 74 :113–149, 1998.

[Pru99] P. Prusinkiewicz. A look at the visual modeling of plants using L-systems. *Agronomie*, 19 :211–224, 1999.

[San92] E. Sanoja. *Essai d'application de l'architecture végétale à la systématique. L'exemple de la famille des Vochysiaceae*. PhD thesis, USTL Montpellier France, 1992.

[Smi99] John Maynard Smith. *Shaping Life : Genes, Embryos and Evolution*. Yale University Press, 1999.

[Ste88] Isabelle Stengers. *D'une science à l'autre. Les concepts nomades*. Le Seuil, 1988.

[TM87] T. Toffoli and N. Margolus. *Cellular automata machines : a new environment for modeling*. MIT Press, Cambridge, 1987.

[VN66] J. Von Neumann. *Theory of Self-Reproducing Automata*. Univ. of Illinois Press, 1966.

[WMS73] M. Wilcox, G. J. Mitchison, and R. J. Smith. Pattern formation in the blue-green alga, *Anabaena*. I. Basic mechanisms. *Journal of Cell Science*, 12 :707–723, 1973.

Part 7

LIST OF ATTENDEES

LIST OF ATTENDEES

(in alphabetic order)

Nadia Abchiche-Mimouni	abchiche@lami.univ-evry.fr
Patrick Amar	pa@lri.fr
Pascal Ballet	pascal.ballet@univ-brest.fr
Pierre Barbier	pierre.barbier@cirad.fr
Georgia Barlovatz-Meimon	Georgia.Barlovatz-Meimon@creteil.inserm.fr
Christophe Baron	chritop@hotmail.com
Vincent Bassano	vbassano@lami.univ-evry.fr
Grégory Batt	batt@inrialpes.fr
Marie-Claire Bellissent-Funel	mcbel@llb.saclay.cea.fr
Arndt Benecke	arndt@ihes.fr
Gilles Bernot	bernot@lami.univ-evry.fr
Marie Beurton-Aimar	aimar@u-bordeaux2.fr
Alexander Bockmayr	Alexander.Bockmayr@loria.fr
Armelle Cabin	armelle.cabin@univ-rouen.fr
David Campard	david.campard@univ-rouen.fr
Daniel Claude	Daniel.Claude@lss.supelec.fr
Emmanuel Collé	colle@ldfc.u-strasbg.fr
Olivier Collin	Olivier.Collin@imag.fr
Jean-Paul Comet	comet@lami.univ-evry.fr
Franck Delaplace	delapla@lami.univ-evry.fr
Jean-Marc Delosme	delosme@lami.univ-evry.fr
Maurice Demarty	maurice.demarty@univ-rouen.fr
Gora Diop	diop@cng.fr
Sarah Djebali	sdjebali@yahoo.fr
Marie Dutreix	marie.dutreix@curie.fr
Eric Fourmentin	fgfondat@club-internet.fr
Jean Fourmentin	fgfondat@club-internet.fr
Paul Francois	francois@lps.ens.fr
Christine Froidevaux	Christine.Froidevaux@lri.fr
Vincent Frouin	vincent.frouin@cea.fr
Jean-Louis Giavitto	giavitto@lami.univ-evry.fr
Christophe Godin	godin@cirad.fr
Yohann Grondin	jg69@le.ac.uk
Janine Guespin	janine.guespin@univ-rouen.fr
Maude Hatano	maudehatano@hotmail.com
Paul Hossenlopp	Paul.Hossenlopp@cnrs-dir.fr
Danièle Hernandez-Verdun	dhernand@ccr.jussieu.fr
Anastassia Iartseva	iartseva@mail.ru
Hassane Karkar	karkar@ccr.jussieu.fr
François Képès	Francois.Kepes@genopole.cnrs.fr
Sébastien Kerdelo	kerdelo@enib.fr
Guillaume Legent	g_legent@hotmail.com
Pierre Maziere	pierre.maziere@ibph.pharma.univ-montp1.fr
Hans Meinhardt	hans.meinhardt@tuebingen.mpg.de
Pascale Mentré	mentre@ibpc.fr
Chaouqi Misbah	chaouqi.misbah@ujf-grenoble.fr

Gradimir Misevic	gradimir@gradimir.com
Franck Molina	franck.molina@ibph.pharma.univ-montp1.fr
René Natowicz	natowicz@esiee.fr
Victor Norris	vjn@univ-rouen.fr
Nicolas Parisey	Nicolas@parisey.com
Sabine Peres	sabine.peres@etud.u-bordeaux2.fr
Aude Pflieger	pflieger@bordeaux.inserm.fr
Gabriel Querrec	gabriel.querrec@enib.fr
Nancie Reymond	nancie.reymond@jouy.inra.fr
Camille Ripoll	camille.ripoll@univ-rouen.fr
Sébastien Roby	roby.sebastien@worldonline.fr
Eduardo Rocha	erocha@abi.snv.jussieu.fr
Vincent Schachter	vs@genoscope.cns.fr
Bertrand Séraphin	Seraphin@cgm.cnrs-gif.fr
Serge Smidtas	sergi@sergi5.com
Fariza Tahi	tahi@lami.univ-evry.fr
Jan Traas	Jan.Traas@versailles.inra.fr
Philippe Tracqui	Philippe.Tracqui@imag.fr
Bernard Vandenbunder	bernard.vandenbunder@ibl.fr
Jean-Pierre Vannier	jean-pierre.vannier@chu-rouen.fr
Conrad Woldringh	woldringh@science.uva.nl
Abdallah Zemirline	zemirlin@univ-brest.fr