Proceedings of the Evry Spring School on

Modelling and simulation of biological processes in the context of genomics

March 29th to April 2nd 2004

Edited by:
Patrick Amar, Jean-Paul Comet,
François Képès, & Victor Norris

FOREWORD

What are the salient features of the new scientific context within which biological modelling and simulation will evolve from now on ? The global project of high-throughput biology may be summarized as follows. After genome sequencing comes the annotation by 'classical' bioinformatics means. It then becomes important to interpret the annotations, to understand the interactions between biological functions, to predict the outcome of perturbations, while incorporating the results from post-genomics studies (of course, sequencing and annotation do not stop when simulation comes into the picture). At that stage, a tight interplay between model, simulation and bench experimentation is crucial. 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 five 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 *G cube (Genomic Graphonomy Group)* subgroup works on networks of molecular interactions. Questions pertaining to the topology, dynamics and partitioning of molecular networks, and statistical inference of networks from post-genomic data, are discussed on a regular basis.

The work of the first four subgroups underpinned the first and second conferences organised in Autrans in 2002, and in Dieppe in 2003. The work of all workgroups underpinned the conference in Évry which, as reported here, brought together over a hundred participants, biologists, physical chemists, physicists, statisticians, mathematicians and computer scientists and gave leading specialists the opportunity to address an audience of doctoral and post-doctoral students as well as colleagues from other disciplines.

This book gathers overviews of the talks, discussions and roundtables, original articles contributed by speakers, and abstracts from attendees. We thank the sponsors of this seminar for making it possible for all the participants to share their enthusiasm and ideas in such a constructive way.

Patrick Amar, Gilles Bernot, Jean-Paul Comet, Franck Delaplace, Jean-Marc Delosme, Marie Dutreix, Jean-Louis Giavitto, Christophe Godin, Janine Guespin, François Képès, Franck Molina, Victor Norris, François Radvanyi, Vincent Schächter, Fariza Tahi, Philippe Tracqui.

CONTENTS

ACKNOWLEDGEMENTS

Part 1

CONFERENCE SYNTHESIS

"Modeling and simulating biological processes in the genomic era": An account of a multidisciplinary thematic school held in Évry (France) in April 2004.

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

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

Introduction

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

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

This five-day thematic school comprised lecture sessions, evening discussions around images, and afternoon workshops on specific topics. The five lecture sessions dealt with the epistemology of

---

[*] Correspondence should be addressed to F. Képès, Epigenomics Project / genopole®, Évry, France (Francois.Kepes@genopole.cnrs.fr).

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

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

Lecture sessions

Epistemology of modeling: interdiscipline or indiscipline ?

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

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

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

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

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

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

Luc Steels (Sony Computer Science Laboratory, Paris & Université libre de Bruxelles) addressed the question of self-organization and language evolution. The main hypothesis is that emergent processes are at the root of language evolution. This is in contrast with Chomski's theory of the " innate structures ". Linguistic arguments in favour of the emergent nature of language evolution are rooted in the study of this evolution itself, through language comparison and assessment of actual speech situations, where communication relies on a common (implicit) knowledge of part of the situation. The relationship between words and meaning is therefore very indirect, including several layers of categorization. The implicit plays a major role, each language differing from the others as to what is implicit (hence explicit). Starting from linguistic processes that give a hint of emergence, Steels uses multi-agent systems and learning robots endowed with a formal communication protocol to study the emergence of linguistic behaviors similar to real ones. In a community of 2 to 400 agents, what instruction must be given to each agent to obtain an evolution of the communication and a consistency of the language between all the agents? This has been tested for the emergence of sounds (one sound can spread into the community) or for a real communication where a common ground of knowledge appears during the experiment. One of the open questions is : what is general enough in these models that might also apply to genetics?

Eric Bonabeau (Icosystem Corp., Cambridge, Mass) discussed the modeling of self-organized systems from the bottom up. The emergent process under scrutiny was organization in social insects, where space must be introduced. Self-organizing processes are complex with many individuals involved, and with a global behavior that is generally non predictable, and may even be anti-intuitive. Thus, the only way to model it is to use an agent based modeling, starting from the "bottom" interactions between agents, to obtain, and therefore understand, the global, emergent, "up" behaviour. 2-D simulations of ant behavior can be achieved with the very simple rule according to which each ant is attracted by the pheromone released by other ants. Some elementary physical rules, such as the drawback of protruding legs of dead bodies may be necessary for proper

cemetery formations. In every instance, the presence of positive and negative feedback circuits is crucial. Simple 3-D agent-based modeling illuminated how coordination may emerge in the collective construction of wasp nests. In sum, the reverse approach, that consists of finding the elementary rules that correspond to a given structure, is quite appealing, but most difficult. This amounts to searching a very large space of rules, with choice criteria that are difficult to formalize. Genetic algorithms may be used, but the choice is most often intuitive and subjective.

Formal models for biological modeling

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

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

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

Christoph Teuscher (EPFL, Lausanne & UCSD, La Jolla) presented examples of computer architectures inspired by biological principles. These new hardware architectures are justified by the needs of new applications, by technological progress, and by new approaches to computation that

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

Macromolecular networks and statistical inference

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

The talk of Florence d'Alché-Buc (Programme d'Épigénomique, Évry & Université Paris 6) aimed at showing the benefits of a Machine Learning approach for modeling gene regulatory networks. The development of microarray technology makes it possible to compare simultaneously the expression of thousands of genes of a given organism (or tissue) in two conditions and offers new valuable pieces of information to construct gene regulatory networks. D'Alché-Buc introduced the task of constructing this network as a search of parameters for a model of gene interactions to be learned from experimental data. Then, she provided a survey of the problems and concepts of statistical learning. The central point was the formulation of the objective of the learning task as an optimization problem. Indeed, once the hypotheses space has been built, learning is tantamount to choosing the best hypothesis in the space. Is it justified to seek an optimal solution ? Not only is this task often unfeasible, but it is sometimes undesirable too. In the second part of her talk, D'Alché-Buc introduced the graphical models for reverse modeling. First, she presented the work by Segal *et al.* who introduce module networks, a probabilistic method for identifying regulatory modules from microarray data. Then she showed how dynamic Bayesian networks used by Perrin *et al.* for gene networks inference were especially well suited to tackle the stochastic nature of gene regulation and gene expression measurement. D'Alché-Buc concluded her talk by listing a number of perspectives, including the elaboration of a set of benchmark problems.

Wolfgang Banzhaf (Memorial University of Newfoundland, Canada) reminded us of the ubiquity of networks. To understand them, we should investigate their similarities and differences, the mechanisms for their generation and the interrelations between their structure and dynamics. Few steps have been made thanks to global characterization (scale free / small world topology) or local characterization (network motifs). Networks remain our best bet to catch emergent phenomena. Banzhaf then presented a method to compare natural regulatory networks from *E. coli*

and *S. cerevisiae*, and artificial ones. The artificial regulatory network comprises 32 genes and is generated by a neutralist duplication and divergence process. Banzhaf extracted structural elements of the networks (86 elementary 3-node sub-graphs) which represent basic elements of complex networks. Very few motifs occurred with significantly higher probability than in random networks. There was a clear relation between the natural distribution of motifs and the distribution in artificial networks generated by a duplication and divergence process. No evolutionary selection pressure had been applied to the artificial system, though. Thus, it can be stated that the distribution outcome is more a reflection of the mechanism of its generation than a result of evolutionary pressures.

Alessandro Vespignani (CNRS / Université Paris-Sud, Centre d'Orsay) introduced networks as a system that allows its abstract/mathematical representation as a graph. He argued on the ubiquity and variety of networks in physics, cybernetics, sociology, biology, etc. These networks are interconnected to form complex networks. They could form layers or have interdependencies. Vespignani provided an introduction to the general properties and possible generative mechanisms of such complex networks. By contrast to a complex system, a complicated system was described as a set of many elements assembled following a predefined blueprint imposed from outside. Consequently, the system shows expected properties and performs predefined tasks. On the other hand, complex systems which are composed of many interacting units show dynamical evolution and the capacity of self-organization. The outcome is a non-trivial architecture, unexpected emergent properties and cooperative phenomena. This was illustrated with protein-protein interaction networks (PIN) obtained from the two-hybrid technique. Finally, Vespignani proposed to use the global information provided by the PIN to functionally annotate proteins. This approach was based on maximizing unclassified interacting proteins with shared functionality and minimizing interacting classified proteins with different functionalities. It may serve as a general method to obtain statistical prediction of protein function from their interactions with reliability around 80% for highly interacting proteins.

Organization and morphogenesis

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

In his talk, Vincent Hakim (LPS, ENS, Paris) examined the synchronization properties of neuron networks. He investigated how the instantaneous firing rate of a neuron can be modulated by

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

Workshops

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

Short talks and posters

Some posters were highlighted by a short talk.

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

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

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

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

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

Conclusion

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

Poem

La log vraisemblance pénalisée ou comment se débarrasser des biologistes…

On a des tigres qui peuvent avaler plusieurs personnes.
On fait appel aux informaticiens pour trouver quel est le tigre qui doit être pénalisé.
Ces tigres sont discrets, et se cachent souvent dans les modèles de Markov
      c'est pour cela qu'on les appelle « tigres Markov cachés ».
On utilise le modèle graphique probabiliste
      c'est à dire qu'on dessine un animal qui est probablement un tigre.
Souvent il se prend les pieds dans un réseau bayésien car on prononce "x1, x2, x3"
      (terme simple et magique très utilisé par les fées locales qu'on appelle aussi
      inféesformatiques et qui essaient par ce moyen de rendre l'invraisemblable probable …)

On trouve grâce au graphe (ou au dessin)
quel était le tigre à l'origine du carnage.
On peut réitérer autant de fois que nécessaire …

Il faut bien entendu vérifier l'hypothèse.

Pour cela on envoie le biologiste vérifier et comme il se fait bouffer,
on conclut que le modèle était correct
ou que notre hypothèse était log-vraisemblable* !
* clairement la pénalité est pour le biologiste…

Georgia Barlovatz

Further reading

Acknowledgments

Part 2

CONFERENCE SCHEDULE

CONFERENCE SCHEDULE

MONDAY MARCH 29TH   **GENERAL INTRODUCTION - INTERDISCIPLINE OR INDISCIPLINE ?**

8:30 - 10:30	**Registration**
10:30 - 12:00	**The robot and the forest** *Pierre Sonigo*
12:00 - 13:30	**Lunch**
13:30 - 15:30	**Round table discussion about epistemology of modelling** *Lead by : Pierre Sonigo and Marc Van Regenmortel*
15:30 - 17:00	**Emergence in biology** *Marc Van Regenmortel*
17:30 - 18:00	**Short talks** *Chairmen : Marie Dutreix, Patrick Amar*
18:30 - 19:00	**Poster session**
19:00 - 21:00	**Dinner**
21:00 - 22:00	**Introduction to experimental approaches of the animal morphogenesis** *Nadine Peyrieras*

TUESDAY MARCH 30TH   **SIMULATION AND EMERGENT PROPRERTIES IN BIOLOGY (FROM MICROSCOPIC TO MACROSCOPIC)**

8:30 - 10:00	**Modelling the circadian clock in Drosophila and mammals** *Jean-Christophe Leloup – Albert Goldbeter*
10:30 - 12:00	**Self-organisation and language evolution** *Luc Steels*
12:00 - 13:30	**Lunch**
13:30 - 15:30	**Views on the choice of simulation tools, compared in the context of the same model: the bacteriophage Lambda (1)** *Speakers : Denis Mestivier, Pierre-Yves Boëlle* *Chairmen: Pascal Ballet, Abdallah Zermirline*
15:30 - 17:00	**Modelling self-organising systems from the bottom up** *Eric Bonabeau*
17:30 - 18:00	**Short talks** *Chairmen : Marie Dutreix, Patrick Amar*
18:30 - 18:00	**Poster session**
19:00 - 21:00	**Dinner**
21:00 - 22:00	**How the cell converts energy** *Giovanni Cappello*

WEDNESDAY MARCH 31ST   FORMAL MODELS FOR BIOLOGICAL MODELLING

 8:30 - 10:00 **Modelling of biochemical networks of interactions**
                Gordon Plotkin
10:30 - 12:00 **The Combinatorics of Membrane Interactions**
                Vincent Danos
12:00 - 13:30 **Lunch**
13:30 - 15:30 **Views on the choice of simulation tools, compared in the context of the same model: the bacteriophage Lambda (2)**
                Speakers: Pascal Ballet, Abdallah Zermirline, Adrien Richard
                Chairmen: Denis Mestivier, Kashayar Pakdaman
15:30 - 17:00 **The biochemical abstract machine BIOCHAM**
                François Fages
17:30 - 19:00 **Biologically-Inspired Cellular Computing Machines**
                Christof Teuscher
19:00 - 21:00 **Dinner**
21:00 - 22:00 **Models for axes formation in higher organisms**
                Hans Meinhardt

THURSDAY, APRIL 1ST   MACROMOLECULAR NETWORKS (STATISTICAL INFERENCE)

 8:30 - 10:00 **Network motifs in natural and artificial transcriptional regulatory networks**
                Wolfgang Banzhaf
10:30 - 12:00 **The Topology of Protein Interaction Networks**
                Alessandro Vespignani
12:00 - 13:30 **Lunch**
13:30 - 15:30 **Group 1 : Virtual reality immersion into the 3-D conformation of DNA**
                Presenter : Rachid Gherbi
                Group 2 : Structures depending of their functioning
                Presenter : Michel Thellier
15:30 - 17:00 **Technological developments for genetic network inference**
                Xavier Gidrol
17:30 - 19:00 **Machine learning for gene networks modeling**
                Florence d'Alché-Buc
19:00 - 19:45 **Cocktail party**
19:45 - 21:00 **Concert**
21:00   **Buffet**

FRIDAY, APRIL 2ND   ORGANIZATION AND MORPHOGENESIS

 8:30 - 10:00 **Microtubule self-organisation as an example of the development of order in living systems**
                James Tabony
10:30 - 12:00 **Synchronization and oscillations in neuronal networks**
                Vincent Hakim
12:00 - 12:30 **Conclusion and handing over of satisfaction questionnaires**
12:30 - 13:30 **Lunch**
13:30 - 16:00 **The Dynamic Geometry of Developing Organisms**
                Brian Goodwin

Part 3

ARTICLES

Network motifs in natural and artificial transcriptional regulatory networks

Wolfgang Banzhaf & P. Dwight Kuo

Memorial University of Newfoundland, St. John's, NL, Canada A1B 3X5,
e-mail: {banzhaf,kuo}@cs.mun.ca, Tel: (709) 737-8652, Fax: (709) 737-2009.

Abstract

We show that network motifs found in natural regulatory networks may also be found in an artificial regulatory network model created through a duplication / divergence process. It is shown that these network motifs exist more frequently in a genome created through the aforementioned process than in randomly generated genomes. These results are then compared with a network motif analysis of the gene expression networks of *Escherichia Coli* and *Saccharomyces cerevisiae*. In addition, it is shown that certain individual network motifs may arise directly from the duplication / divergence mechanism.

1 Introduction

Analysis of network motifs has recently become of interest with respect to transcriptional regulation networks. Methods are mainly based on searching for connection patterns among small numbers of nodes. Here we shall introduce a class of artificial regulatory networks which can be used to compare results obtained through the same methods as have been applied to natural regulatory networks of *Escherichia Coli* and *Saccharomyces Cerevisiae*. We shall see that the high frequency of certain network motifs detected in natural systems can be found in artificial systems as well, provided they are generated by a gene duplication and divergence process. This leads us to believe that the actual frequency distribution of motifs ("motif fingerprint") in natural regulatory networks is as much if not more a consequence of the process of network generation than of subsequent evolutionary selection.

The artificial regulatory network model presented here has previously been shown to generate networks which exhibit scale–free and small world network topologies [12]. Specifically, if the network generation process is one of duplication and divergence (similar to that presented in [16], though working on an actual genome level) we can show such global connectivity statistics. In this contribution, we extend those observations to network motifs and demonstrate that certain motifs frequent in natural regulatory systems also occur with regularity in this model.

It has also been shown in the past that the regulatory network model is able to reproduce dynamic phenomena found in natural genetic regulatory networks, for instance shifts in onset and offset of gene expression (heterochrony) based on single bit-flip mutations [3]. As such, this model can relate changes in time and intensity to tiny pattern changes on bit strings which could possibly provide the algorithmic " missing link " between genotypes subject to constant evolutionary changes and the remarkably stable phenotypes found in the real world.

2 Background

2.1 Regulatory Networks

Regulatory networks are an important new research area in biology [6, 8]. With the realization that in higher organisms only a tiny fraction of DNA is translated into proteins, the question of determining the function of the remaining DNA becomes all the more pertinent. A reasonable

answer for the function of this remaining unexpressed DNA appears to be regulation. According to Neidthardt et al. [15], 88% of the genome of the bacterium *E. Coli* is expressed with 11% suspected to contain regulatory information (also see Thomas [19]). Given the selective pressures on bacterial genomes, this would point to a very prominent role for regulation in general.

In addition, it has been recognized that understanding the differences between species and thus the key to evolution lies in the DNA information controlling gene expression [11]. Since many evolutionary effects can be traced back to their regulatory causes, regulatory networks mediate between development and evolution and thus serve to help unfold the patterns and shapes of organism morphology and behavior [10, 2].

Studying models of regulatory networks can help us to understand some of these mechanisms by providing lessons for both natural and artificial systems under evolution.

2.2   Network Motifs

There has recently been significant interest in studying static network motifs as a tool for understanding regulatory networks [14, 17, 23]. Complex networks have previously been classified by global characteristics such as scale–free [1, 4, 5, 9, 22] and small world network connection topologies [20, 21]. In order to investigate networks further beyond their global features requires an understanding of the potential basic structural elements which make up complex networks.

It has been proposed that studying so called "network motifs" can lead towards such an understanding [13, 14, 17]. Network motifs may be defined as the structural elements (subgraphs) which form the basic elements of more complex networks. Whereas an edge usually connects two nodes, network motif analysis starts with three nodes and their corresponding connections.

Interestingly, certain network motifs occur with significantly higher probability in natural regulatory networks than in random networks [14]. It has also been shown that network motifs may be conserved over evolutionary time, for instance in the yeast protein interaction network [23].

In order to detect all n–node network motifs, we have implemented an algorithm similar to one devised by Milo et al. [14]. The algorithm scans all rows of the adjacency matrix M of connections between nodes searching for non–zero elements (i, j) which represent a connection from node i to node j. The algorithm then recursively traverses the neighboring vertices connecting vertex i and j until a specific n–node motif is detected. The constituent vertices and edges of a motif are then compared to previously found motifs in order to ensure that none have been overcounted. It must also be noted that the total number of motifs of a given type is counted and possible isomorphisms are considered the same motif type. Table 1 in the Appendix lists all 3-node connection patterns in directed graphs, including auto-connections, up to isomorphism. We shall later refer to particular motifs with their motif ID (given in the table) only.

3   Artificial Regulatory Network Model

The artificial regulatory network (ARN) model presented here is based on work by one of the authors [3, 2]. The ARN consists of a bit string representing a genome with direction (i.e. 5' $\rightarrow$ 3' in DNA) and mobile "proteins" which interact with the genome through their constituent bit patterns. In this model, proteins are able to interact with the genome most notably at "regulatory" sites located upstream from genes. Attachment to these sites produces either inhibition or activation of the corresponding protein. It can thus be interpreted as a regulatory network with proteins acting like transcription factors.

The genome itself can be created through a series of duplication / divergence events. First, a random 32–bit string is generated. This string is then used in a series of length duplications followed by mutations in order to generate a genome of length L_G. A "promotor" bit sequence of 8–bits was then arbitrarily selected to signal the start of a gene on the genetic string analogous to

an open reading frame (ORF) on DNA. The actual gene length is set to a fixed length of $l_g = 5$ 32–bit integers which results in an expressed bit pattern of 160 bits per gene. Therefore, genes can be created by complete duplications of previously created genes, mutation, and / or combinations of the end and starting sequences of the genome during duplication.

Immediately upstream from the promotor sites exist two additional 32–bit segments which represent the enhancer and inhibitor sites. As previously mentioned, attachment of proteins (transcription factors) to these sites results in changes to protein production for the corresponding genes (regulation). In this model, we assume only one regulatory site for the increase of expression and one site for the decrease of expression of proteins. This is a radical simplification since natural genomes may have 5–10 regulatory sites that may even be occupied by complexes of proteins [2].

Processes such as transcription, and elements such as introns, RNA–like mobile elements and translation procedures resulting in a different alphabet for proteins are neglected in this model. This last mechanism is replaced as follows: Each protein is a 32–bit sequence constructed by a many–to–one mapping of its corresponding gene which contains five 32–bit integers. The protein sequence is created by performing the majority rule on each bit position of these five integers so as to arrive at a 32–bit protein. Ties (not possible with an odd number for l_g) for a given bit position are resolved by chance.

Proteins may then be examined to see how they may "match" with the genome. This comparison is implemented by using the XOR operation which returns a "1" if bits on both patterns are complementary. In this scheme, the degree of match between the genome and the protein bit patterns is specified by the number of bits set to "1" during an XOR operation. In general it can be expected that a Gaussian distribution results from measuring the match between proteins and bit sequences in the random genome [2].

By making the simplifying assumption that the occupation of both of a gene's regulatory sites modulates the expression of its corresponding protein, we may deduce a gene–protein interaction network comprising the different genes and proteins which can be parameterized by strength of match.

By examining the interaction networks at different matching strengths (we call thresholds) we may obtain different network topologies for the same connected network components. An example is shown in Figs. 1 and 2. Each node in the diagram represents a gene found in the genome along with its corresponding protein forming a gene–protein pair. Edges in the diagram represent some form of influence of one gene's protein on another gene. For the diagrams presented, a random genome was created by the previously mentioned duplication and mutation procedure with the network interaction diagrams being created at threshold levels of 21 and 22. Here and later we do not discern between enhancer and inhibitor sites, although such an analysis would be necessary to understand the actual function of motifs.

It must be stressed that although the actual genome has not changed, by simply changing the threshold parameter, we can obtain different network topologies. It may be noted by the more astute reader that the diagrams in Figs. 1 and 2 possess different numbers of genes and proteins. This is due to the fact that only connected gene–protein pairs are displayed in the diagrams. Should a change in the parameterized threshold lead to the creation of an isolated node, it is deleted from the diagram. Also note that only the largest network of interactions is displayed here.

It is possible to have multiple clusters of gene–protein interactions that are not interconnected. This is likely to occur as the threshold level is increased. As connections between gene–protein pairs are lost due to the threshold, each cluster of gene–protein pairs begins to become isolated from the others. This often occurs abruptly indicating a phase transition between sparse and full network connectivity.

The end result of this process would be first isolated pairs of nodes, then nodes without connections which would disappear from the network completely.
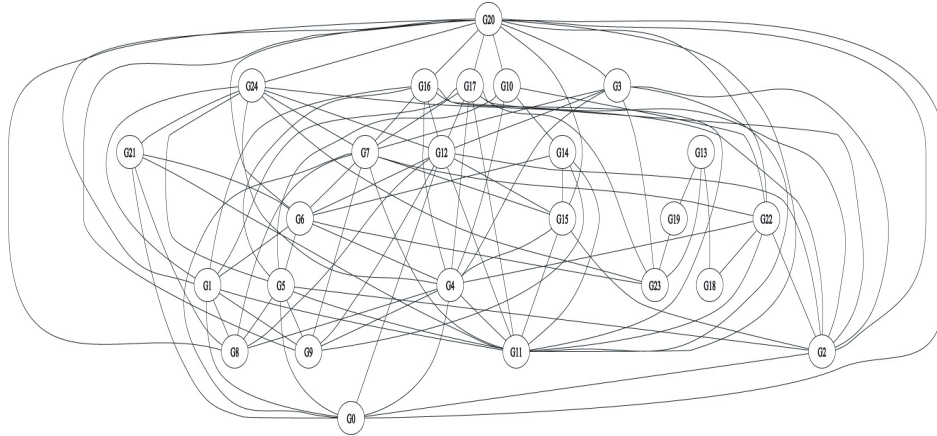
Figure 1: Sample of a gene–protein interaction network for a duplication/divergence genome at a threshold of 21 bits.
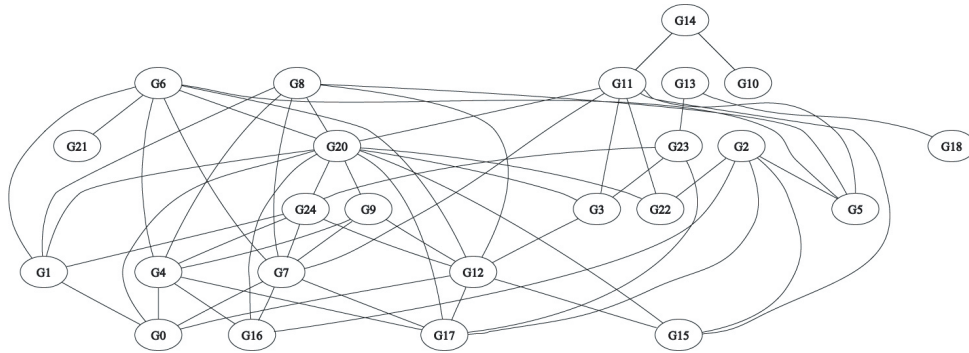
Figure 2: The same gene–protein interaction network at a threshold of 22 bits.

4   Results

The network motif finding algorithm was applied to 800 instances of the artificial regulatory model generated by the duplication / divergence process. As a control it was additionally applied to 800 networks whose genomes were generated randomly (by choosing the full number of bits at random). Results of motif counting are shown in Figs. 3 and 4. For both methods of network generation, the genome length was set at 131072 (12 duplication events in the case of duplication / divergence). For networks generated by duplication / divergence the mutation rate was set at 1%. In both cases the threshold had to be determined. We observed that the ratio of the number of edges to the number of vertices for the two natural regulatory networks was approximately 2 to 1. Therefore, in our artificial regulatory network framework, the threshold was chosen by iteratively raising the threshold until the network generated had a ratio that was equal to or less than 2 to 1.

This was then compared to the results of applying the algorithm to the transcriptional networks of *Escherichia Coli* [17, 18] and *Saccharomyces cerevisiae* [7], see Figures 5 and 6. Milo et al. defined network motifs as n–node subgraphs which occur significantly more than at random [14]. We prefer a more general definition here, and speak instead of a characteristic motif fingerprint if talking about the count in regard to a particular network. It can be seen in Figs. 3 - 6 that the most frequent natural motifs (ID 22 and ID 12) are both well represented in duplication/divergence type artifical networks whereas only one of them can be detected in fully random networks.

Table 2 in the Appendix lists all regulatory networks looked at for this manuscript, and lists
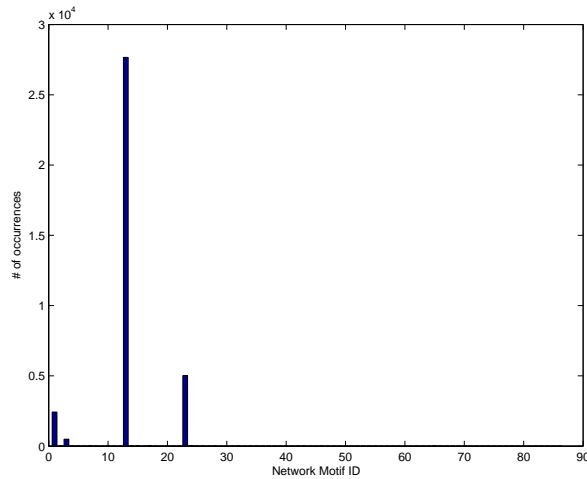
Figure 3: Average of frequency of occurrence of network motifs in 800 instances of the artificial network model generated by a duplication / divergence procedure.
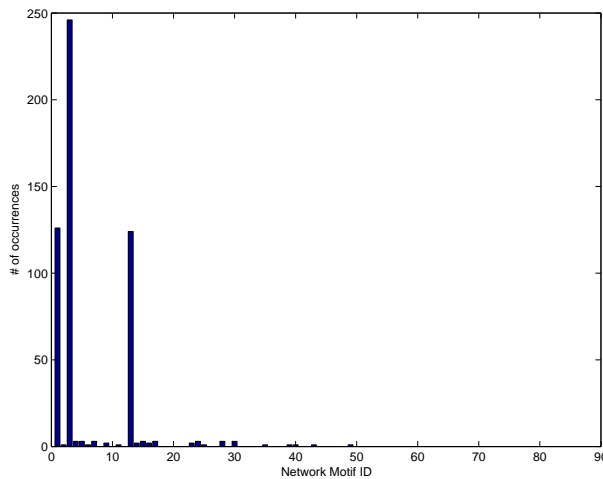
Figure 4: Average of frequency of occurrence of network motifs in 800 randomly generated instances of the artificial network model.

the distribution of all motifs. Note that for artificial networks we have chosen average numbers of counts, whereas there is only one example each for the natural regulatory systems.

5 Conclusions

A look at the table might convince us that there is a clear relation between the natural distribution of motifs and the distribution in artificial networks generated by a duplication and divergence process. No evolutionary selection pressure has been applied in artificial systems, though. Thus it can be stated that the distribution outcome is more a reflection of the mechanism of its generation than a result of evolutionary pressures (although evolutionary pressures are certainly responsible for fine-tuning of distributions).[1]

It might be pointed out that from the simplest elements of a network, two nodes connected by a

---

[1]The authors are aware of the fact that duplication / divergence events might interweave with selective pressure.
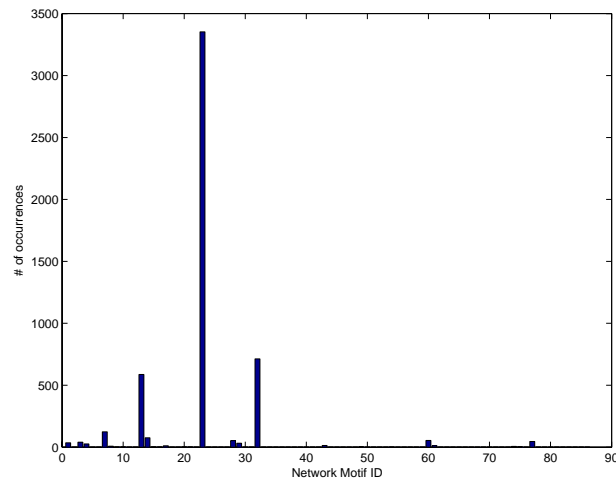
Figure 5: Frequency of occurrence of network motifs in the transcriptional network of *Escherichia Coli*
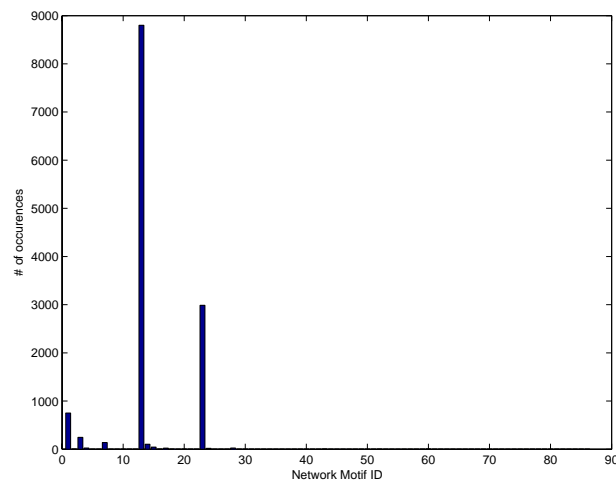
Figure 6: Frequency of occurrence of network motifs in the transcriptional network of *Saccharomyces cerevisiae* .

link (or one node, connected by a link to itself), an arbitrary number of patterns can be generated by duplication and divergence. The Bi-Fan (a 4-node) motif found in abundance in natural regulatory networks, can be easily generated from the above element by simple duplication. In the same vain, the feedforward loop consisting of 3 nodes (ID 14) can be generated from a 2-node motif with an autoconnection by a partial duplication and two mutations effecting a loss of the autoconnections.

So far we have not examined the case of enhancing and inhibiting connections, a fact that further complicates motif analysis. It remains to be seen whether dividing networks into their smallest components will teach us something useful about the overall structure of regulatory networks.

Acknowledgements
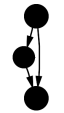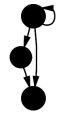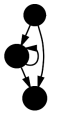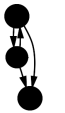
References

[1] R. Albert, H. Jeong, and A. Barabási. The internet's achilles' heel: Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000.

[2] W. Banzhaf. On the dynamics of an artificial regulatory network. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, and J. Ziegler, editors, *Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life (ECAL)*, volume 2801 of *Lecture Notes in Artificial Intelligence*, pages 217–227. Springer Verlag Berlin, Heidelberg, 2003.

[3] Wolfgang Banzhaf. Artificial regulatory networks and genetic programming. In Rick L. Riolo and Bill Worzel, editors, *Genetic Programming Theory and Practice*, chapter 4, pages 43–62. Kluwer, 2003.

[4] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[5] A. L. Barabási, H. Jeong, R. Ravasz, Z. Neda, T. Vicsek, and A. Schubert. Evolution of the social network of scientific collaborations. *Physica A*, 311:590–614, 2002.

[6] J.M. Bower and H. Boulouri, editors. *Computational Modeling of Genetic and Biochemical Networks*. MIT Press, Cambridge, MA, 2001.

[7] M. C. Costanzo, M. E. Crawford, J. E. Hirschman, J. E. Kranz, P. Olsen, L. S. Robertson, M. S. Skrzypek, B. R. Braun, K. L. Hopkins, P. Kondu, C. Lengieza, J. E. Lew-Smith, M. Tillberg, and J. I. Garrels. YPD^{TM}, PombePD^{TM} and WormPD^{TM}: model organism volumes of the BioKnowledge^{TM} library, an integrated resource for protein information. *Nucleic Acids Research*, 29(1):75–79, 2001.

[8] E. H. Davidson. *Genomic Regulatory Systems*. Academic Press, San Diego, CA, 2001.

[9] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM*, pages 251–262, 1999.

[10] F. Harold. *The Way of the Cell*. Oxford University Press, 2001.

[11] L. Hood and D. Galas. The digital code of DNA. *Nature*, 421:444–448, 2003.

[12] P. D. Kuo and W. Banzhaf. Scale–free and small world network topologies in an artificial regulatory network model. In *Artificial Life IX*, volume (in press), 2004.

[13] S. Mangan and U. Alon. Structure and function of the feed–forward loop network motif. *Proceedings of the National Academy of Science*, 100(21):11980–11985, 2003.

[14] R. Milo, S. Shen-Or, S. Itzkovitz, N Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002.

[15] F. C. Neidhardt. *Escherichia Coli and Salmonella typhimurium*. ASM Press, Washington, DC, 1996.

[16] P.S. Romualdo, E.D. Smith, and R.V. Solé. Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology*, 222:199–210, 2003.

[17] S. S. Shen-Or, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of Escherichia Coli. *Nature Genetics*, 31:64–68, 2002.

[18] D. Thieffry, A. M. Huerta, E. Pérez-Rueda, and J. Collado-Vides. From specific gene regulation to global regulatory networks: a characterisation of Escherichia Coli transcriptional network. *BioEssays*, 20:433–440, 1998.

[19] G. H. Thomas. Completing the E.Coli proteome: a database of gene products characterised since completion of the genome sequence. *Bioinformatics*, 7:860–861, 1999.

[20] D. Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton, NJ: Princeton University Press, 2003.

[21] D. Watts and S. Strogatz. Collective dynamics of small-world networks. *Nature*, 363:202–204, 1998.

[22] S. Wuchty. Scale-free behavior in protein domain networks. *Molecular Biology & Evolution*, 18(9):1694–1702, 2001.

[23] S. Wuchty, Z. N. Oltvai, and A.-L. Barabási. Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nature Genetics*, 35(2):176–179, 2003.

Appendix

Motif										
Motif ID	0	1	2	3	4	5	6	7	8	9
Motif ID	10	11	12	13	14	15	16	17	18	19
Motif ID	20	21	22	23	24	25	26	27	28	29
Motif ID	30	31	32	33	34	35	36	37	38	39
Motif ID	40	41	42	43	44	45	46	47	48	49
Motif ID	50	51	52	53	54	55	56	57	58	59
Motif ID	60	61	62	63	64	65	66	67	68	69
Motif ID	70	71	72	73	74	75	76	77	78	79
Motif ID	80	81	82	83	84	85				

Table 1: Network motifs and their ID

Network IDs		Count in				Network IDs		Count in			
ID	AlonID	D/D	Rand	*E.Coli*	*S.Cerv*	ID	AlonID	D/D	Rand	*E.Coli*	*S.Cerv*
0	6	2424	126	35	751	45	A	1	0	0	0
1	A	4	1	0	1	46	110	0	0	0	0
2	12	490	246	40	246	47	A	0	0	0	0
3	A	11	3	26	24	48	A	0	1	3	0
4	14	6	3	0	0	49	A	0	0	0	0
5	A	0	1	0	0	50	A	0	0	0	0
6	A	12	3	124	138	51	A	0	0	0	1
7	A	0	0	8	0	52	A	0	0	0	0
8	A	0	2	1	0	53	A	0	0	1	0
9	A	0	0	2	0	54	A	0	0	0	0
10	A	0	1	0	0	55	A	0	0	0	0
11	A	0	0	0	0	56	A	0	0	0	0
12	36	27659	124	587	8800	57	A	0	0	0	0
13	A	8	2	76	104	58	A	0	0	0	0
14	38	15	3	2	44	59	A	0	0	54	4
15	A	0	2	1	1	60	A	0	0	12	0
16	A	20	3	11	22	61	A	0	0	0	0
17	46	0	0	0	1	62	A	0	0	0	0
18	A	0	0	0	0	63	A	0	0	0	0
19	A	0	0	2	1	64	A	10	0	0	0
20	A	0	0	1	0	65	A	0	0	0	0
21	A	0	0	0	0	66	A	0	0	0	0
22	A	5016	2	3353	2987	67	A	0	0	0	0
23	74	36	3	0	18	68	238	0	0	0	0
24	A	5	1	0	0	69	A	0	0	0	0
25	78	3	0	0	0	70	A	0	0	0	0
26	A	0	0	0	0	71	A	0	0	0	0
27	A	6	3	53	25	72	A	0	0	0	0
28	A	0	0	32	0	73	A	0	0	6	0
29	A	0	3	0	0	74	A	0	0	3	0
30	A	0	0	0	0	75	A	0	0	0	0
31	A	14	0	713	0	76	A	0	0	46	0
32	A	0	0	0	0	77	A	0	0	0	0
33	A	3	0	0	0	78	A	0	0	0	0
34	A	0	1	0	0	79	A	0	0	0	0
35	A	0	0	0	0	80	A	0	0	0	0
36	A	0	0	0	0	81	A	0	0	0	0
37	A	0	0	0	0	82	A	0	0	0	0
38	98	0	1	0	0	83	A	0	0	0	0
39	A	0	1	0	0	84	A	0	0	0	0
40	102	0	0	0	0	85	A	0	0	0	0
41	A	0	0	0	0						
42	A	6	1	14	3						
43	A	0	0	0	0						
44	108	0	0	0	0						

Table 2: Network motifs and their distribution. D/D: Duplication/Divergence genomes; Rand: Random genomes. AlonIDs shown as A are autoconnected nodes without a code in Alon's system.

The Biochemical Abstract Machine BIOCHAM

Nathalie Chabrier-Rivier, François Fages & Sylvain Soliman

Projet Contraintes, INRIA Rocquencourt, BP105, 78153 Le Chesnay Cedex, France.
`http://contraintes.inria.fr`

Abstract

In this article we present the Biochemical Abstract Machine BIOCHAM and advocate its use as a formal modeling environment for networks biology. Biocham provides a precise semantics to biomolecular interaction maps. Based on this formal semantics, the Biocham system offers automated reasoning tools for querying the temporal properties of the system under all its possible behaviors. We present the main features of Biocham and report on our modeling experience with this language.

1 Introduction

In networks biology, the complexity of the systems at hand (metabolic networks, extracellular and intracellular networks, networks of gene regulation) clearly shows the necessity of software tools for reasoning globally about biological systems. Several formalisms have been proposed in recent years for modeling biochemical processes either qualitatively [23, 22, 11] or quantitatively [20, 14, 1, 13, 3]. State-of-the-art tools integrate a graphical user interface and a simulator, yet few formal tools are available for reasoning about these processes and proving properties of them. Our focus in Biocham has been on the design of a biochemical rule language and a query language of the model in temporal logic, that are intended to be used by biologists.

Biocham is a language and a programming environment for modeling biochemical systems, making simulations, and querying such models in temporal logic. Biocham is composed of :

1. a rule-based language for modeling biochemical systems,

2. a simple simulator,

3. a powerful query language based on Computation Tree Logic CTL,

4. an interface to the NuSMV [9] model checker for automatically evaluating CTL queries.

Biocham shares several similarities with the Pathway Logic system [11] implemented in Maude. Both systems rely on an algebraic syntax and are rule-based languages. One difference is the use in Biocham of CTL logic which allows us to express a wide variety of biological queries, and the use of a state-of-the-art symbolic model checker for handling the complexity of highly non-deterministic models.

The first experimental results of this approach for querying models of biochemical networks in temporal logic have been reported in [4, 5], on a qualitative model of the mammalian cell cycle control [8, 16] and in [5] on a quantitative model of gene expression [3]. In this paper we describe the Biocham system which provides a modeling environment supporting this methodology.

2   A Simple Example

The Mitogen-Activated Protein Kinase (MAPK) cascades are a well-known example of signal transduction, since they appear in many receptor-mediated signal transduction schemes. They are actively considered in pharmaceutical research, for their applications to cancer therapies. The MAPK/ERK pathway is indeed hyperactivated in 30% of all human cancer tumours [17].

   The structure of a MAPK cascade is a sequence of activations of three kinases in the cytosol. The last kinase, MAPK, when activated, has an effect on different substrates in the cytosol but also on gene transcription in the nucleus.

   Since this cascade has been studied a lot, mathematical models of it appear in most model repositories, like for instance that of Cellerator [25] or the SBML repository page [12], both coming from [18]. This cascade was also the first example treated by Regev, Silverman and Shapiro [23] in the pi-calculus process algebra which was an initial source of inspiration for our own work.

   Models based on ordinary differential equations (ODE) allow us to reproduce simulation results like the one pictured out in Figure 1, where the concentration of the visualized compounds is represented on the vertical axis and time on the horizontal axis. In Figure 2, the concentrations axis has been simply split and rescaled to the maximum value for each compound.
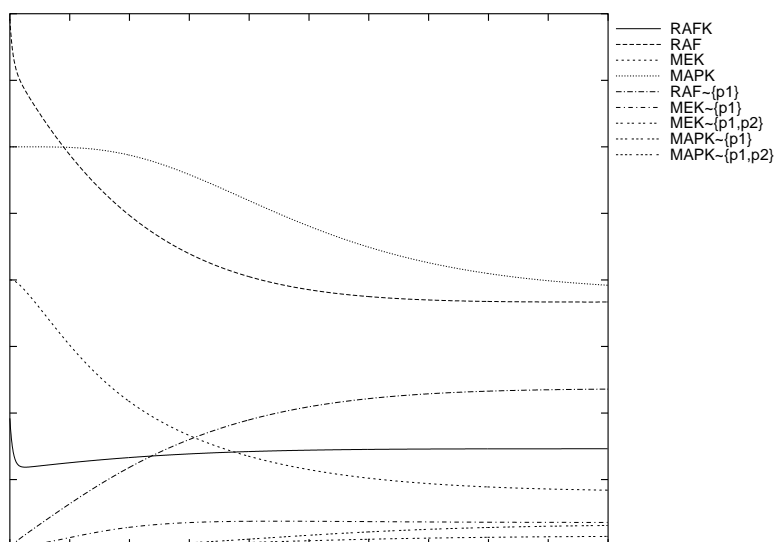
Figure 1: Simulation result of an ODE model of the MAPK cascade.

   It is possible to see from such simulations how the cascade evolves in time. It is possible to change input quantities to check for a significant change in the outcome of the simulation. Similarly, the sensibility of the system to the values of the parameters can be checked by running different simulations with different values of the parameters.

   Our aim in Biocham is to introduce complementary techniques to automate reasoning on all possible behaviors of the system modeled in a purely qualitative way. Taking the above model, one sees that it is built quite directly from the enzymatic reactions and Michaelis-Menten kinetics. Abstracting the kinetics part, one gets a system of biochemical reactions that can be interpreted as a non-deterministic transition system over boolean variables denoting the presence or absence of the compounds in the signaling cascade. The source code of this example is given in the appendix in Biocham syntax (explained in the next section) and in graphical form in Figure 4. The semantics of Biocham (explained in sections 3.2 and 3.3) ensures that the set of the possible behaviors of the boolean model *over-approximates* the set of all behaviors of the system for all kinetic parameter' values.
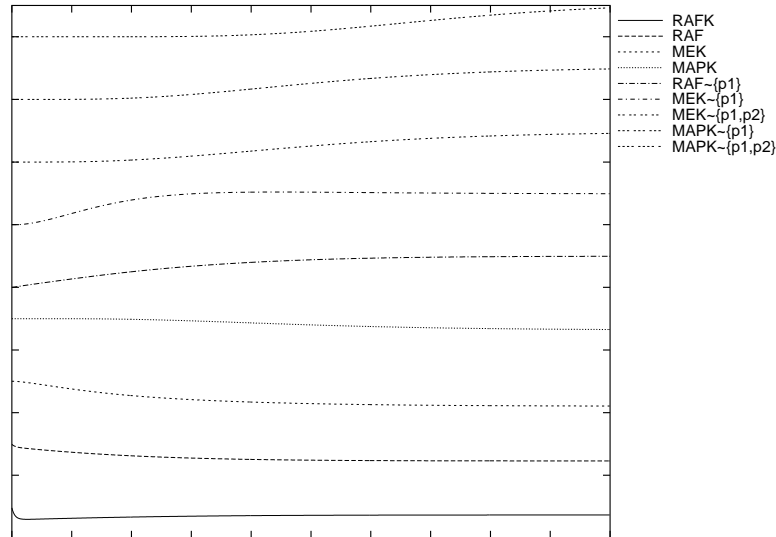
Figure 2: Same simulation as figure 1, side by side rescaled view.

Biocham uses the Computation Tree Logic (CTL) [10] as a query language for querying the temporal properties of the system under all possible conditions. A biological query like for example "Is the activation of the second kinase of the cascade (MEK) compulsory for the cascade ?" asks whether the phosphorylated form of MEK, noted in Biocham `MEK~{p1}`, is necessary to the production of the activated MAPK, noted `MAPK~{p1,p2}`, which is the output of the cascade, that is whether `MEK~{p1}` is a checkpoint. In Biocham, one expresses this query by the CTL formula

```
biocham: !(E(!(MEK~{p1}) U MAPK~{p1,p2}))
true
```

This formula expresses the non (!) existence (E) of a path on which `MEK~{p1}` is absent (!) until (U) `MAPK~{p1,p2}` becomes present, that is to say that `MEK~{p1}` is a checkpoint. This formula is checked automatically by the system.

The same query about a complex with a phosphatase, such as the complex `MEK~{p1}-MEKPH`, is false. These complexes are thus not checkpoints. The `why` command computes a counterexample in the form of a pathway which validates the negation of the query:

```
biocham: !(E(!(MEK~{p1}-MEKPH) U MAPK~{p1,p2}))
false
biocham: why
Step 1     Initial state
Step 2     rule 1    RAF-RAFK                 present
Step 3     rule 21   RAF~{p1}                 present
Step 4     rule 5    MEK-RAF~{p1}             present
Step 5     rule 24   MEK~{p1}                 present
Step 6     rule 7    MEK~{p1}-RAF~{p1}        present
Step 7     rule 23   MEK~{p1,p2}              present
Step 8     rule 13   MAPK-MEK~{p1,p2}         present
Step 9     rule 27   MAPK~{p1}                present
Step 10    rule 15   MAPK~{p1}-MEK~{p1,p2}    present
Step 11    rule 28   MAPK~{p1,p2}             present
```

This means that the complexes with a phosphatase (xxxPH) are intermediate products that do not strictly participate in the signal transduction. They are here to regulate the cascade, but they are not mandatory for the signal transduction in this model. A similar trace is obtained when asking a simple accessibility query like `EF(MAPK~{p1,p2})`, that is the existence (E) of a path on which at some time point (F) MAPK is fully phosphorylated.

It is worth noting that imposing the absence of an intermediate product is generally difficult in an ODE based simulation tool, without touching the model. Complex CTL queries thus have no natural counterpart in a numerical model and complement the information that can be deduced from interaction maps.

Querying a Biocham model in CTL temporal logic provides a mean to analyze exhaustively all possible behaviors of the system from first principles of enzymatic reactions, in particular when numerical data are not available. The simulation of Biocham models is also possible. Since a Biocham model is highly non-deterministic, simulations are randomized, which means that at each time step, one of the possible reactions happens. Figure 3 depicts one random simulation of the MAPK cascade, that is one but only one possible behavior of the system at the boolean abstraction level.
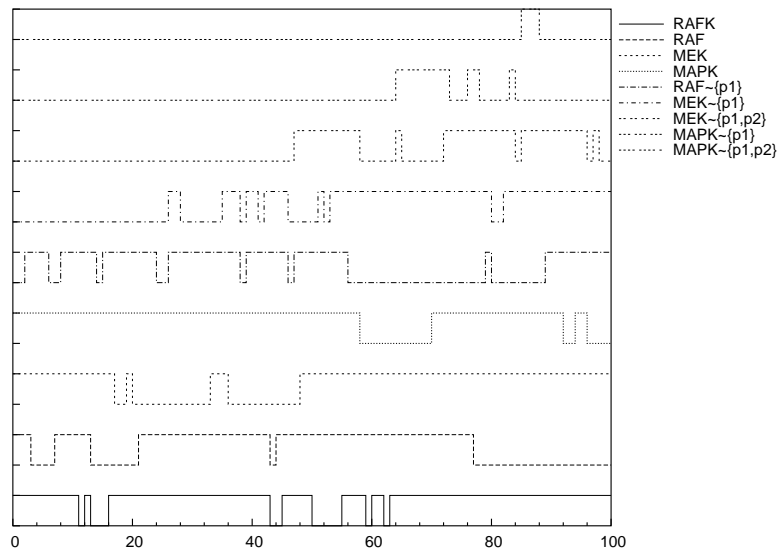
Figure 3: Random simulation of the Biocham model of the MAPK cascade

Biocham has been designed in the framework of the ARC CPBIO on "Process Calculi and Biology of Molecular Networks" [2] which aims at pushing forward a declarative and compositional approach to modeling languages in systems biology. The largest example treated so far with Biocham is a model of the Mammalian cell cycle control [4] developed after Kohn's map [16], involving 500 proteins and genes and 147 rule patterns which expand into 2733 rule instances. The computational results reported in [4] show the feasibility of this approach on such large examples as the CTL queries can be evaluated in a few seconds using state-of-the-art symbolic model-checking tools.

3 Modelling Biochemical Processes in Biocham.

3.1 A simple algebra of biochemical compounds

Biocham manipulates formal objects which represent chemical or biochemical compounds, ranging from ions, to small molecules, macromolecules and genes. Biocham objects can be used also to represent control variables and abstract biological processes.

Syntax:

object	=	molecule | abstract
molecule	=	name | molecule-molecule | molecule~{name,...,name} | gene | (molecule)
gene	=	#name
abstract	=	@name

In the simplest and the most flexible syntactical form, a molecule is simply given a name. Multimolecular complexes are denoted with the linking operator -. This binary operator is assumed to be associative and commutative, hence the order of the elements in a complex does not matter. Note that the same hypothesis is made in Pathway Logic [11] and other systems [19]. In the cases where one would like to distinguish between different orders of association, one can denote the different complexes with specific names. A third syntactical form serves to write modified forms of molecules, like attaching the set of phosphorylated sites with the operator $\sim$. Several sets can be attached. The order of the elements is irrelevant.

Example: `cdk1`, `cdk1-cycB` and `cdk1~{tyr15,thr161}-cycB` are valid Biocham notations for, respectively, the cyclin dependent kinase one, the complex cdk1 with cyclin B, and the phosphorylated form at sites tyrosine 15 and threonine 161 of cdk1 in the complex cdk1-cycB.`(cdk1-cycB)~{tyr15,thr161}` is another notation for the same phosphorylated form of the complex without making precise the constituent which is phosphorylated. Precising or not the phosphorylated constituent defines two formally different complexes.

The fourth syntactical form is used to denote genes or gene promotors, with a name beginning with #. These objects are assumed to be *unique*, which has a consequence on the way reactions involving such objects are interpreted by Biocham, as explained in the next section.

Example: `DMP1-#p19ARF` can be used to denote the binding of protein DMP1 on the promotor of the gene producing protein p19ARF noted `#p19ARF`.

The same assumption of uniqueness is made on abstract objects that are noted with a '@'. Abstract objects can be used to represent particular phases of a process, complete subsystems or abstract biological processes.

Example: `@UbiPro` can be used to denote the Ubiquitine/Proteasome subsystem and write this formal object as a catalyst in degradation reaction rules.

3.2 Reaction rules

Biocham reaction rules are used primarily to represent biochemical reactions. They can be used also to represent state transitions involving control variables or abstract processes, or to represent the main effects of complete subsystems such as protein synthesis by DNA transcription without introducing RNAs in the model.

Syntax:

reaction	=	name: reaction
		| solution => solution
		| solution =[object]=> solution
		| solution =[solution => solution]=> solution
		| solution <=> solution
		| solution <=[object]=> solution
solution	=	_ | object | solution + solution | (solution)

A solution is thus a sum of objects, the character _ denotes the empty solution. The order and multiplicity of molecules in a solution are ignored, only the presence or absence of objects are considered.

The following abbreviations can be used for reaction rules: A<=>B for the two symmetrical rules, A=[C]=>B for the rule A+C=>B+C with catalyst molecule C, and A=[C=>D]=>B for the rule A+C=>D+B.

Example: cdk1 + cycB => cdk1-cycB is a complexation rule. cdk1-cycB =[Myt1]=> cdk1~{thr14}-cycB is a phosphorylation rule with catalyst Mytosine 1. This rule is equivalent to cdk1-cycB + Myt1 => Myt1 + cdk1~{thr14}-cycB.

A reaction transforms one solution matching the left-hand side of the rule, into another solution in which the objects of the right-hand side have been added. The molecules in the left-hand side of the rule which do not appear in the right-hand side may be non-deterministically present or consumed in the resulting solution. This convention defines the *boolean abstraction* of stoichiometric models used in Biocham. It reflects the capability of Biocham to reason about all possible behaviors of the system with unknown concentration values and unknown kinetics parameters [1, 3].

Following the uniqueness assumption, molecule parts marked as "genes" with the '#' notation, or any compound built on such a molecule (such as DMP1-#p19ARF for instance) are not multiplied. These objects remain unique and they are deterministically consumed in the form in which they appear in the left-hand side of the rule. The same goes for control variables, noted with a '@', which are deterministically consumed.

Biocham has also a rich pattern language with constraints which is used to specify molecules and sets of reaction rules in a concise manner. Patterns introduce the special character ? and variables noted with a name beginning with a $ to denote unspecified parts of a molecule. These variables can be constrained with simple set constraints. The description of Biocham patterns [6] is however beyond the scope of this paper. The appendix contains the Biocham model of the MAPK cascade of the introductory example written with a set of 16 reaction rule patterns which expand into 30 rule instances.

3.3 Kripke semantics

A Biocham model is a set of reaction rules given with an initial state. The formal semantics of a Biocham model is a Kripke structure, that is a triple formed of a set of states, a transition relation between states and a labeling function associating to each state the set of atomic propositions true in that state. This formal semantics gives a mathematical meaning to Biocham models and provides a firm ground for :

- comparing different modeling formalisms and languages,

- comparing different models of a same biological system,

- importing models from other sources,

- and designing and implementing automated reasoning tools.

A *Kripke structure K* is thus a triple (S, R, L) where S is the set of states, $R \subseteq S \times S$ is a total relation (i.e. for any state $s \in S$ there exists a state $s' \in S$ such that $(s, s') \in R$), and $L : S \rightarrow 2^A$ is a labeling function over A the set of atomic propositions. A path in K starting from a state s_0 is an infinite sequence of states $\pi = s_0, s_1, \ldots$ such that $(s_i, s_{i+1}) \in R$ for all $i \geq 0$.

Clearly, one can associate to a Biocham model a Kripke structure, where the set of states S is the set of all tuples of boolean values denoting the presence or absence of the different biochemical compounds (molecules, genes and abstract processes), the transition relation R is the union (i.e. disjunction) of the relations associated to the reaction rules, and the labeling function L simply associates to a given state the set of biochemical compounds which are present in the state. Reaction rules in Biocham are asynchronous in the sense that one reaction rule is fired at a time (interleaving semantics), hence the transition relation is the union of the relations associated to the reaction rules. On the other hand, in a synchronous semantics for Biocham, the transition relation would have been defined by intersection. The choice of a synchronous semantics has been rejected in Biocham as it would bias fundamental biological phenomena such as the masking of a relation by another one and the resulting inhibition or activation of biological processes. Note that as explained in the previous section, the boolean abstraction of enzymatic reactions used in Biocham associates several transitions to a single Biocham reaction rule, one for each case of possible consumption of the molecules in the left-hand side of the rule.

That Kripke structure defines the semantics of a Biocham model as a non-deterministic transition system where the temporal evolution of the system is modeled by the succession of transition steps, and the different possible behaviors of the system are obtained by the non-deterministic choice of reactions.

3.4 *Importing biochemical models from other formalisms*

Since the basic building block of a Biocham model is an (enzymatic) reaction, it is quite easy to import any model based on such reactions into Biocham. This is the case of most graphical map-based models, but also of some ODE models, derived from the mass-action law or Michaelis-Menten kinetics. A well known source of such models is KEGG [15], which provides (graphical) maps of metabolic and signaling pathways. Biocham has been designed to provide such maps with a simple yet precise semantics.

In this respect, the Biocham project is part of the workpackage entitled "Towards a Bioinformatics Semantic Web" in the EU network REWERSE[1]. In parallel to this effort, the CMBSlib [26] web site[2] has been created as an open repository of computational models of biological systems, in order to:

- compare different *models* expressed in the same formalism,

- compare different *formalisms* and *tools* for a same model,

- cross-fertilize modeling experience and language issues between designers.

This library currently includes models of biological processes obtained from the literature and by translation from KEGG maps or ODE models into different formalisms. It is open to all contributions in any (ascii) format and in most exotic formalisms.

---

[1]The 6th EU Framework Programme Network of Excellence REWERSE stands for REasoning on the WEb with Rules and SEmantics, see `http://www.rewerse.net`

[2]`http://contraintes.inria.fr/CMBSlib`

4 Querying Biocham Models in Temporal Logic CTL.

Thanks to its simple Kripke semantics, Biocham supports the use of the Computation Tree Logic CTL [10] as a query language for querying the temporal properties of Biocham models. This methodology introduced in [4, 5] is implemented in Biocham with an interface to the state-of-the-art symbolic model checker NuSMV [9].

CTL basically extends propositional logic used for describing states, with operators for reasoning over time and non-determinism. Several temporal operators are introduced in CTL: $X\phi$ meaning ϕ is true at next transition, $G\phi$ meaning ϕ is always true, $F\phi$ meaning finally true, and $\phi U \psi$ meaning ϕ is always true until ψ becomes true. For reasoning about non-determinism, two path quantifiers are introduced: $A\phi$ meaning ϕ is true on all paths, $E\phi$ meaning ϕ is true on some path. In CTL, all temporal operators must be immediately preceded by a path quantifier (e.g. $AFG\phi$ is not in CTL, but $AF(EG\phi)$ is).

$$
\begin{array}{lll}
s \models \alpha & \text{iff} & \alpha \in L(s), \\
s \models E\psi & \text{iff} & \text{there is a path } \pi \text{ from } s \text{ such that } \pi \models \psi, \\
s \models A\psi & \text{iff} & \text{for every path } \pi \text{ from } s, \pi \models \psi, \\
\pi \models \phi & \text{iff} & s \models \phi \text{ where } s \text{ is the starting state of } \pi, \\
\pi \models X\psi & \text{iff} & \pi^1 \models \psi, \\
\pi \models F\psi & \text{iff} & \text{there exists } k \geq 0 \text{ such that } \pi^k \models \psi, \\
\pi \models G\psi & \text{iff} & \text{for every } k \geq 0, \pi^k \models \psi, \\
\pi \models \psi U\psi' & \text{iff} & \text{there exists } k \geq 0 \text{ such that } \pi^k \models \psi' \text{ and } \pi^j \models \psi \text{ for all } 0 \leq j < k.
\end{array}
$$

Table 1: Inductive definition of the truth relations $s \models \phi$ and $\pi \models \phi$ in a given Kripke structure K.

CTL is expressive enough to express a wide range of biological queries. Simplest queries are *about reachability*: is there a pathway for synthesizing a protein P, $EF(P)$? *About pathways*: can the cell reach a state s while passing by another state s_2, $EF(s_2 \wedge EF(s))$? is state s_2 a necessary checkpoint for reaching state s, $\neg E((\neg s_2)\ U\ s)$? can the cell reach a state s without violating certain constraints c, $E(c\ U\ s)$? from an initial state I, is it possible to synthesize a protein P without creating nor using protein Q, $I \Rightarrow E(\neg Q\ U\ P)$? *About steady states and permanent states*: is a certain (partially described) state s of the cell a steady state, $s \Rightarrow EG(s)$? a permanent state, $s \Rightarrow AG(s)$? can the cell reach a given permanent state s from the initial state I, $I \Rightarrow EF(AGs)$? must the cell reach a given permanent state s from the initial state I, $I \Rightarrow AF(AGs)$? can the system exhibit a cyclic behavior w.r.t. the presence of a product P, $EG((P \Rightarrow EF\ \neg P) \wedge (\neg P \Rightarrow EF\ P))$? The latter formula expresses that there exists a path where at all time points whenever P is present it becomes eventually absent, and whenever it is absent it becomes eventually present. This formula is not expressible in LTL [10], where formulas are of the form $A\phi$ with ϕ containing no path quantifier.

The formal semantics of CTL in a fixed Kripke structure K is given in table 1, as the inductive definition of the truth relation stating that a CTL formula ϕ is true at state s, written $s \models \phi$, or true along path π, written $\pi \models \phi$ (the clauses for ordinary boolean connectives are omitted). π^i denotes the suffix of π starting at s_i.

5 The Challenge of the Virtual Cell

High-throughput technologies addressing cell functions at a whole genome scale are revolutionizing cell biology. The challenge of virtual cell projects is to map molecular interactions within the cell, and to build virtual cell models predicting the effects of a drug on a given cell.

Virtual cell environments, like for instance the Virtual Cell project [24] or Cellerator [25], maintain a library of models of different parts of the cell, among different living organisms. ODE models typically range from a tenth of variables to 50 variables like in the Budding Yeast cell Cycle model of [7]. On the other hand, qualitative models represented by interaction maps allow for the global modeling of a large number of interacting subsystems.

A formal model of the Mammalian cell cycle control has been developed in the ARC CPBIO [4, 2] after Kohn's map [16]. This model transcribed in Biocham involves 500 proteins and genes and 147 reaction rule patterns which expand into 2733 reaction rule instances. Performance results of CTL querying in this model are reported in [4]. Symbolic model checking techniques used in Biocham are efficient enough to automatically evaluate CTL queries about biochemical networks of several hundreds or thousands of rules and variables [4, 5]. It is worth noting however that this is far below the size of digital circuits that the same model checking algorithms can treat. The reason for this discrepancy in performance comes from the high level of non-determinism which results from the competition between reaction rules and the soup aspect of biochemical solutions.

Combining ODE models with purely qualitative models like current Biocham models is an important issue for managing the complexity of concurrent interacting models. This combination is under investigation within the framework of non-deterministic hybrid systems.

6   Learning Reaction Rules from Temporal Properties

With such a simple syntax and semantics for describing reaction rules in Biocham, it is possible to apply learning techniques to reaction rules discovery. We have done some preliminary experiments using the inductive logic programming system Progol [21] for the automatic discovery of missing Biocham reaction rules in a simple model of the cell cycle with 10 variables, given a set of accessibility properties. The basic experiment consists in furnishing a set of examples of accessibility relations and a set of counterexamples, and letting the inductive logic program search for a set of reaction rules satisfying the accessibility properties of the system. In the first phase of validation of the learning technique, where we are, the models we use are known models, from which we compute a set of temporal properties, and remove one or more reaction rules to check whether the missing rules can be recovered by learning from the temporal properties.

More generally, the basic idea is to specify the intended or observed temporal properties of the system with CTL formulas, and apply learning techniques such as inductive logic programming, in order to correct the model by suggesting to add or modify Biocham rules in the model[3].

7   Conclusion and perspectives.

Biocham is a free software[4] for modeling biochemical processes and querying these models in temporal logic. The largest example treated so far is a model of the mammalian cell cycle control [4] after Kohn's diagram [16]. Other models have been imported from interaction maps available on the Web and ODE models. This shows the simplicity of the scheme and the flexibility of this approach.

The Pathway Logic of [11] is close to Biocham for the algebraic representation of cell compounds and the representation of molecular interactions by rewriting rules. However, the boolean abstraction used in Biocham and the state-of-the-art symbolic model checker NuSMV permit the handling of potentially larger models. The choice of CTL for expressing biological queries provides also more expressiveness than LTL, which is used in Pathway Logic. Much can be gained by

---

[3]We investigate this approach in the 6th PCRD EU project APRIL 2 "Applications of Probabilistic Inductive Logic Programming", http://www.aprill.org.

[4]Biocham system can be downloaded from http://contraintes.inria.fr/BIOCHAM

exchanging Biocham and Pathway Logic models, cross-fertilizing our modeling experiences and comparing language issues in particular w.r.t. the pattern language. The CMBSlib open repository [26] has been created for this purpose as well as for comparison with very different formalisms.

Currently, Biocham is primarily oriented towards the qualitative modeling of biochemical processes and the querying of the temporal properties of boolean models. This approach can be generalized however to numerical models by relying on to constraint-based model checking techniques [5]. In this extension, called Biocham2, variables can denote real values expressing the concentrations of molecules, and rules are extended with constraints to denote the relationship between the old and the new values of the variables. In particular, biochemical systems described by differential equations can be handled in this framework using time discretization methods, and can be combined with boolean models. The modeling power of such non-deterministic hybrid systems is under investigation.

Acknowledgments

References

[1] Rajeev Alur, Calin Belta, Franjo Ivanicic, Vijay Kumar, Max Mintz, George J. Pappas, Harvey Rubin, and Jonathan Schug. Hybrid modeling and simulation of biomolecular networks. In Springer-Verlag, editor, *Proceedings of the 4th International Workshop on Hybrid Systems: Computation and Control, HSCC'01*, volume 2034 of *Lecture Notes in Computer Science*, pages 19–32, Rome, Italy, 2001.

[2] ARC CPBIO. Process calculi and biology of molecular networks, 2002–2003. http://contraintes.inria.fr/cpbio/.

[3] Alexander Bockmayr and Arnaud Courtois. Using hybrid concurrent constraint programming to model dynamic biological systems. In Springer-Verlag, editor, *Proceedings of ICLP'02, International Conference on Logic Programming*, pages 85–99, Copenhagen, 2002.

[4] Nathalie Chabrier, Marc Chiaverini, Vincent Danos, François Fages, and Vincent Schächter. Modeling and querying biochemical networks. *Theoretical Computer Science*, To appear, 2004.

[5] Nathalie Chabrier and François Fages. Symbolic model cheking of biochemical networks. In Corrado Priami, editor, *CMSB'03: Proceedings of the first Workshop on Computational Methods in Systems Biology*, volume 2602 of *Lecture Notes in Computer Science*, pages 149–162, Rovereto, Italy, March 2003. Springer-Verlag.

[6] Nathalie Chabrier, François Fages, and Sylvain Soliman. *BIOCHAM's user manual*. INRIA, 2003–2004. http://contraintes.inria.fr/BIOCHAM/.

[7] Katherine C. Chen, Attila Csikász-Nagy, Bella Györffy, John Val, Bela Novàk, and John J. Tyson. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Molecular Biology of the Cell*, 11:396–391, 2000.

[8] Marc Chiaverini and Vincent Danos. A core modeling language for the working molecular biologist. In Corrado Priami, editor, *CMSB'03: Proceedings of the first Workshop on Computational Methods in Systems Biology*, volume 2602 of *Lecture Notes in Computer Science*, page 166, Rovereto, Italy, March 2003. Springer-Verlag.

[9] Alessandro Cimatti, Edmund Clarke, Fausto Giunchiglia Enrico Giunchiglia, Marco Pistore, Marco Roveri, Roberto Sebastiani, and Armando Tacchella. Nusmv 2: An opensource tool for symbolic model checking. In *Proceedings of the International Conference on Computer-Aided Verification, CAV'02*, Copenhagen, Danmark, July 2002.

[10] Edmund M. Clarke, Orna Grumberg, and Doron A. Peled. *Model Checking*. MIT Press, 1999.

[11] Steven Eker, Merrill Knapp, Keith Laderoute, Patrick Lincoln, José Meseguer, and M. Kemal Sönmez. Pathway logic: Symbolic analysis of biological signaling. In *Proceedings of the seventh Pacific Symposium on Biocomputing*, pages 400–412, January 2002.

[12] Michael Hucka et al. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003. http://sbml.org.

[13] Ronojoy Ghosh and Claire Tomlin. Lateral inhibition through delta-notch signaling: A piecewise affine hybrid model. In Springer-Verlag, editor, *Proceedings of the 4th International Workshop on Hybrid Systems: Computation and Control, HSCC'01*, volume 2034 of *Lecture Notes in Computer Science*, pages 232–246, Rome, Italy, 2001.

[14] Ralf Hofestädt and S. Thelen. Quantitative modeling of biochemical networks. In *In Silico Biology*, volume 1, pages 39–53. IOS Press, 1998.

[15] Minoru Kanehisa and Susumu Goto. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

[16] Kurt W. Kohn. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Molecular Biology of Cell*, 10(8):703–2734, August 1999.

[17] Walter Kolch, Ashwin Kotwaliwale, Keith Vass, and Petra Janosch. The role of raf kinases in malignant transformation. In *Expert Reviews in Molecular Medicine*, volume 25. Cambridge University Press, April 2002. http://www.expertreviews.org/02004386h.htm.

[18] Andre Levchenko, Jehoshua Bruck, and Paul W. Sternberg. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *PNAS*, 97(11):5818–5823, May 2000.

[19] Ron Maimon and Sam Browning. Diagrammatic notation and computational structure of gene networks. In Tau-Mu Yi, Michael Hucka, Mineo Morohashi, and Hiroaki Kitano, editors, *Proceedings of the 2nd International Conference on Systems Biology*. Online Proceedings, 2001. http://www.icsb2001.org/toc.html.

[20] Hiroshi Matsuno, Atsushi Doi, Masao Nagasaki, and Satoru Miyano. Hybrid petri net representation of gene regulatory network. In *Proceedings of the 5th Pacific Symposium on Biocomputing*, pages 338–349, 2000.

[21] Stephen H. Muggleton. Inverse entailment and progol. *New Generation Computing*, 13:245–286, 1995.

[22] Masao Nagasaki, Shuichi Onami, Satoru Miyano, and Hiroaki Kitano. Bio-calculus: Its concept, and an application for molecular interaction. In *Currents in Computational Molecular Biology*, volume 30 of *Frontiers Science Series*. Universal Academy Press, Inc., 2000. This book is a collection of poster papers presented at the RECOMB 2000 Poster Session.

[23] Aviv Regev, William Silverman, and Ehud Y. Shapiro. Representation and simulation of biochemical processes using the pi-calculus process algebra. In *Proceedings of the sixth Pacific Symposium of Biocomputing*, pages 459–470, 2001.

[24] J. Schaff and L. M. Loew. "the virtual cell". In *Proceedings of the fourth Pacific Symposium on Biocomputing*, pages 228–239, 1999.

[25] Bruce E. Shapiro, Andre Levchenko, Elliot M. Meyerowitz, Barbara J. Wold, and Eric D. Mjolsness. Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics*, 19(5):677–678, 2003. http://www-aig.jpl.nasa.gov/public/mls/cellerator/.

[26] Sylvain Soliman and François Fages. CMBSlib: a library for comparing formalisms and models of biological systems. In Vincent Danos and Vincent Schächter, editors, *CMSB'04: Proceedings of the second Workshop on Computational Methods in Systems Biology*, Lecture Notes in Computer Science. Springer-Verlag, 2004.

Appendix: Biocham model of the MAPK signaling cascade

Here is the full code of the MAPK example given in section 2. The phosphorylation sites for MEK and MAPK are declared first, and then the Biocham rules are given, sometimes with pattern variables (noted $P) which are constrained in the where part of the rules. In this model, the first rules are reversible, the other ones are directional. The initial state is partially defined.

```
% MAPK cascade in solution (no scaffold)
%
% adapted from:
% http://www-aig.jpl.nasa.gov/public/mls/cellerator/notebooks/MAPK-in-solution.html
% by Sylvain Soliman <Sylvain.Soliman@inria.fr>
% Nov. 26, 2003
%
% original source:
% Levchenko, A., Bruck, J., Sternberg, P.W. (2000) .Scaffold proteins may
% biphasically affect the levels of mitogen- activated protein kinase
% signaling and reduce its threshold properties. Proc. Natl. Acad. Sci. USA
% 97( 11):5818-5823.  http://www.pnas.org/cgi/content/abstract/97/11/5818

declare MEK~parts_of({p1,p2}).
declare MAPK~parts_of({p1,p2}).

RAF + RAFK <=> RAF-RAFK.

RAF~{p1} + RAFPH <=> RAF~{p1}-RAFPH.

MEK~$P + RAF~{p1} <=> MEK~$P-RAF~{p1}
   where p2 not in $P.

MEKPH + MEK~{p1}~$P <=> MEK~{p1}~$P-MEKPH.

MAPK~$P + MEK~{p1,p2} <=> MAPK~$P-MEK~{p1,p2}
   where p2 not in $P.
```

```
MAPKPH + MAPK~{p1}~$P <=> MAPK~{p1}~$P-MAPKPH.

RAF-RAFK => RAFK + RAF~{p1}.

RAF~{p1}-RAFPH => RAF + RAFPH.

MEK~{p1}-RAF~{p1} => MEK~{p1,p2} + RAF~{p1}.
MEK-RAF~{p1} => MEK~{p1} + RAF~{p1}.

MEK~{p1}-MEKPH => MEK + MEKPH.
MEK~{p1,p2}-MEKPH => MEK~{p1} + MEKPH.

MAPK-MEK~{p1,p2} => MAPK~{p1} + MEK~{p1,p2}.
MAPK~{p1}-MEK~{p1,p2} => MAPK~{p1,p2} + MEK~{p1,p2}.

MAPK~{p1}-MAPKPH => MAPK + MAPKPH.
MAPK~{p1,p2}-MAPKPH => MAPK~{p1} + MAPKPH.

% Are present in the initial state, the following molecules:

present({
RAFK,
RAF,
MEK,
MAPK,
MAPKPH,
MEKPH,
RAFPH
}).

% All the other ones, which are complexed forms or phosphorylated forms
% are absent

absent({?-?,?~{p1}~?}).
```

The last line of the file uses patterns to declare absent from the initial state all molecules that are complexes (?-?) or phosphorylated at p1 (?~{p1}~?). It is equivalent to the following sequence:

```
absent(RAF-RAFK).
absent(RAFPH-RAF~{p1}).
absent(MEK-RAF~{p1}).
absent(MEK~{p1}-RAF~{p1}).
absent(MEKPH-MEK~{p1}).
absent(MEKPH-MEK~{p1,p2}).
absent(MAPK-MEK~{p1,p2}).
absent(MAPK~{p1}-MEK~{p1,p2}).
absent(MAPKPH-MAPK~{p1}).
absent(MAPKPH-MAPK~{p1,p2}).
absent(RAF~{p1}).
absent(MEK~{p1}).
absent(MEK~{p1,p2}).
absent(MAPK~{p1}).
absent(MAPK~{p1,p2}).
```

When loading the model into BIOCHAM, one can see the expanded rules and the rule numbering used to explain the answer to the CTL query of section 2.

```
BIOCHAM 1.1 (C) 2003, 2004 INRIA, France,
```

```
by N. Chabrier-Rivier, F. Fages and S. Soliman.
http://contraintes.inria.fr/BIOCHAM
biocham: load_biocham('EXAMPLES/MAPK/mapk.bc').

biocham: expand_rules.
1  RAF+RAFK=>RAF-RAFK.
2  RAF-RAFK=>RAF+RAFK.
3  RAF~{p1}+RAFPH=>RAFPH-RAF~{p1}.
4  RAFPH-RAF~{p1}=>RAF~{p1}+RAFPH.
5  MEK+RAF~{p1}=>MEK-RAF~{p1}.
6  MEK-RAF~{p1}=>MEK+RAF~{p1}.
7  MEK~{p1}+RAF~{p1}=>MEK~{p1}-RAF~{p1}.
8  MEK~{p1}-RAF~{p1}=>MEK~{p1}+RAF~{p1}.
9  MEKPH+MEK~{p1}=>MEKPH-MEK~{p1}.
10 MEKPH-MEK~{p1}=>MEKPH+MEK~{p1}.
11 MEKPH+MEK~{p1,p2}=>MEKPH-MEK~{p1,p2}.
12 MEKPH-MEK~{p1,p2}=>MEKPH+MEK~{p1,p2}.
13 MAPK+MEK~{p1,p2}=>MAPK-MEK~{p1,p2}.
14 MAPK-MEK~{p1,p2}=>MAPK+MEK~{p1,p2}.
15 MAPK~{p1}+MEK~{p1,p2}=>MAPK~{p1}-MEK~{p1,p2}.
16 MAPK~{p1}-MEK~{p1,p2}=>MAPK~{p1}+MEK~{p1,p2}.
17 MAPKPH+MAPK~{p1}=>MAPKPH-MAPK~{p1}.
18 MAPKPH-MAPK~{p1}=>MAPKPH+MAPK~{p1}.
19 MAPKPH+MAPK~{p1,p2}=>MAPKPH-MAPK~{p1,p2}.
20 MAPKPH-MAPK~{p1,p2}=>MAPKPH+MAPK~{p1,p2}.
21 RAF-RAFK=>RAFK+RAF~{p1}.
22 RAFPH-RAF~{p1}=>RAF+RAFPH.
23 MEK~{p1}-RAF~{p1}=>MEK~{p1,p2}+RAF~{p1}.
24 MEK-RAF~{p1}=>MEK~{p1}+RAF~{p1}.
25 MEKPH-MEK~{p1}=>MEK+MEKPH.
26 MEKPH-MEK~{p1,p2}=>MEK~{p1}+MEKPH.
27 MAPK-MEK~{p1,p2}=>MAPK~{p1}+MEK~{p1,p2}.
28 MAPK~{p1}-MEK~{p1,p2}=>MAPK~{p1,p2}+MEK~{p1,p2}.
29 MAPKPH-MAPK~{p1}=>MAPK+MAPKPH.
30 MAPKPH-MAPK~{p1,p2}=>MAPK~{p1}+MAPKPH.
```

It is possible to export a .dot file of the rules, to use with the Graphviz [5] visualization suite. The generated map is depicted in Figure 4.
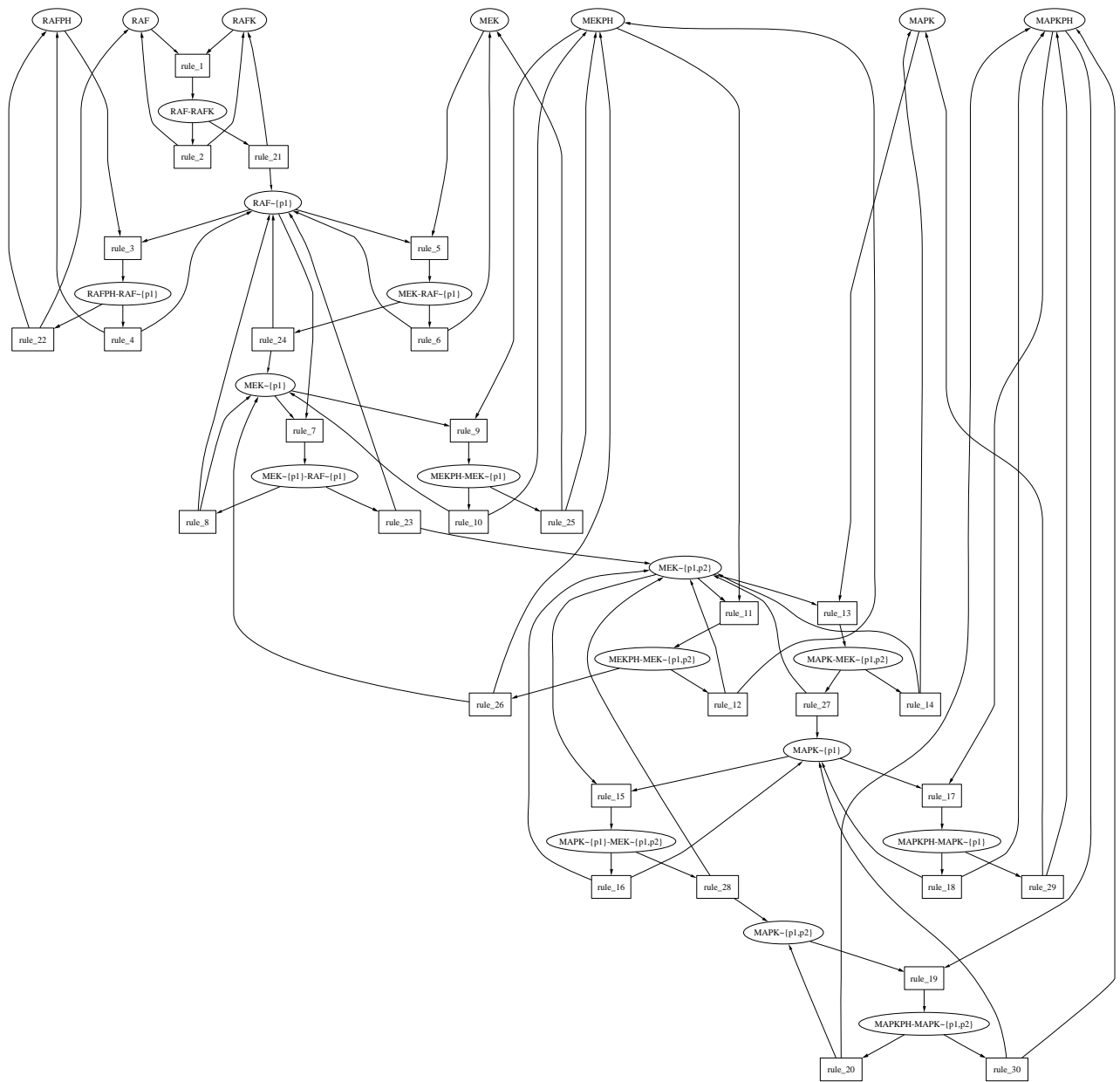
Figure 4: Interaction map generated from BIOCHAM rules for the MAPK cascade.

Kinesin step without external force takes less than 70 microseconds

Aurélie Dornier[1], Lorenzo Busoni[1], Konstantin Zeldovich[1], Boris Guirao[1], Gauthier Robin[2], Giovanni Cappello[1] & Jacques Prost[1]

[1]Physico Chimie Curie, UMR CNRS/IC 168, 26 rue d'Ulm, 75248 Paris Cedex 05, France.
[2]Institut de Biologie Structurale (CEA-CNRS-UJF), 41 rue Jules Horowitz, 38027 Grenoble Cedex 01, France.

Key words

Kinesin, Step, Total Internal Reflection, Interference.

Abstract

The kinesin-microtubules dynamics is characterized by *Interference Total Internal Reflection Microscopy* (ITIRM). The Neurospora kinesin step is measured with nanometer precision and a time resolution of six microseconds, in absence of external load. We observe that, without external force, the kinesin 8-nanometer step occurs in less than 70 microseconds and the step has no internal structure at this time scale.

1 Introduction

The kinesin-microtubule is a protein complex, which catalyzes ATP hydrolysis: $ATP \Rightarrow ADP + P_i$. During the enzymatic reaction, the kinesin-microtubule complex converts chemical energy into mechanical work. Together with dynein-microtubule and myosin-actin complexes, it is responsible for intracellular transport, mitosis and many other biological processes. Kinesin is a dimer of two identical subunits; it contains two motor heads and a coiled-coil stalk. This two-headed motor moves processively [1] along microtubules towards the plus-end, and travels over a mean distance of more than 1 μm without releasing from the microtubule. Recent experiments [2] also indicate that small mutations in the amino-acid sequence induce a change, or a loss, of the motor directionality. Weak mutations also affect the processivity of the enzyme. Optical tweezers experiments have shown that kinesin develops a force of a few picoNewton (stall force $\approx$ 6-7 pN) and that the motion is achieved by discrete steps of 8 nanometers [3, 4]. This distance corresponds to the periodicity of the $\alpha\beta$-tubulin arrangement in microtubules. Each step requires the hydrolysis of one ATP molecule [5]. Energy studies [6] show that for a load of 5 pN the kinesin efficiency can reach 50%.

The mechanism by which the molecular motors achieve such a high efficiency is still poorly understood. As well, the effect of the amino-acid mutations on the dynamics of kinesin cycle, and kinesin directionality has to be investigated. A detailed description of the kinesin motion is essential to understand those mechanisms [7, 8]: it includes the measurement of the power-stroke time scale, the shape of the step and its internal structure. It is also necessary to determine how the energy is dissipated and in which part of the kinesin cycle the dissipation occurs.

In order to measure the step details of kinesin we develop a novel technique for imaging small particles with high spatial and temporal resolution: the Interference Total Internal Reflection Microscopy ($ITIRM$) [9, 10]. This technique, described in section 2, allows us to follow the kinesin motion without external manipulation and it makes possible measurements at zero force.

The experiments described in this paper are performed on kinesin carrying a small cargo (200 nm bead) and without external load.

2 Interference Total Internal Reflection Microscopy

A single particle (i.e. a small bead) moves through a sine-like spatially modulated light. The particle necessarily scatters the light proportionally to its local intensity: going across the modulation with constant speed, the particle appears as a blinking dot. The modulation is obtained by interference of two laser beams, with opposite wave vectors $\overrightarrow{k_x}$, undergoing total internal reflection at the glass/water interface. In this setup, the fringe periodicity is:

$$d = \frac{\lambda_0}{2 n_{glass} \cos \theta} \tag{1}$$

where λ_0 is the laser wavelength, θ is the incidence angle at the glass/water interface, and n_{glass} is the glass refraction index (usually 1.52).
The resulting light is:

$$I(x,z) \propto \left(1 + \cos \frac{2\pi \cdot x}{d} \right) e^{-z/\zeta} \tag{2}$$

where ζ is the penetration length of the evanescent wave.
Measurement of the temporal variations of the bead luminosity allows us to estimate its position in the fringes. If the collected light scattered by the bead varies between I_{min} and I_{max}, the bead position $x(t)$ can be expressed as a function of the scattered intensity $I(t)$:

$$x(t) = \frac{d}{2\pi} \sin^{-1} \left[\frac{I(t) - (I_{\max} + I_{\min})/2}{(I_{\max} - I_{\min})/2} \right] \tag{3}$$

The particle is localized with a precision that only depends on our ability to realize a pure sine-like modulation. The time resolution of this setup depends on the detection speed, which can easily go up to many megahertz with fast detectors (avalanche photo diodes or photomultipliers). On the other hand, the main limitation to the time resolution is imposed by the number of photons collected per sample: use of very high sample rates is actually limited by the shot-noise.
The experimental setup is described in Fig. 1. The laser beam is compressed (L_1 and L_2 in Fig. 1) and splitted by using the beam splitter BS_1: two beams are produced, identical in intensity, phase and polarization. The lens L_3 focuses the beams on two, diametrically opposed, points of the objective back focus plane. In such a way a small region of the specimen (15×15 μm^2) is illuminated by two parallel laser beams and a standing wave is established at the glass/water interface.
The light scattered from the probe is collected by the same objective, focused by the lenses L_4, L_5 and L_6 and measured with a bandwidth of 500 kHz by the avalanche photo-diode (`Hamamatsu`). Data are acquired using an analog-digital converter (`Keithley instruments`), with a sample rate of 150 ksample/s. Experiments are carried out using an Olympus 60X TIRF objective (numerical aperture 1.45) and a CCD camera. The laser source is a second harmonic YAG (`Coherent Verdi`), with a wavelength λ_0=532 nm and a longitudinal coherence length of several hundreds meters. Its power is typically set to 500 mW. In our experiments the period of the fringes (Eq. 1) is typically $\sim$200 nm and the field depth ζ is around 100 nm.
The resolution of this setup has been calibrated using a piezoelectric stage, which moves a bead stuck to the coverglass. Fig. 2a shows the signal from a 100 nm bead, moving through the fringes: the piezo-electric stage makes a step of 10 nm every 200 ms. The motion is reflected by sudden variations of the intensity and the bead position can be calculated from Eq. (3). Fig. 2b shows the bead position: it can be determined with an accuracy of $\sim$ 1 nm at that time scale (66 μs/sample). This accuracy is limited by the mechanical instabilities, whose power spectrum is plotted in Fig. 3. The noise spectrum is dominated by a broad peak around 300 Hz, whose the amplitude is few nanometers. The main reason of this noise is the extreme sensitivity of the optical setup to the vibrations. In fact, we observe that the noise is strongly reduced near the sine antinodes, where
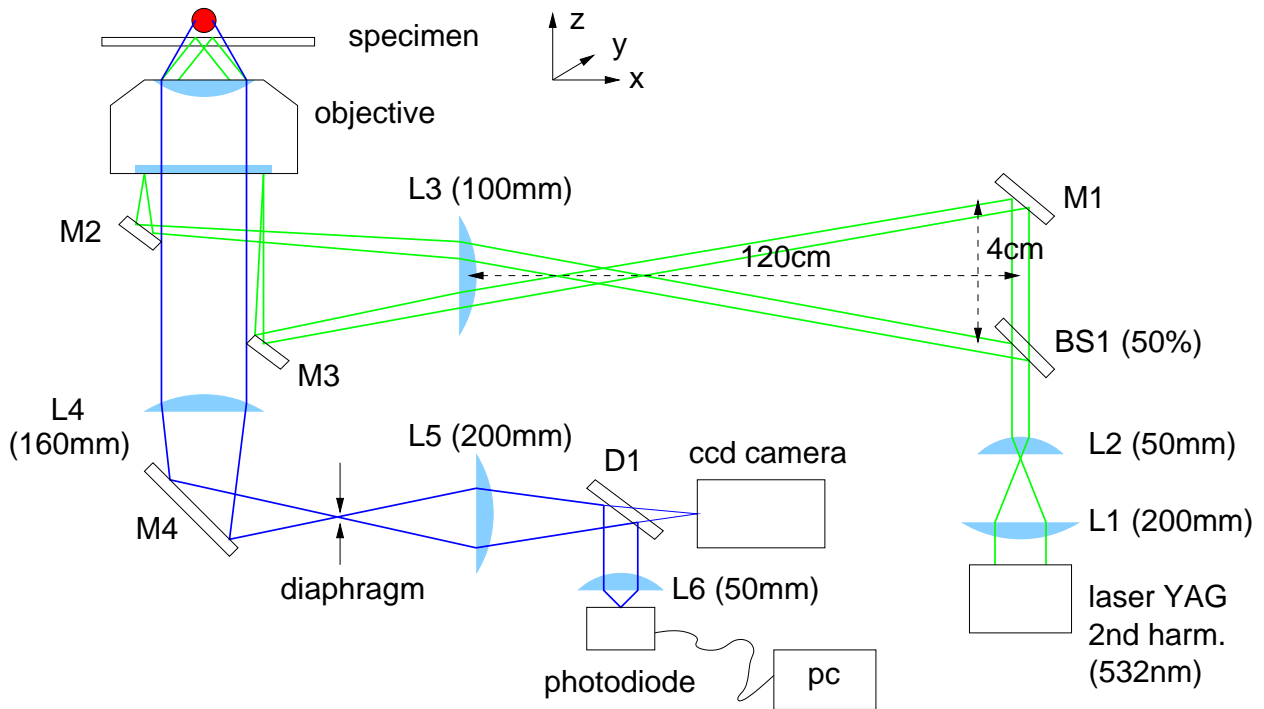
Figure 1: Schematic of the experimental setup. Interference Total Internal Reflection configuration through the TIRF objective. The two laser beam are focused on the diameter of the back focus plane, in order to light the sample with two parallel beams with opposite wave vector k_x. The light scattered by the probe is recorded using an avalanche photo-diode.

the positional resolution decreases. The signal/noise ratio should be improved by reducing the acoustic noise and by stiffening the setup.

3   Material and Methods

Nkin460GST is expressed in BL21 (DE3) *E. coli* and is purified as described in [11, 12].

Preparation of the microtubules : The MAP-rich tubulin (from Cytoskeleton ML113) is polymerized at a concentration of 2 mg/ml (concentration adjusted with BRB80 : 80 mM K-PIPES pH 6.9, 1 mM MgCl$_2$, 1 mM EGTA, 1 mM GTP) and at a temperature of 37°C for 30 minutes. Then taxol is added at a final concentration of 20 μM to stop the polymerization-depolymerization process for 20 minutes at 37°C. Unpolymerized tubulin is removed by centrifugation at 30000 RPM, for 15 minutes at 30°C. The pellet is rinsed and resuspended with BRB80 supplemented with 10 μM taxol).

Bead assay : Microtubules are introduced in a flow chamber and incubated for 3 minutes to adsorb on the poly-L-lysine-coated coverglass. The chamber is rinsed with BRB80+taxol and with filtered casein solution (2 mg/ml) to avoid non specific interaction. After a last wash with MOPS buffer (MOPS 20 mM - pH 7.2, MgCl$_2$ 5 mM, NaCl 200 mM), the kinesin-coated beads are mixed with the motility buffer and injected into the chamber. Kinesin-coated beads are prepared in mixing highly diluted preparation of kinesins (diluted in filtered casein at equimolar concentrations) and carboxylated latex beads (200 nm diameter; Polysciences cat 07304). After 5-10 minutes incubation, filtered casein at 1 mg/ml is added to this mix. The motility buffer is MOPS buffer supplemented with 20 μM ATP, 1 mM phosphocreatine, 50 μg/ml creatine phosphokinase and an oxygen scavenging system (3 mg/ml glucose, 100 μg/ml glucose oxidase and 20 μg/ml catalase),
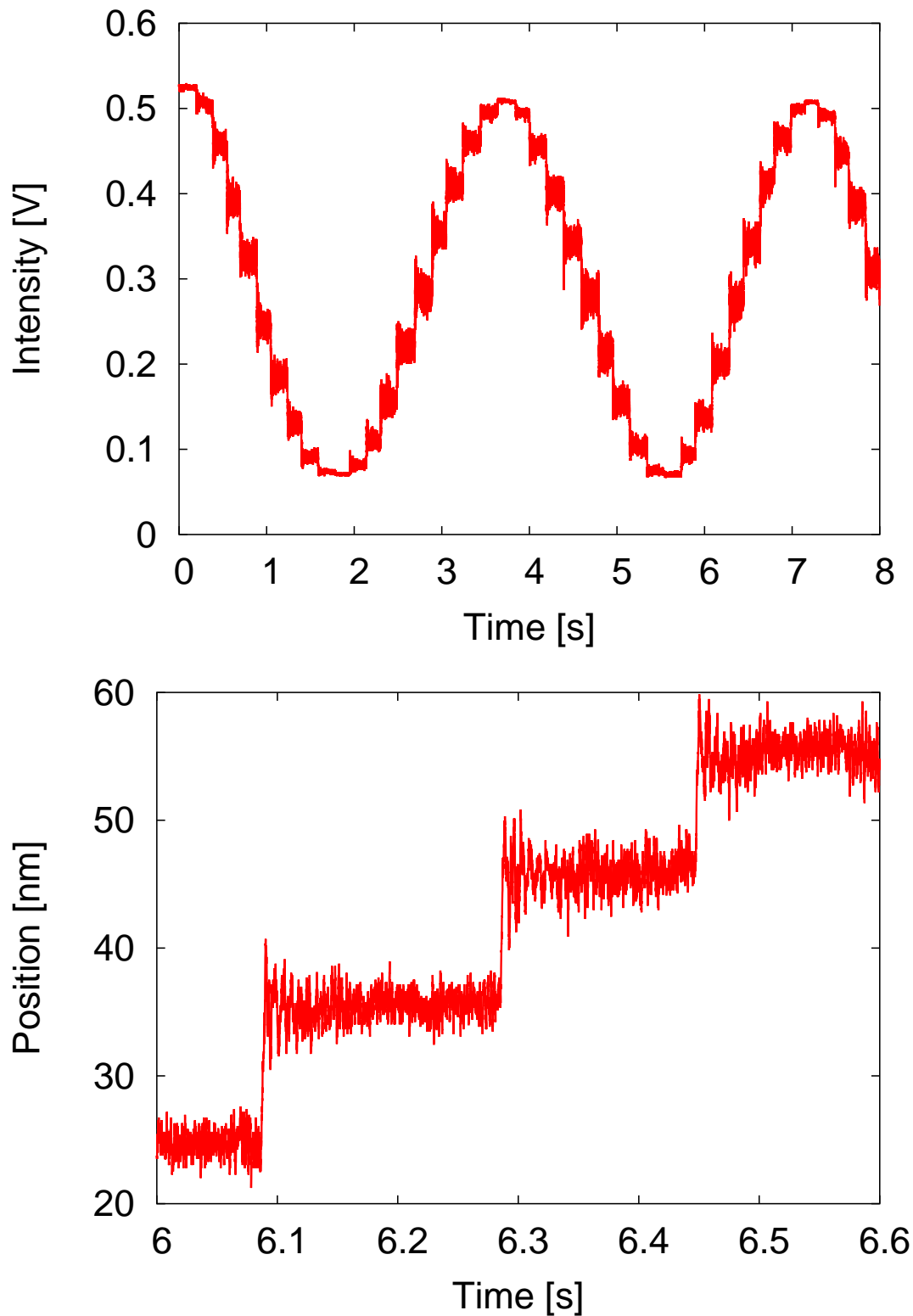
Figure 2: A single colloid (diameter 200 nm) is fixed on the coverglass, while the sample holder is moved in 10 nm steps by means of a calibrated piezoelectric stage. The bead blinks moving through interference fringes in the near field and its luminosity is recorded by the Photo-detector. The data are acquired with a time resolution of 60 μs. **a**) light intensity (in arbitrary units) collected by the avalanche photodiode. **b**) the position (in nanometers) of the bead is calculated from this luminosity according to Eq. (3).
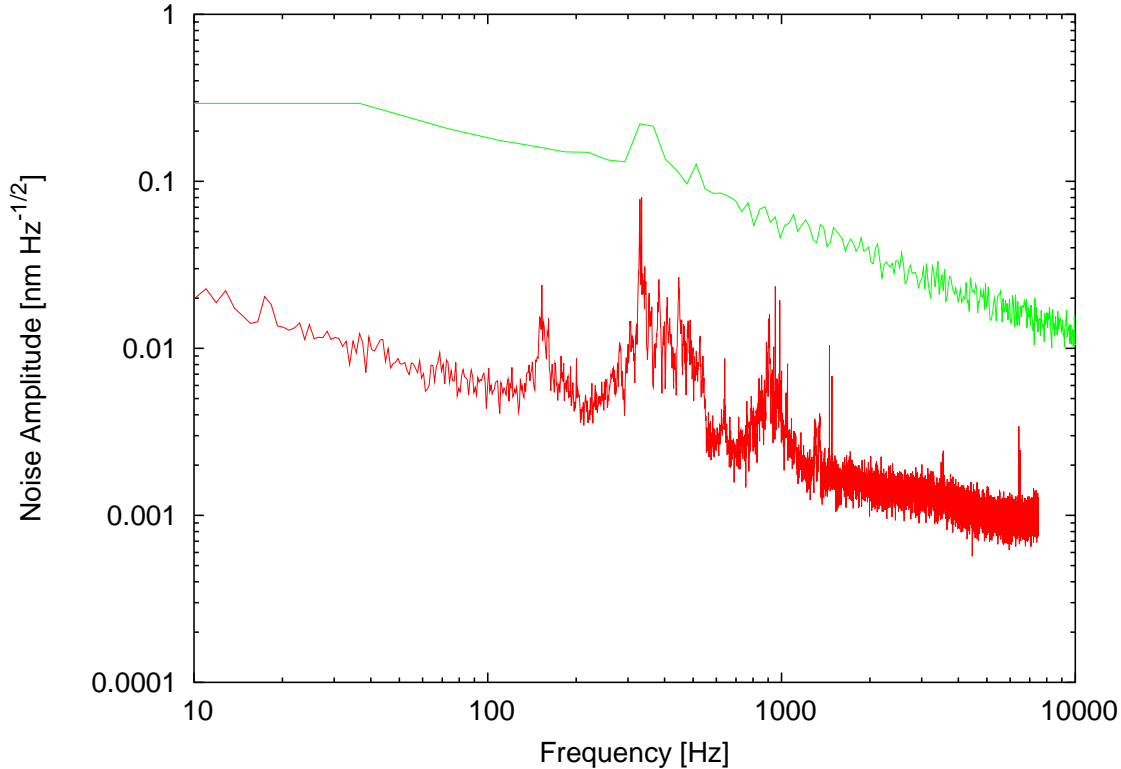
Figure 3: Noise Power Spectrum of a bead stuck on the coverglass (bottom line), driven by a piezoelectric stage. Top line is the noise power spectrum calculated for a bead linked to a kinesin.

2 mM DTT. The coverslip is sealed with VALAP (vaselin, lanolin, paraffin at 1:1:1). Observations are performed at room temperature (approximately 25°C).

In order to obtain a sufficient low concentration of kinesins per bead, we worked with a concentration just above the threshold under which we do not observe any motility events.

4   Results

When the microtubules are injected into the observation chamber, they are partially aligned by the flow. We take advantage of this effect to roughly line them up, perpendicularly to the interference fringes. The kinesin-coated beads come randomly in touch with the microtubules and they move along of them through the interference pattern.

As described previously, the bead blinks moving through the interference pattern and its luminosity is measured. Fig. 4a shows the intensity $I(t)$ as a function of time (dots ·), acquired with 6 μs of time resolution. In order to improve the signal/noise ratio, the signal is smoothed by performing a sliding average with a time average of 5 milliseconds (continuous line in Fig. 4).

A comparison between Fig. 2a and Fig. 4a highlights a qualitative difference: the curve measured with the molecular motor is more noisy than the one obtained with the piezo-electric stage. This observation seems to be obvious, because the bead-kinesin link has a finite stiffness, and the tethered bead wiggles because of the brownian motion. In Fig. 2a we also observe that the noise decrease significantly around the minima and the maxima of the intensity. Of course, the signal from a bead wiggling only in the xy-plane (the plane parallel to the surface) should appear less noisy near the sine antinodes, where the lateral sensitivity (along the x-direction) is reduced. Conversely, in Fig. 4a the noise does not significantly depend on the x-position. This difference can be explained taking in account the bead motion in the evanescent field perpendicularly to the surface
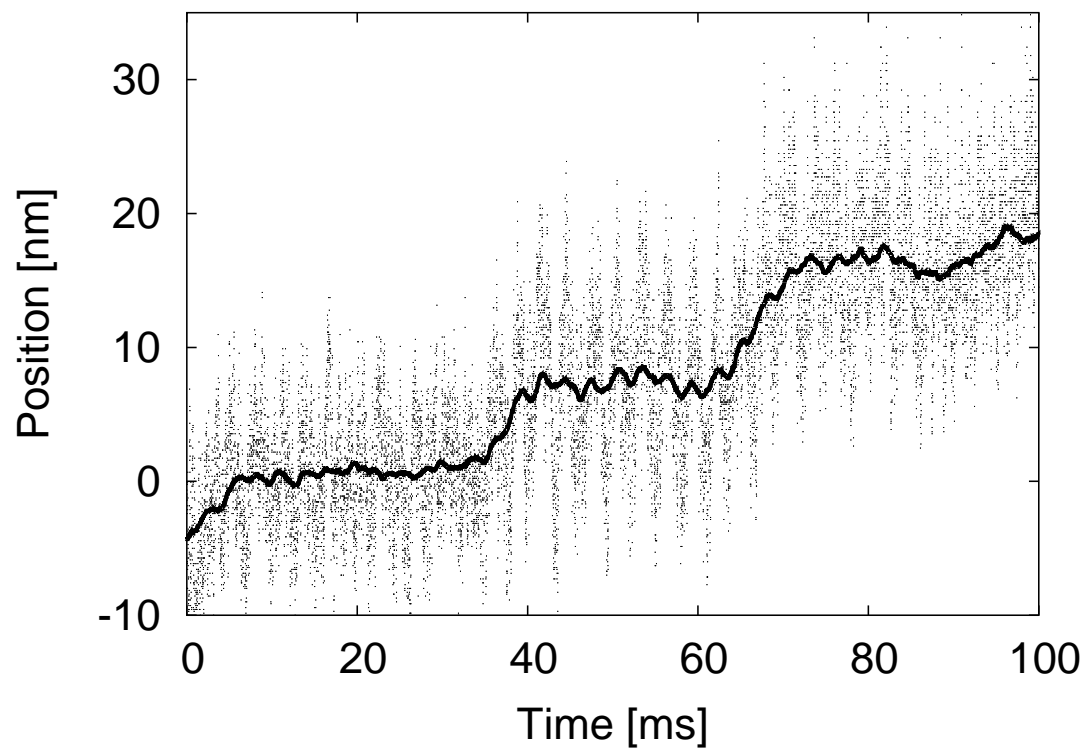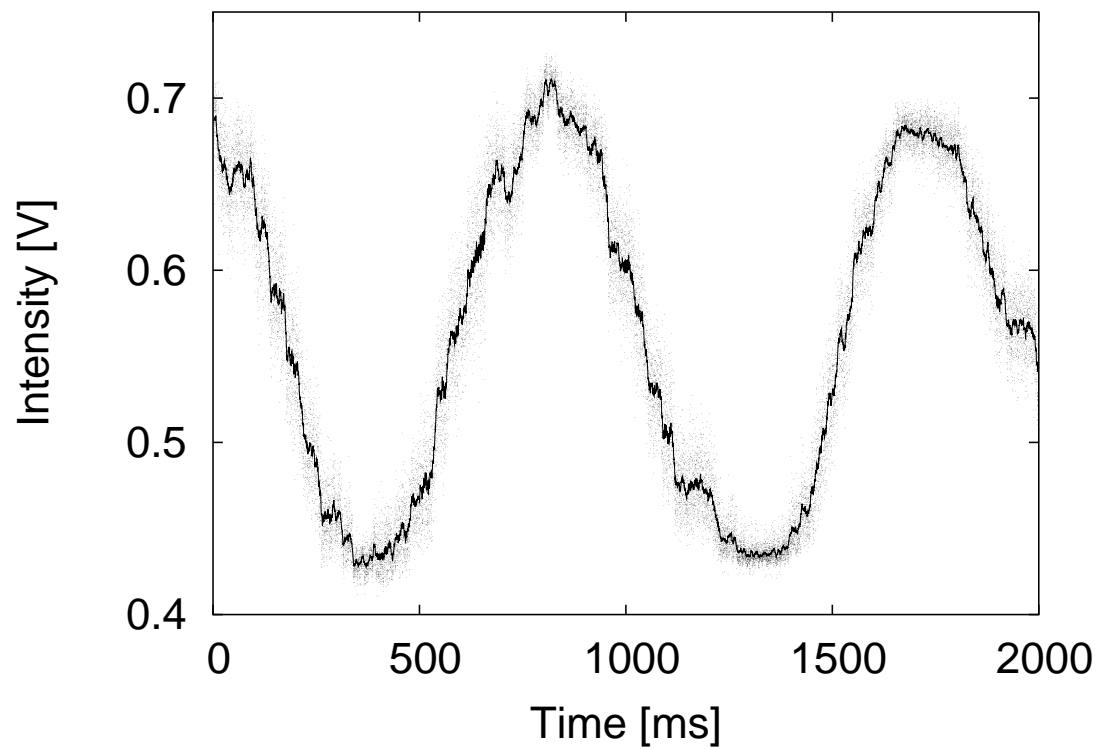
Figure 4: Measured intensity as a function of time. **a**) Data recorded with time resolution of 6 microseconds (red line) and smoothed with a 5 milliseconds sliding average. Notice the discontinuities in intensity variations, corresponding to the kinesin steps. **b**) Bead position as a function of time, calculated from a portion of curve *a* according to Eq. (3).

(along the z-direction). This suggests that a significant part of the noise is due to the brownian motion in the direction perpendicular to the surface.

The analysis of the brownian motion of the tethered bead supplies an important information about the bead response time. This time determines the physical limit of the time resolution of the experiment. It can be determined by performing the *intensity-intensity* autocorrelation function, as described in Ref. [9, 10]. The autocorrelation analysis (data not shown) indicates that the temporal resolution is $\sim 70\mu s$ in our experiments.

Fig. 4b shows a detail of curve **a**, where the bead position $x(t)$ has been calculated according to Eq. (3), over intervals of half a period. We observe the typical motion of kinesin, with discrete 8 nanometer steps and plateaus alternately. These observation is consistent with the previous experiments [3, 11].

The data are too noisy to determine the step details. In order to overcome this problem, several steps are averaged. Of course, different steps can be averaged only after synchronization and the accuracy in synchronization determine the final resolution of the mean step and it is basically limited by the signal/noise ratio of the data. We first proceed by an automatic detection of the 8-nanometers steps, using the algorithm proposed by D. Smith [13], then each step is fitted to a step-like function (Eq.4) to determine its middle position t_0 (Fig. 5a) and its rising time α^{-1}.

$$\Theta(t) = \frac{8}{\pi} \arctan \left[\alpha(t - t_0) + \frac{\pi}{2} \right] \ nm \tag{4}$$

Steps with a rising time > 1 ms are considered anomalous and are discarded.

Fig. 5b shows the kinesin step after averaging: the noise is appreciably reduced and the step time can be evaluated. As the step ends are not well defined, we measure the time necessary to move from 10% to 90% of the total step height. The experimental data supply a step time $\tau_s = (60 \pm 10)\mu s$ to move from 0.8 nm to 7.2 nm. Of course, the actual movement of kinesin might be faster, because τ_s is essentially limited by the bead response time and to our ability to synchronize the steps before averaging. However, the measured time τ_s is an upper limit to the step time in absence of external load. This result is consistent with earlier experiments [8] performed under load. In addition, we also notice that, at this time scale and within the experimental uncertainty, there is no evidence neither for an internal structure of the step nor for sub-steps.

5 Conclusion

We present a simple and versatile technique to track a small bead, with nanometer precision and a time resolution of six microseconds. This technique is used to characterize the motion of kinesin-coated beads along of microtubules, in absence of external forces.

The typical kinesin motion, with discrete 8 nanometer steps and relatively long dwell times, is observed. At the same time, the brownian motion of the bead, wiggling around the kinesin position, hides the short-time details of the steps. In order to overcome this effect, and improve the signal/noise ratio, we develop a simple algorithm to line up the steps and average them. In such a way, brownian motion is averaged over many events and the final resolution is only limited by the number of available steps and the bead response time, which is $\sim 70 \ \mu s$ in our experiments.

Summarizing, we can state that the kinesin takes less than 70 μs to make 8-nanometer steps, and that at this time scale no fine step details are observed.

Acknowledgments
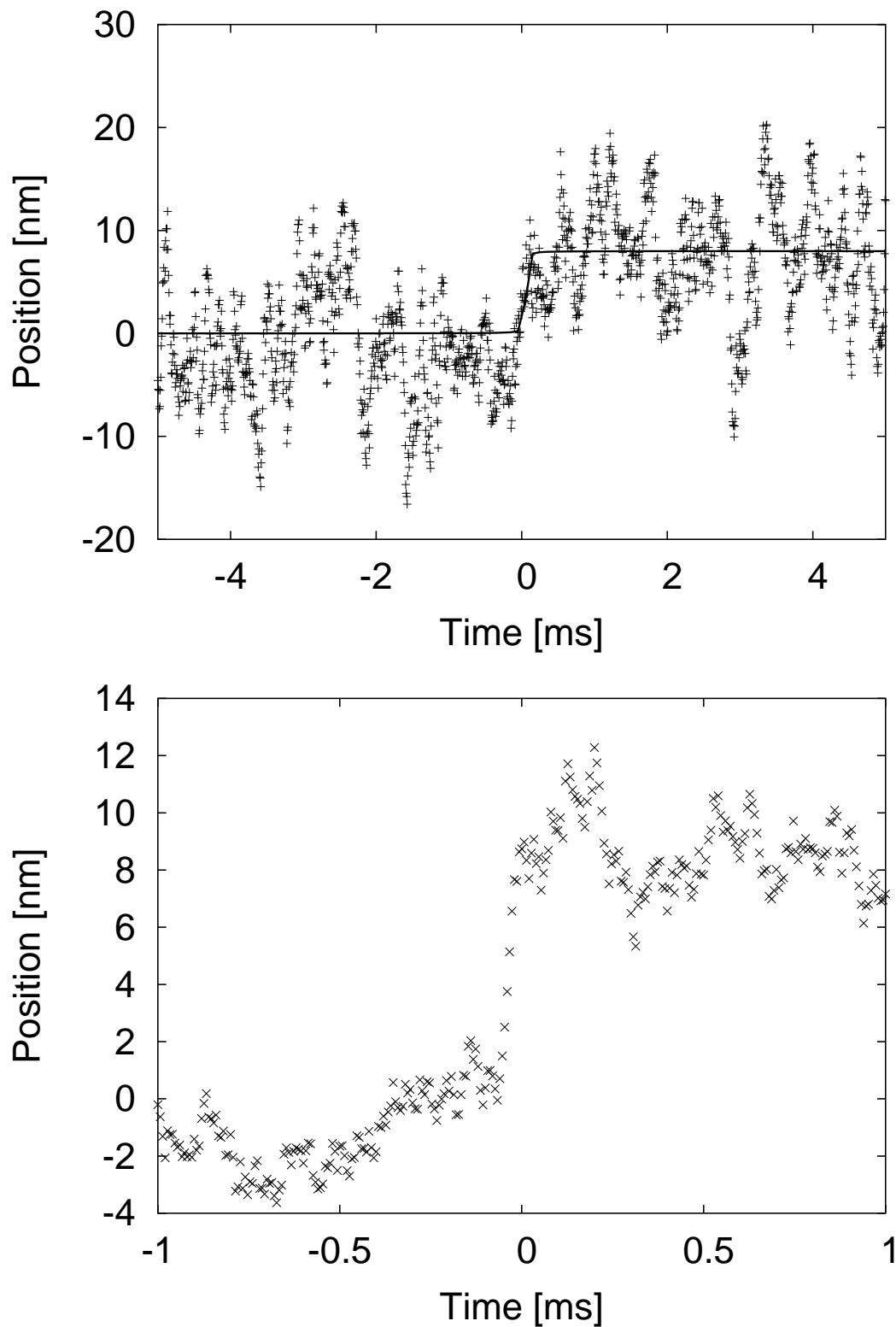
Figure 5: Analysis of the rising time of the kinesin step: **a**) A single 8 nm step step: the continuous line is the best fit between data and Eq.(4). **b**) the average 8-nm step: the rising time τ_s is lower than 70 μs.

References

[1] Y. Inoue, A. H. Iwane, T. Miyai, E. Muto, and T. Yanagida, *Biophys. J.* **81**, 2838 (2001).

[2] S. A. Endow and H. Higuchi, *Nature* **406**, 913 (2000).

[3] K. Svoboda, C. F. Schmidt, B. J. Schnapp, and S. M. Block, *Nature* **365**, 721 (1993).

[4] H. Higuchi, E. Muto, Y. Inoue, and T. Yanagida, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4395 (1997).

[5] M. J. Schnitzer and S. M. Block, *Nature* **388**, 386 (1997).

[6] M. Nishiyama, H. Higuchi, and T. Yanagida, *Nat. Cell. Biol.* **4**, 790 (2002).

[7] K. .Visscher, M. J. Schnitzer, and S. M. Block, *Nature* **400**, 184 (1999).

[8] M. Nishiyama, E. Muto, Y. Inoue, T. Yanagida, and H. Higuchi, *Nat. Cell Biol.* **3**, 425 (2001).

[9] G. Cappello, M. Badoual, L. Busoni, A. Ott, and J. Prost, *Phys. Rev. E* **68**, 021907 (2003).

[10] M. Badoual, G. Cappello, K. Zeldovich, and J. Prost, *European Physical Journal E* p. In press (2004).

[11] I. Crevel, N. Carter, M. Schliwa, and R. Cross, *EMBO J.* **18**, 5863 (1999).

[12] S. Lakamper, A. Kallipolitou, G. Woehlke, M. Schliwa, and E. Meyhofer, *Biophys J.* **84**, 1833 (2003).

[13] D. A. Smith, *Phil.Trans. R. Soc. Lond. B* **353**, 1969 (1998).

Integration and Visualisation of Textual and Factual Genomic Data

Nicolas Férey, Pierre-Emmanuel Gros, Joan Hérisson & Rachid Gherbi

LIMSI-CNRS, Université Paris-Sud XI, BP 133, 91403 ORSAY CEDEX, France.
e-mail : {ferey, gros, herisson, gherbi}@limsi.fr, http://www.limsi.fr

Abstract

Biologists are now in front of the result of a large work on the genomes (sequencing, alignment, transcriptome). The genomic information resulting from this, show however some characteristics that make them very difficult to interpret and to exploit. They are stored in huge amount of data, are heterogeneous, and are geographically distributed. These data are recorded in structured or semi-structured formats within public or private databanks, and constitute an important factual data source (GenBank, SwissProt, or Decrypthon). But, genome knowledge could not be limited to DNA or protein annotated sequences. Indeed, there is a significant quantity of information relating to these genes, recorded in an unstructured format within millions publications (Medline, PubMed). This paper presents GenomeExplorer, a new modeling and software solution to integrate textual and factual genomic data based on adapted federator description format. Moreover, GenomeExplorer offers us a user-friendly visualization of this information within an immersive environment. The visualization is based on a well-adapted graphical paradigm that automatically helps to build a graph-based representation. This solution allows biologist to suitably explore huge sets of genomic data, but it could be applied to other application fields. This kind of graphical exploration has the advantage to highlight some global topological characteristics, which are uneasily visible using traditional exploration tools. Finally, some results produced by GenomeExplorer software on various sets of biological data, are presented.

1   Introduction

The *ADN-Viewer* [8] [9] software offers us to visualize the complex spatial trajectory of huge naked DNA augmented by genetic annotated information provided by *GenoMedia* [10] platform. The third dimension in DNA visualization, allows biologists a new representation to go from genetic approach (studying targeted genes) to a genomic ones (studying whole genome). In this kind of immersive visualization, the biologists interact more directly and intuitively with their objects of interest. The fact to visualize a whole annotated genome, thanks to *ADN_Viewer*, shows the possibility of exploring large genomic data within an immersive environment. However a greatest part of genome knowledge is not in annotated DNA or protein sequences. Indeed, there is a significant quantity of information relating to these genes, stored in unstructured manner in the million genomic publications.

Our objective is to elaborate new solutions in order to virtually explore both textual and factual genomic data. The textual data result from *Medline* [16] and factual ones come from the many other databases, such as *GenBank* [14], *SwissProt* [15], or *Decrypthon* [13].

In this paper, we firstly situate this work compared to existing ones, in order to highlight the interest of our approach. This approach is mainly based on the definition of a genomic data federator language, answering the requirements and specificities of genomic databases. This language must take into account the heterogeneity resulting from textual and factual data. Then we explain the representation methods to view these data within an immersive framework. Finally, we present some results produced by *GenomeExplorer* software on various sets of biological data.

2 Situation

A synthetic and short state of the art relating to our problems is difficult to do, because the field is on the crossing point of three main ones: automatic text processing, heterogeneous data integration, and visualization. However, the focusing on genomic applications restricts this work. So we only present here below the references that are on the intersection of the former fields.

2.1 Textual information extraction

J. T. Chang and S. Raychaudhuri [6] have written a clear tutorial about the problems concerning the natural language processing (NLP) in the specific Bioinformatics field. Several problems of the NLP are particularly present in the publications biology field, such as these characteristics:
- A very frequent use of the acronyms.
- A strong polysemia in the biological terms and their acronyms.

It is necessary to analyze the context to solve ambiguity on these acronym or terms. In *ACROMED* [4] system developed by J Pustejovsky, the author proposed a solution to automatically build biological acronyms databases. Moreover in the purpose to extract semantic relations between the biological entities from text (such as positive retroaction), we need to efficiently solve the anaphora problem [5].

2.2 Genomic factual data exploration

Two main obstacles prevent from doing a simple exploration in factual biological data:
- The variability and heterogeneity of the genomic data.
- The huge genomic data mass that permanently grows.

The huge mass makes difficult the selection, visualization and interpretation of these data. In addition, heterogeneity and multiplicity of data sources reduce transparent integration for the user, within applications such as *GenoMEDIA* [10] or *GenoStar* [18].

2.3 Genomic data visualization

2.3.1 ADN-Viewer

The main characteristic of *ADN-Viewer* [8] [9], compared to other visualization applications, lies in the fact that such visualization is carried out within an immersive environment. Moreover, it allows us to apprehend whole genome, but also by fragments, as in the majority of the DNA visualization tools. The immersive environment facilitates also collaborative work: several people can interact with the objects while discussing. The third dimension offers users additional degrees of freedom, and helps the use of concepts that are not revealed in two dimensions. It is important for us to preserve *ADN-Viewer* user-interface modalities, in order to build coherent and integrated software dedicated to the analysis of genomes within an immersive environment. Finally, the immersive

exploration of textual and factual data must take into account the biologist needs. Many other visualization tools exist focused on proteins such as **Rasmol** [18], **VMD** [19]. However, the protein data are relatively small and do not need a complex processing.

2.3.2 Sequence Word

The closest work that could correspond to our approach concerns the representation of factual genomic databases in a virtual reality context [1]. In order to use the human orientation capabilities, the authors present a solution where genes are positioned in a landscape ground, allowing actors to explore the data. The genes can be gathered spatially according to several clustering criteria, by using **BLAST** [11] algorithm calculating an alignment score. In the virtual world, a distance between objects represents this score. The paper highlights the following points:

- The approach of immersive visualization brings new knowledge, which is latent in the data, and difficult to extract without adequate representation.
- The representation facilitates the modeling of the crossed references, frequently used in biological data, thanks to additional freedom degree in the immersive environment.
- It is significant for biologist to represent certain data credibility in this space.

2.3.3 BioBiblioMetrics

B. J. Staplet and G Benoit [2] propose a 2D visualization prototype of textual and factual biological information. The genes are extracted from the **Medline** [16] documents starting from **MeSH** [17] thesaurus terms. In this paper, they propose to use the mutual information coefficient to measure bibliographical proximity between two genes names or aliases. This genes bibliographical information is visualized by a graph. A node represents each gene; the minimal bibliographical relation (user-defined co-occurrence threshold) between two genes is represented by an edge. Finally the length of these edges represents the proximity degree between each pair of genes. The genes references, found automatically by the system in the factual database, are used to increase or to decrease edges values. The integration of the textual and factual data is treated with an interesting way, but it would be interesting to look further the linguistic analysis of textual documents.

3   A federator genomic data description language

In order to efficiently integrate textual and factual data, we need to define a common format that must accommodate and integrate knowledge resulting from structured databanks (**GenBank** [14], **SwissProt** [15], or **Decrypthon** [13]), as unstructured information extracted from texts relating to genomic publications (**Medline** [16]). We describe in this section how we used the specific characteristics of the factual and textual genomic data to find an adapted description format for this kind of data.

3.1   Factual genomic databases specificities

The factual genomic databases are very heterogeneous (format or quality), but involve some specific characteristics. Indeed, they are often focused on biological object of interest (protein, gene...), described by an attribute set. Moreover, these objects are often compared one to another by a measurement (sequence alignment score, functional similarity…).

3.2 Requirements in text information extraction

It seems furthermore that the most interesting textual data for biologists are the interaction relationships between biological objects (genes expression, genes regulation, proteins interaction…). So, formal modeling as binary relationships between biological objects, as well as in factual databases case, comply with biologist's requirements. For example, using text corpora, they could extract co-occurrence relationship between two biological terms, or more specific semantic relationships (like inhibition).

3.3 Definition of federator data format

In both above cases, these binary relationships are valuated, by co-occurrence measurements or alignments (numerical values), or by semantic relations extracted from texts (symbolic values). It

```
<data>
        <object id ="0" value0 ="" ... valueN=""/>
        <object id ="1" value0 ="" ... valueN=""/>
        ...
        <relation id1="0" id2="1" value0 ="" ... valueN=""/>
        <relation id1="0" id2="1" value0 ="" ... valueN=""/>
        ...
</data>
```

seems that this format, which is based on the concept of multi-valuated objects and relationships, is particularly adapted to describe both textual and factual genomic data.

These values can be numeric, symbolic, or textual (Table 1). So we define a format with the following characteristics: the identifiers represent objects. The objects and relationships are associated to a values list, which characterizes them. The values can be textual (label, genetic sequence…), numeric (like co-occurrence score), or symbolic (type of interaction, like positive retroaction…).

	symbolic	numeric	textual
Object	kind of sequence (protein, DNA) function	size, entropy, curvature	textual sequence, label, id, term name
(binary) Relationship	*textual relation extraction* (inhibition, interaction…)	*co-occurrence,* alignment	consensus sequence

Table 1: Genomic data translation the federator language (textual data in italic)

4 Representation modalities in an immersive framework

We recall the goal of this work: find a genomic data exploration system, as much textual as factual. Even if the characteristics of the federator format allow us to integrate these two kinds of data, it remains to define a visualization paradigm. This paradigm must be adapted at the same time to this format and to the user's needs. We present in this section how we map the data to a visual representation.

4.1 Graph visualization

The selected federator format is a list of objects with their binary relationships. These data may be interpreted as a graph, where biological objects are nodes and relationships between us are edges. So we choice to visualize data by 3D graph. The advantage of this system is the visualization has no reality references (on the contrary to metaphoric representation) and is independent from data.

4.2 Visualization description format

Visualizing data by a 3D graph results from the following motivation: invent a system that ensures independence between data and their representation. This motivation was transformed into requirements because of the data source format heterogeneity. Indeed we didn't want to import this heterogeneity in the visualization system. However it remains to choice how graphically represents each value (both object and relationship values) within a 3D graph. In order to map dynamically data to their 3D graph representation, a language was defined. To do that, we started to build a short inventory of the graphic 3D object characteristics.

Federated data / Graphic attributes	Object (Node)		Relationship (Edge)	
	Symbolic	numeric	symbolic	numeric
Placement		x, y, z		
Size		dx, dy, dz		weight
Length				length
Color	pink, green, purple…	r, g, b	pink, green, purple…	r, g, b
Shape	sphere, cube		line, cylinder	
Transparence		a		a

Table 2: Graphic characteristics summary which may be used in description format

Taking into account the description data format, these graphic characteristics can be classified in the three following groups: symbolic, numeric, or both. So within 3D graph visualisation, numeric object values may be represented by numeric graphic properties, like node placement, node size, node color or node transparency. Numeric relationship values may be vizualized by edge length, edge weigth, edge color, or edge transparency.

```
<view>

        <object objectorm="sphere" color="blue"> (default color and shape)

                <label map="value1"/>
                (value1 (textual) is the node label in visualisation)
                <size map="value2"/>
                (value2 (numeric) is the node size)

                <objectcolor map="value3">
                        <objectdomaincolor value="human" mapto="white"/>
                        <objectdomaincolor value="mouse" mapto="black"/>
                </objectcolor>
                (la value3 (discrete : {human,mouse}) is the node color, according to
                defined domain values)
        </object>

        <relation relationform="line" color="green"> (default color and shape)
                <label map="value1"/>
                (value1 (textual) of the relations is an edge label in visualisation)

                <length map="value2"/>
                (value2 (numeric) of the relations is the edge length in visualisation, e.g.
                distance between node)
                ...
                <a map="value3"/>
                (value3 (numeric) of the relation is edge transparency in visualisation)
        </relation>
```

Each symbolic values may be represented by a predefined shape or color, according to the kind of value (object or relationship value). Finally, the textual values may be visualized by a label within the 3D graph representation.

4.3 Graphs representation problems

4.3.1 Edge-valued graphs representation problems

We see in the preceding section that an edge length, within the 3D graph representation, can represent numerical relationships values. We are faced of the following problem: mapping numeric data value to an edge length (distance between two graph nodes) adds new geometrical constraints in 3D Euclidian space. Sometimes these new constraints cannot be solved. For example, the Figure 1 represents a graph with the wished distances between its edges. This triangle can never be drawn on 3D Euclidian space respecting the wished distance constraints between its edges.
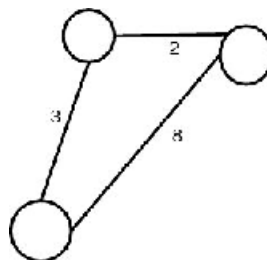
Figure 1: Distance constraints without graphic solutions

An approach, proposed by Eades [12], consists in simulating two kinds of force between each graphic object. In order to place two nodes that are in relations respecting distance constraints, he proposed to apply them an attraction force, to minimize the difference between an initial random distance in the graphic world, and the wished distance. Moreover, a repulsion force is applied on two close nodes, which are not in a relation. After several iterations, this dynamic property allows

system to converge into a satisfactory solution where all the distances are as closed as possible than desired edges lengths.

4.3.2   Huge graph visualization problems

The main disadvantage of above approach is its complexity. Each node reacts to the presence of all its connected neighbors by an attraction force, and moves according to the presence of all the other nodes per repulsion. The complexity strongly decreases by applying a visibility threshold on the not connected nodes. All the nodes, which are too much far according to this fixed threshold, do not repulse. This complexity decreasing is not enough efficient, in order to represent huge biological data (a million nodes or edges). In front of this kind of data (***Decrypthon*** [13]), it was necessary to use segmentation algorithms in order to make some partitioning on the graph. However, in very huge graphs, classical clustering algorithms on graph are unsuitable.

We firstly use an algorithm to extract connected components. Indeed, this is the easiest way to segment huge graphs, where the topological structure is generally unknown. However, the result obtained on ***Decrypthon*** was disappointing: the ***Decrypthon*** graph is relatively connected, except some satellite nodes. So this clustering was insufficient to really decrease complexity. But we have a first result on the ***Decrypthon*** topology.

Another algorithm to detect articulation points was used to segment the graph in its bi-connected components. When an articulation point is removed from a connected graph, this one becomes not connected. This clustering was then tested on a part of ***Decrypthon*** and allowed us to extract several bi-connected components (see Figure 4 and Figure 5). Each component has only single interaction with the others, so nodes placement in each component could be much more quickly processed. The components are then placed in Euclidean space and connected by their articulation point. A complementary approach using Kruskal's algorithm, extracts a maximum spanning tree from a graph. The force-directed placement complexity depends on edges number. Graphs like ***Decrypthon*** [13], contain million of edges, so it is necessary to use an effective method in order to place the nodes. The Kruskal's algorithm extracts the weightiest edges and preserves the graph connectivity. The force-directed placement process is then applied on this maximum spanning tree. This method makes it possible to deal with huge graphs, but this is a weak approximation, if the edge value is not a transitive function.

5   Results

We present in this section some examples about results obtained with ***GenomeExplorer*** for exploring factual and textual biological data. The first example deals with block duplications between the 16 *S. cerevisiae* chromosomes. The second example concerns with data from public databases, like ***Decrypthon*** [13]. In the last example, the data come from textual processing on ***Medline*** [16] papers abstracts.

5.1   Factual data: Yeast gene block duplications

The first results show the visualization of block duplications in the Yeast chromosomes. In this experiment, each object is one of the two Yeast chromosome arm. For instance, *1R* is the right arm of the first Yeast chromosome, and *2L* is the left arm of the second chromosome... For each object, the value *value1* gives the chromosome name; the value *value2* gives the chromosome size, the value *value3* gives right or left arm. In each binary relationship between chromosomes, the value *value1* is the number of same blocks shared by two chromosomes (if this number is null, this relationship is not created).

```
<data>
  <object id="0" value1="1R
          value2="0.2" value3="right"/>
    ...
  <object id="31" value1="16L"
          value2="0.9" value3="left"/>
    ...
  <relation id1="2" id2="6" value0="14.0" />
    ...
  <relation id1="27" id2="29" value0="8.0" />
</data>
```

```
<view>
  <object objectform="cube" color="red">
    <label map="value0"/>
    <size map="value1"/>
    <objectcolor map="value3">
      <objectdomaincolor value="left" mapto="red"/>
      <objectdomaincolor value="right" mapto="blue"/>
    </objectcolor>
  </object>
  <relation relationform="line" color="green">
    <length map="value0"/>
  </relation>
</view>
```

Data translation

Representation settings
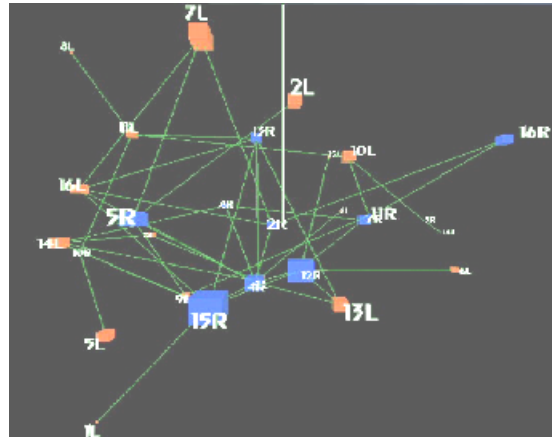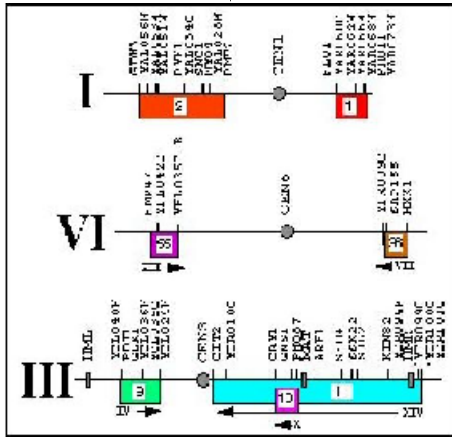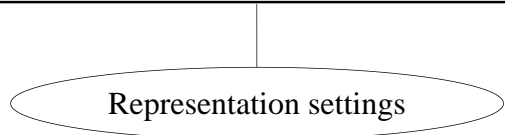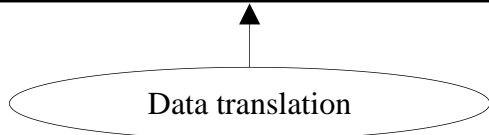
Figure 2: Partial traditional 2D visualization
(3 of 16 Yeast chromosomes)

Figure 3: Immersive synthetic visualization
(32 Yeast chromosome arms)

5.2 Factual data: a huge genomic set from Decrypthon

Decrypthon [13] is a huge protein alignment network, where each protein is compared to another (500000 proteins). Here is a sample of the data available on the *Decrypthon* web site.

ID=GM_0341259 550 405 438 0 0 ID=GM_0139128 261 43 76 0 0 51 34 11 16 34 26 43 5.55 F922A317DB8B53AE
ID=GM_0524625 314 141 194 0 0 ID=GM_0507857 1778 948 1001 0 0 62 54 19 27 54 29 46 6.42 AB4A23F6531FF818
ID=GM_0491325 99 46 85 0 0 ID=GM_0000235 324 67 106 0 0 45 40 10 20 40 21 40 5.18 E42839C174C39C17
ID=GM_0278014 123 23 97 1 8 ID=GM_0027567 607 345 427 0 0 58 83 22 33 75 26 46 5.39 47D9C19CD99FB3C4
ID=GM_0348030 146 48 114 1 2 ID=GM_0048942 796 552 620 0 0 53 69 14 32 67 25 42 6.24 AA4747D4EBC53C7A
...
ID=GM_0029820 705 444 647 1 13 ID=GM_0258468 1162 929 1127 2 18 139 217 50 93 186 31 54 23.97 378C78B1D558D830
ID=GM_0388572 362 313 353 1 1 ID=GM_0433870 101 6 47 0 0 58 42 16 24 41 23 44 5.50 8D7919A24AED51ED
ID=GM_0388572 362 27 289 3 8 ID=GM_0387259 273 2 264 4 8 313 271 85 133 255 27 52 60.83 16DB267495B93023

These data are interesting but too huge to be correctly exploited (39 Go). A work about large graph management was carried out in order to visualize this kind of data. Here is an output of the first results.
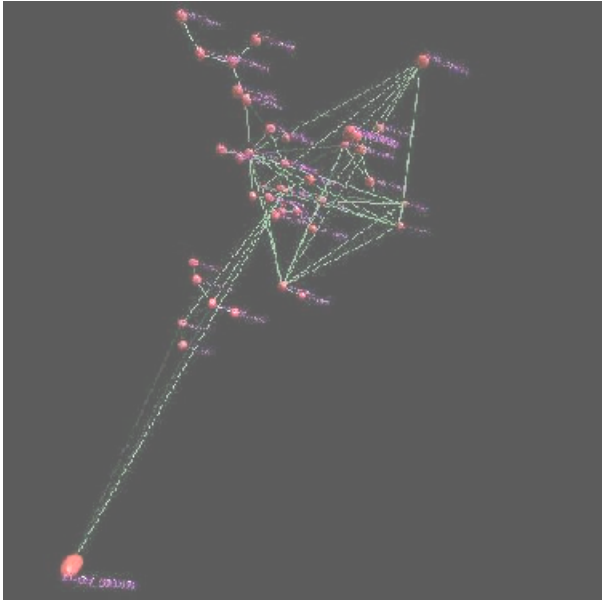
Figure 4: Global proteins sequences alignment of a bi-connexe component of *Decrypthon (1)*
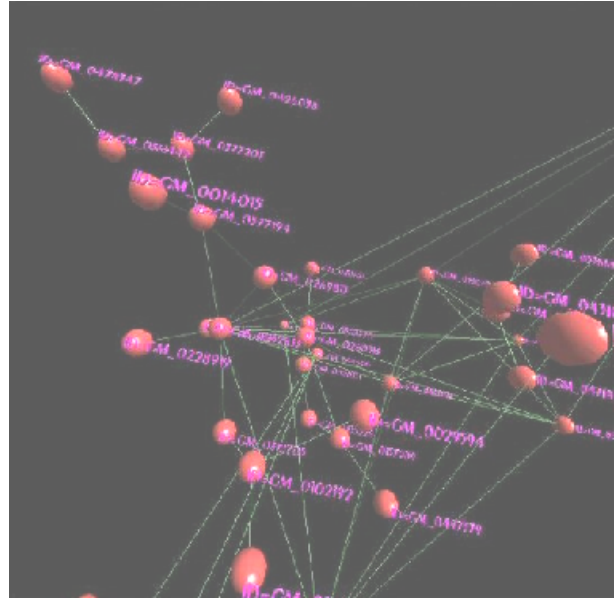
Figure 5: Global proteins sequences alignment of a bi-connexe component of *Decrypthon (2)*

5.3 Textual data

We did another experiment by visualizing co-occurrence relationships between several biological terms used in ***Medline*** abstracts corpora [16]. This co-occurrence relationship is a mutual information measurement defined by the following formula:

$$IM(X,Y) = \frac{2 \times |X,Y|}{|X| + |Y|}, \quad \textit{Where X and Y are terms}$$

In this experiment, nodes represent the terms and lines represent the co-occurrence relationships measurements. The distance between nodes is inversely proportional to the co-occurrence score between terms.
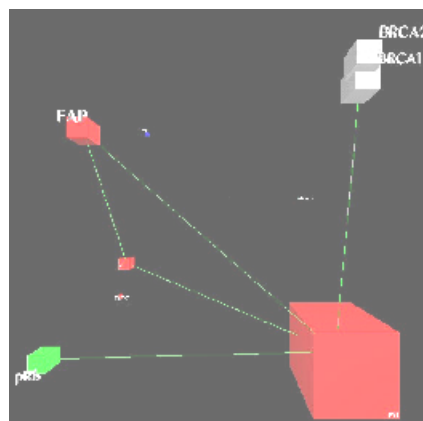
Figure 6: Bibliographic proximity network (from PubMed) of terms relating to cancer (pRb, FAP, P53, BCRA1...)

6   Conclusion and future work

In this paper, our objective was to elaborate new solutions in order to virtually explore both textual and factual genomic data. This approach is mainly based on the definition of a genomic data federator language, answering the requirements and specificities of genomic databases. This language takes into account the heterogeneity resulting from textual and factual data. The representation methods to view these data within an immersive environment were presented and successively tested on various sets of biological data.

Some aspects of this work must be discussed. On the one hand, the federator genomic data description format can describe textual data (extracted from abstracts) as much as factual data (from genomic databases), separately or simultaneously. On the other hand, the independence between data description format and visualization methods offers biologists a total freedom in their representation choices. Moreover, this independence allows users to experiment different visual representations for the same genomic data. This allows users to extract interesting characteristics, visible by comparing several representations, thanks to human cognitive skills, but invisible in a single representation.

Furthermore, both the immersive aspect and the possibility of exploring huge data in a synthetic way constitute the strong points of our system, because it offers a global point of view of the data subjacent structure. These characteristics are particularly interesting when biologists wish to explore a mass of data without precisely knowing what they seek. For example, the partial analysis of **Decrypthon** data shows directly several clusters within the representation.

The co-occurrence measurement of terms is a rather good starting point for a textual abstract corpus exploration. On the one hand, it is a very synthetic representation of the lexical relations existing between the terms, and on the other hand, it reduces considerably the search space for the following phase of semantic relations extraction from genomic texts. It is necessary in the future to look further into the textual linguistic analysis.

Finally, this study was concretized by a software development, named **GenomeExplorer**, which was used to generate the results presented in this paper.

References

[1] I. Rojdestvenski, D. Modjeska, and F. Pettersson. Sequence World: A Genetics Database in Virtual Reality. *International Conference on Information Visualisation (IV 2000),* London, England, 2000.

[2] B.J. Staplet and G. Benoit. Biobibliometrics: Information Retrieval and Visualization from Co-occurrences of Genes Names in Medline Abstracts. *International Pacific Symposium on Biocomputing*, 5:526-537, 2000.

[3]   J. Castano, J. Zhang, and J. Pustejovsky. Anaphora resolution in biomedical literature. *International Symposium on Reference Resolution.* Alicante, Spain, 2002.

[4] J. Pustejovsky, J. Castano, B. Cochran, M. Kotecki, M. Morrell, and A. Rumshisky. Linguistic Knowledge Extraction from MedLine: Automatic Construction of an Acronym Database. An updated version of the paper presented at Medinfo, 2001.

[5] J. Pustejovsky, J. Castano, and J. Zhang. Robust Relational Parsing over Biomedical Litteratures: Extracting Inhibit Relations. *Pacific Symposium on Biocomputing*, 2002.

[6] J.T. Chang and S. Raychaudhuri. Natural Language Processing for Bioinformatics: The time is ripe. International Conference on Intelligence System for Molecular Biology. Copenhagen, Denmark, 2001.

[7] J. Thomas, D. MildWard, C. Ouzounis, S. Pulman and M. Caroll. Automatic Extraction of Protein Interaction from Scientific Abstract. *Proceedings of the Pacific Symposium on Biocomputing (PSB'2000),* pp 538–549, 2000.

[8] R. Gherbi and J. Hérisson. ADN_Viewer, A Software Framework for 3D Modeling and Stereoscopic Visualization of The Genome. *International Conference on Computer Graphics and Computer Vision.* Moscow, Russia, 2000.

[9] R. Gherbi and J. Hérisson. Representation and Processing of Complex DNA Spatial Architecture and its Annotated Content. *Proceedings of the International Pacific Symposium on Biocomputing* (PSB'2002), Hawaii, USA, 2002.

[10] P.E. Gros, J. Hérisson, R. Gherbi. A Human-Machine Platform for 3D Visualization and Management of Genomic Databases. *Sequence Databases and Ontologies, ECCB 2003 satellite Workshop*, Paris, France, 2003.

[11] S.F. Altschul, W. Gish, W. Miller, E.W. Myers and D.J. Lipman. Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215:403–410, 1990.

[12] P. Eades. A heuristic for graph drawing. *Congressus Nutnerantiunt,* 42:149–160, 1984.

[13] Decrypthon. http://www.infobiogen.fr/services/decrypthon/index.html

[14] GenBank. http://www.ncbi.nlm.nih.gov/Genbank/index.html

[15] SwissProt. http://us.expasy.org/sprot

[16] Medline. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi

[17] MeSH. http://www.nlm.nih.gov/mesh/meshhome.html

[18] RasMOL. http://www.genostar.org/

[19] VMD. http://www.ks.uiuc.edu/Research/vmd/

The Dynamic Geometry of Developing Organisms

Brian C. Goodwin

Schumacher College, Dartington, Devon, UK.

As an organism develops from an egg or a bud into the adult form there is an orderly and progressive change of shape, revealing a dynamic that is organised in the four dimensions of space-time. The generic scientific term for such dynamic order in space is a field. Familiar examples are electromagnetic, hydrodynamic, and gravitational fields. These are described by equations that define the state of a field at any point in terms of a specific mathematical relationship to other spatial points in the field, specifying its intrinsic relational order.

In biology the space-time order of a developing organism is described as a morphogenetic field, in analogy with the fields of physics, but there is as yet no consistent theory that defines the specific type of order that underlies morphogenesis. However, there are mathematical models that provide significant insight into the type of dynamic process that is at work, suggesting what type of ordering relationships may be involved in bringing about the cellular and tissue-level changes of form that transform the morphology of the organism during its development. In this article I shall describe briefly work on morphogenetic fields in plants and in animals that give insights into the ways in which molecular processes may be controlled to produce macroscopic changes. This can guide research that allows us to connect large-scale to small-scale dynamics in developmental processes. The basic ideas about morphogenetic fields have been presented [8,18], but the present article looks for a more direct connection with the genetic and molecular level in developing cells and tissues.

1   Plant Morphogenesis

The phenomenon that I shall concentrate on in plants is the process known as phyllotaxis: the generation of spatial patterns of leaf order by the vegetative meristem of the growing plant. There are essentially three different leaf arrangements in flowering plants, of which there are about 250,000 different species. Each species follows one of these patterns, known as distichous or opposite ; whorled or decussate, tricussate, etc depending on how many leaves there are in a whorl ; and spiral. Spiral phyllotaxis is the most frequently observed pattern, found in about 80% of higher plant species. It has been known for many years that these patterns, although stably inherited, are not determined by genes: the vegetative meristem of a species with decussate phyllotaxis can, for instance, be disturbed by an oblique surgical incision that results in the production of two meristems each of which generates spiral phyllotaxis [14]. Also, plants with spiral phyllotaxis can undergo transition to a whorled pattern of flower organs in the floral meristem, and vice-versa [10]. The general question that this leads us to is : what kind of field dynamic in meristems is capable of generating the variety of patterns observed as alternative states of the field ? Included in these alternatives are infrequent but inherited patterns that do not belong to the three major patterns described for the majority of plants. The problem posed by phyllotaxis at the macroscopic level is, then, to describe a morphogenetic field that, with appropriate initial and boundary conditions, can generate the observed variety of phyllotactic patterns and explain their different frequencies across the diversity of higher plant species.

There have been many contributions to this problem by different scientists over the course of the past 100 years, but a particularly comprehensive treatment has been provided by Douady and Couder in a series of articles published in 1996 (*J. theoret. Biol.*, 178, pp 255, 275, and 295). They show how a morphogenetic field in the meristem, governed by a particular type of spatial

relationship that describes how existing leaf primordia inhibit new leaf initiation as a function of distance, together with the conical geometry of the meristem, can account for the various patterns of leaf formation observed. They also show that the differential abundance of the different patterns among the different species, with spiral the most common, can be understood in terms of the intrinsic stability of the different phyllotactic modes. This model shows us, in principle, how leaf phyllotaxis can be generated by a single type of morphogenetic field, the various patterns arising as a result of different initial and boundary conditions governed by parameters of the model. The parameter values that stabilise different phyllotactic patterns are assumed to be specified by genes whose action can select the particular trajectory followed by the morphogenetic field so that one species follows one pattern and another a different one. However, a specific morphogenetic pattern can be transformed into another by a change in the boundary conditions of the field, as in a surgical cut through the meristem, or by change of parameter values during development, governed by a change in gene activity, as in a switch from spiral leaf pattern to a whorled pattern of organs in the flower.

Despite its capacity to provide a dynamic explanation of phyllotaxis in principle, the limitation of a theoretical model of morphogenesis of the type studied by Douady and Couder is that it does not tell us precisely what type of field we are dealing with, nor does it inform us about the molecular nature of the processes involved. It operates at a level of abstraction that gives us important insights into the type of explanation required in terms of spatial ordering principles, but it necessarily fails to tell us the details of how the process is realised in terms of molecules and forces involved. In the same way, the electromagnetic field fails to explain the nature of the entities that express it, such as photons, or the origin of the forces whose effects are observed. Nevertheless, it gives us very significant insights into the process we seek to understand, defining many of its constraints. Maxwell's equations for electromagnetic fields capture essential relationships underlying the phenomena, although they need to be significantly modified to describe quantum electrodynamics.

Recently there have been studies by Reinhardt et al. [13] indicating the nature of the molecules involved in generating phyllotactic patterns that reveal a field dynamic different in detail to that assumed by Douady and Couder (1996). They assumed an inhibitory influence, possibly carried by a morphogen from a leaf primordium, that decreases with distance, allowing a new leaf to form at a critical distance from a developing leaf. However, the evidence presented by Reinhardt et al. is that the influence is an induction of primordium formation by auxin. A new leaf acts as a sink for auxin that gets transported away from the site, thus preventing new leaf formation in its neighbourhood. The inhibition assumed by Douady and Couder is then replaced by removal of an inducer from the new leaf rather than a direct inhibitory action by a morphogen. The field rule is therefore effectively inverted so that instead of permissive regions of low inhibition allowing new leaf formation in a scalar field, we have domains of elevated auxin, a morphogen that induces leaf initiation, and transport of auxin away from the new leaf and its boundaries in the meristem as the leaf develops, preventing new leaf initiation in its neighbourhood. The dynamics is more complex than a scalar field of inhibition because there is now a vector component that governs where the peaks of auxin occur in the field, arising from polarised transport of the morphogen. The field is therefore closer to a hydrodynamic flow field, with sources and sinks and a vector field of flux direction and rate throughout the meristem.

The elegant studies of Reinhardt et al. take the question of morphogenetic fields in plants in some very interesting and challenging directions. Clearly they do not explain phyllotaxis, which requires both a coherent model of the whole and an experimentally valid molecular dynamic, as in the construction of the Navier-Stokes equations for hydrodynamic fields in liquids. The vector field component underlying auxin flow direction is governed by the spatial location of auxin pumps in the cell membrane, such as PIN1, a protein that is localised to the apical side of cells in the outer

layers of the meristem, transporting auxin apically, but orientated for basal transport in developing vascular cells of the developing leaf. Thus there is a field of regulatory influence that underlies the production and localisation of PIN1 in cells that needs to be explained in terms of molecular organising forces in order to model this morphogenetic field. Modelling the close coupling between oriented pumping across cell walls, cell differentiation, and auxin flow patterns throughout the meristem to produce the coherent macroscopic patterns of leaf arrangements observed in leaf phyllotaxis presents an interesting challenge for those seeking to provide an integrated understanding of this robust morphogenetic process. Clearly any model that seeks to explain the phenomena must achieve the same range of integrated understanding of phyllotactic patterns and their transformations as the model of Douady and Couder does.

2   Morphogenesis in Animals

Animal morphogenesis has deep similarities with the process in plants, but also has significant differences. There is a coherent dynamic throughout the embryonic domain that gives rise to a particular structure such as a limb or an eye, a heart or a sex organ, similar to the coherence that gives to the leaf primordium its spatial order. There is also a higher-order spatial pattern that generates the spatial relationships between the various organs within the normal animal body as a whole, as there is in the plant meristem, producing the pattern of relationships that results in phyllotaxis. There is thus a hierarchy of ordering relations in the developing organism that can be partially uncoupled, revealing what are called developmental modules. The partial autonomy of these modules results in the possibility of inducing ectopic eyes or limbs or flower organs that are recognisable units of form though they are out of place in relation to the normal pattern. My concern here is to examine possible connections between intracellular distributions of particular classes of molecule that can be regarded as morphogens in animal embryos and the macroscopic patterns they may produce at the tissue level. In this sense I am looking at relationships similar to those between polar transport proteins like PIN1 and a morphogen such as auxin, which affect growth and form in the meristem, but now asking what type of molecular action may be involved in generating the patterns of geometrical change that are observed in animal embryos.

I shall base the model that follows on a ideas presented by Fred Cummings, a research colleague, in a series of papers connecting geometrical transformations in animal embryos to the action of morphogens that affect cell-cell adhesion and tissue deformation [2-6]. The core idea underlying Cummings' work is to connect the spatial patterns of morphogens localised within cell membranes to the systematic changes in the curvature of cell sheets that underlies morphogenetic processes such as budding and tentacle formation in hydroids, bryozoa, and ascidians ; gastrulation, neurulation, and segmentation in chordates ; and formation of the heart, lungs and gastrointestinal tract in vertebrates. This approach was described briefly in [9]. Cummings has also applied the general ideas of his model to plant phyllotaxis [7]. This model works at a high level of abstraction but seeks to link changing patterns of gene activity and molecular distibutions to morphogenetic movements. The core idea is a closed dynamic cycle in which cell adhesion molecules localised in membranes of polarised cells comprising sheets of embryonic tissue induce shape changes in the tissue, and these geometrical changes cause differential change in the concentration and distribution of cell adhesion molecules.

Cummings assumes that there are two different types of cell adhesion molecule, present in concentration N_A and N_B , activated by diffusible ligands A and B. The rate of production of ligand A is proportional to the concentration of activated adhesion molecule N_A in the cell, while its production is inhibited by N_B, and similarly for ligand B, so that the two together act as a kind of toggle switch : where N_A is high, N_B is low, and vice versa. Since ligands can diffuse between cells, activated adhesion molecules are governed by partial differential equations of the type familiar to

biologists through the work of Turing [17] and Meinhardt [11] on morphogen patterns. However, Cummings' model generalises this so that there is no need for a pair of morphogens with different diffusion constants to produce spatial patterns, as in reaction-diffusion models. Spatial patterns can form simply as a result of the kinetics of production of the cell adhesion molecules in the partial differential equations describing their production and spatial patterning.

While the kinetics of these cell adhesion molecules is hypothetical, there are categories of signalling molecules in cells that have a similar type of behaviour to those described above as a switch, in particular members of the Wnt family (Wg or wingless in Drosophila), which are accompanied by modulating G proteins. Wnt/Wnt interaction may be the primary example of the switch described above [16,12,15,19], while Delta and Notch constitute another example [1]. Since Wnt proteins control many developmental pathways, including cell adhesion molecules [19], these identified regulatory molecules may underlie the kinetics assumed for the reciprocally coupled pair of cell adhesion molecules N_A and N_B, the morphogens of the model.

In general, as the geometry of the developing tissue changes under the action of the cell adhesion molecules, the pattern of their distribution in space will change in accordance with the property that for any particular geometry there will be a particular spatial solution of the partial differential equations, starting with defined initial and boundary conditions. There is thus a closed causal loop from morphogen distribution to geometry and back to morphogen pattern, which can change as the cycle proceeds. The system has dynamic autonomy. Cummings [6] has shown that various morphogenetic patterns, such as gastrulation and segmentation, can be generated by this type of model. The computations require the use of conformal coordinates for position in the tissue, and coupling between morphogen concentrations and functions defining local curvature of cell sheets. However, very few parameters are used and solutions appear to be robust. There are also ways in which one can include switching between gene regulators as size and geometry of the developing embryo change.

The value of such a model is the realisation of a basic pattern generator for animal embryos that produces three dimensional morphologies, doing so by a continuous unfolding for shape from simple initial conditions such as a symmetric ball of cells (e.g., an asexual bud in hydroids or ascidians), or as a blastula in higher organisms undergoing sexual reproduction. There are many molecular candidates for molecules that can influence cell shape such as those suggested involving Wnt, and including interactions between the Wnt/β-catenin branch and a Wnt/Ca^{2+} branch that extends the influence to the cytoskeleton and modulating proteins. Cummings' model is intended to do no more than provide a candidate for a new type of morphogenetic dynamic that is autonomous and includes specification of geometry in the developing organism. As such, it is an original contribution to thinking that links the different levels of morphogenesis from genes and molecules to the three-dimensional form of the whole developing organism, using plausible mechanisms.

References

[1]   J. Collier, N. Monk, P. Maini and J. Lewis. Pattern formation by lateral inhibition with feedback: Mathematical model of Delta-Notch intercellular signalling. *J. Theor. Biol.,* 183:429-446, 1996.

[2]   F.W. Cummings. Aspects of growth and form. *Physica D,* 79:146-163, 1994.

[3]   F.W. Cummings. Geometrical concepts in epithelial sheets. *J. Theor. Biol.,* 179:41-49, 1996.

[4]   F.W. Cummings. A model of pattern formation based on signaling pathways. *J. Theor. Biol.,* 207:107-116, 2000.

[5] F.W. Cummings. The interaction of surface geometry with morphogens, *J. Theor. Biol.*, 212:303-313, 2001.

[6] F.W. Cummings. Coupled morphogens and geometry. *J. Theor. Biol.,* 2004 (in press).

[7] F.W. Cummings and J.C. Strickland. A model of phyllotaxis, *J. Theor. Biol.*, 192:531-544, 1998.

[8] B.C. Goodwin. How The Leopard Changed Its Spots. Weidenfeld and Nicolson, London, UK, Princeton University Press, Princeton (NJ), USA, 1994.

[9] B.C. Goodwin, The life of form. Emergent patterns of morphological transformation. *C.R. Acad. Sci. Sciences de la Vie*, 323:15-21, 2000.

[10] P.B.Green, A. Havelange and G. Bernier. Floral morphogenesis in *Anagallis* : scanning electron micrograph sequences from individual growing meristems before, during and after the transition to flowering. *Planta,* 185:502-512, 1991.

[11] H. Meinhardt. *Models of Biological Pattern Formation.* Academic Press, 1982.

[12] C. Niehus. Solving a sticky problem, *Nature,* 413:787-788, 2001.

[13] D. Reinhardt, , E.-R. Pesce, P. Stieger, T. Mandel, K. Baltensperger, M. Bennett, J. Traas, J. Friml, and C. Kuhlemeier. Regulation of phyllotaxis by polar auxin transport. *Nature,* 426:255-260, 2003.

[14] M. Snow and R. Snow. Experiments on phyllotaxis. III. Diagonal splits through decussate species. *Phil. trans. Roy. Soc. Lon. B.* 139:545-566, 1935.

[15] J. Taipale and P.A. Beachy. The Hedgehog and Wnt signalling pathways in cancer. *Nature,* 411:349-354, 2001.

[16] O. Tetsu and F. McCormick. β-Catenin regulates expression of cyclin D1 in colon carcinoma cells, Nature **398**, 422-423, 1999

[17] A.M. Turing. The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. B,* 237:37-72, 1952.

[18] G. Webster and B. Goodwin. Form and Transformation ; *Generative and Relational Principles in Biology.* Cambridge University Press, Cambridge, UK, 1996.

[19] A. Wodarz and R. Nusse. Mechanisms of Wnt signalling in development, *Ann. Revs. Cell Biol.,* 14:59-88, 1998.

Modelling the circadian clock in *Drosophila* and mammals

Jean-Christophe Leloup & Albert Goldbeter

Unité de Chronobiologie théorique, Faculté des Sciences, Université Libre de Bruxelles, Campus Plaine, C.P. 231, B-1050 Brussels, Belgium.

1 Introduction

Rhythmic phenomena occur at all levels of biological organization, with periods ranging from less than a second to years [1]. Among these rhythms, circadian oscillations, which occur with a period of about 24 h, play a key physiological role in the adaptation of living organisms to their periodically varying environment. These circadian rhythms can occur in constant environmental conditions, e.g. constant darkness, and are therefore endogenous. Experimental advances during the last decade have permitted to largely unravel the molecular bases of circadian rhythms in a number of organisms, the most studied so far being *Drosophila* [2]. Significant progress has also been made on the mechanism of circadian rhythms in *Neurospora* [3], cyanobacteria [4], plants [5] and mammals [2].

Oscillatory behaviour often originates at the cellular level from regulatory feedback loops which involve many parameters and interacting variables. Relying only on sheer intuition to predict the dynamics of such complex regulatory systems rapidly meets with limitations. Analyzing the origin of oscillations has therefore much to gain from theoretical models closely related to experimental observations. Some of the roles and advantages of theoretical models in biology are listed in Table 1 [6]. These considerations on the use of theoretical models apply to the study of biological processes in general, but pertain with particular weight to biological rhythms that only occur in precise conditions. Determining these conditions is a primary goal of a modelling approach.

- Provide a unified theoretical framework accounting for available experimental observations, that corroborates or not experimental conclusions.
- Conceptualization leads to clarification of hypotheses.
- Analyze complex situations involving multiple, coupled variables, for which it becomes impossible to rely only on sheer intuition.
- Models show that certain types of behavior only occur in precise conditions, in a domain bounded by critical parameter values, in contrast to what may be predicted by merely verbal descriptions.
- Determine the qualitative and quantitative effects of each parameter and identify key parameters.
- Rapid exploration of different mechanisms and of large ranges of conditions.
- Possibility to ask questions which may be inaccessible to experiments or hard to address experimentally.
- Testable predictions: suggestion of experiments, which will either validate the model or call for its modification.
- Optimal situation: provide counterintuitive explanations or surprising predictions.
- Mathematical structure underlines link with similar phenomena in other contexts.

Table 1: Some roles and advantages of theoretical models in biology [6].

Models have so far been applied primarily to ultradian biochemical oscillations, characterized by periods ranging from seconds to minutes [1]. Theoretical models for circadian rhythms were at first borrowed from the physical literature, as exemplified by the use of the van der Pol oscillator for modelling properties of circadian oscillations. This line of research is still pursued to study, for example, the effect of light on the human circadian system [7].

Besides these abstract models, a complementary approach rests on the study of models more directly related to the biochemical regulatory processes that underlie circadian rhythms. These models can be viewed as extensions of a general three-variable model proposed by Goodwin [8] soon after the principles of genetic regulation were established. Anticipating many of the subsequent experimental findings on circadian clock mechanisms, this author suggested that negative feedback on gene expression can lead to oscillations in protein and mRNA levels. Goodwin's model is still used, e.g. to study the phase shifting of circadian rhythms by inhibitors of protein synthesis in *Neurospora* [9]. Given the increasing availability of experimental data, more detailed theoretical models can now be considered for circadian rhythms. Such models based on transcriptional regulation have so far been proposed for the circadian clock of *Drosophila* [10-14], *Neurospora* [9,14] and mammals [15,16].

2   *Model for circadian rhythms in* Drosophila

The model for circadian oscillations in *Drosophila* is represented schematically in Figure 1. The *per* and *tim* genes are transcribed in the nucleus before the mRNAs are transported into the cytosol where they are translated. The PER and TIM proteins are multiply phosphorylated and form a complex that enters the nucleus [2]. Through interaction with the products of the activator genes *clock* and *cyc* (not shown in the Figure) the PER-TIM complex inhibits the expression of the *per* and *tim* genes. The model for circadian rhythms in *Drosophila* incorporating the formation of the PER–TIM complex is described by a set of ten kinetic equations [11]. Light controls the rhythm by triggering the destruction of the TIM protein through an ubiquitin-proteasome mechanism that requires TIM phosphorylation [2].

The model schematized in Figure 1 predicts the occurrence of sustained oscillations in constant darkness (DD), as observed experimentally. Shown in Panel A of Figure 2 are the oscillations in total PER protein (P_t), *per* mRNA (M_P), and nuclear PER–TIM complex (C_N) obtained in conditions corresponding to DD; such conditions are achieved in the *Drosophila* model by holding at a constant low value parameter v_{dT} which measures the maximum rate of TIM degradation (see Figure 1). Although the environmental conditions remain constant, the PER–TIM control system generates autonomous oscillations with a period close to 24 h for the set of parameter values considered.

When plotting the time evolution of the system as a function of the concentrations of two of the biochemical variables, sustained oscillations obtained in conditions of constant darkness (Figure 2, Panel A) correspond to a trajectory that takes the form of a closed curve (Panel B). For a given set of conditions (i.e. of parameter values), this closed curve can be reached regardless of initial conditions and is generally unique; hence the name of *limit cycle* given to this type of trajectory associated with periodic behaviour. Because perturbations do not change in the long run their period or amplitude, limit cycle oscillations represent a particularly stable mode of periodic behaviour. Such stability holds with the robust nature of circadian clocks that have to maintain their amplitude and period in a changing environment while retaining the capability of being phase shifted by light or temperature.
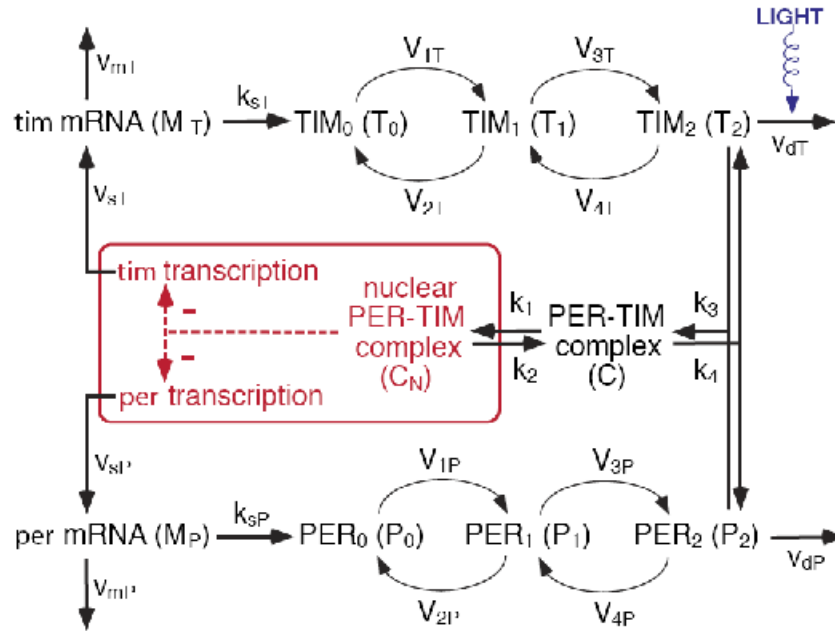
Figure 1: Scheme of the model for circadian oscillations in *Drosophila* [6,11]. The model is based on the negative regulation exerted by the PER–TIM protein complex on the expression of the *per* and *tim* genes; light controls the rhythm by enhancing the rate of TIM degradation, v_{dT}. The model incorporates gene transcription in the nucleus, accumulation of the corresponding mRNAs in the cytosol where the associated protein synthesis takes place, protein and mRNA degradation, phosphorylation-dephosphorylation reactions involving PER and TIM, protein transport into and out of the nucleus, formation of a PER-TIM complex and regulation of gene expression by the nuclear form of this complex. Nuclear processes are shown in red and the effect of light is indicated by the blue arrows.
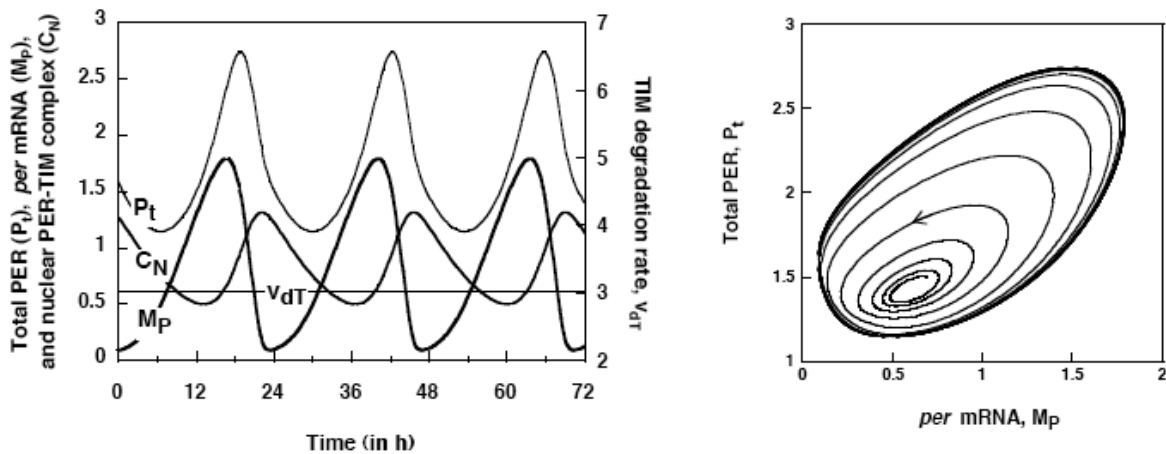
Figure 2: Oscillations in continuous darkness and evolution toward the limit cycle [6]. (A) Shown is the temporal variation in *per* mRNA (M_P) and in the total amount of PER protein (P_t), together with the variation in nuclear PER-TIM complex (C_N). (B) Evolution toward a limit cycle corresponding to sustained oscillations in Panel A. The limit cycle is reached here from initial conditions located near the unstable steady state. The arrow indicates the direction of movement toward and along the limit cycle.

Light triggers degradation of the TIM protein in *Drosophila*. Incorporating a periodic variation of the light-controlled parameter into the model for the *Drosophila* circadian clock allows us to simulate the entrainment of circadian oscillations by light-dark (LD) cycles. In such conditions, the maximum TIM degradation rate v_{dT} varies in a square-wave manner as it increases up to a higher value during each light phase. As the duration of both the light and dark phases is equal to 12 h in the case considered, the system is entrained precisely to the 24 h external periodicity. The effect of continuous light (LL) is simulated by holding parameters v_{dT} at a constant high value. As observed in the experiments, the oscillations in the model are readily damped in LL when increasing v_{dT} up to a higher, constant value.

The induction of phase shifts by light pulses represents one of the most conspicuous properties of circadian rhythms. Since we have incorporated the effect of light into the model for *Drosophila* circadian rhythms [11], we can determine in this model the response to light pulses as a function of the phase at which the system is perturbed, and we can use the results to construct phase response curves (PRCs) that can be compared with those determined experimentally, as for the case of the wild type and the short period *per*^S mutant [11].

Another interest of theoretical models is to shed light on the conditions in which light pulses can suppress circadian rhythmic behavior. The possibility that critical pulses of light may achieve such an effect by bringing the oscillatory system back to the singularity has long been explored both experimentally and theoretically by Winfree [17]. The models considered here allow us to address this issue explicitly [18] and to show that there exist a domain of duration and amplitude for the light pulse that can suppress the oscillations and bring the system to a stable steady state.

3   Model for circadian rhythms in mammals

The model for circadian oscillations in mammals is represented schematically in Figure 3. As the situation in mammals resembles that observed in *Drosophila* [2], this model is closely related to the model for *Drosophila* (Figure 1). Nevertheless, there are several differences between these two models. Instead of TIM, it is the CRY protein that forms a regulatory complex with a PER protein. Several forms of these proteins exist (PER1, PER2, PER3, CRY1, CRY2). The PER-CRY complex inhibits the expression of the *Per* and *Cry* genes in an indirect manner, by binding to the complex CLOCK–BMAL1; the latter, formed by the products of the *Clock* and *Bmal1* genes, activates *Per* and *Cry* transcription. Besides this negative regulation of gene expression, indirect positive regulation is also involved. *Bmal1* expression is subjected to negative autoregulation by BMAL1, via the product of the *Rev-Erbα* gene. The complex between PER2 and CRY1 or CRY2 enhances *Bmal1* expression in an indirect manner, by binding to CLOCK–BMAL1 and thereby reducing the transcription of the *Rev-Erbα* gene. Light can entrain circadian rhythms in mammals by inducing the expression of the *Per* genes.

The model can account for autonomous, sustained circadian oscillations in conditions corresponding to continuous darkness. In agreement with experimental observations, *Bmal1* mRNA oscillates in antiphase with *Per* and *Cry* mRNAs, and the proteins undergo similar oscillations and follow their mRNA by a few hours. Sustained oscillations only occur in an appropriate range of parameter values. Outside this range, rhythmic behavior disappears and the system evolves toward a stable steady state; such an evolution is often accompanied by damped oscillations. Given the large number of parameters considered in the model it is difficult to thoroughly assess its sensitivity to changes in parameter values. Useful insights can nevertheless be obtained by determining, for each parameter, one at a time, the range of values producing sustained oscillations as well as the

variation of the period over this range, while keeping for the other parameters the basal values. This sensitivity analysis allows us to identify the key parameters of the system. The model also accounts for the entrainment by light-dark cycles. However, the phase of the oscillations after entrainment can be very sensitive to the choice of other parameter values. Thus, different phases can be obtained in LD for parameter values yielding comparable periods of circadian oscillations in DD.

The existence of intertwined positive and negative regulations in the scheme of Figure 3 raises the possibility that the mechanism producing sustained oscillations may not be unique. To test whether the oscillations rely primarily, as expected, on the indirect negative feedback loop involving the inactivation of the CLOCK–BMAL1 complex via its binding to PER–CRY, we tried to determine whether oscillations still occur when preventing this inactivation. When silencing the negative feedback loop involving the PER–CRY complex, e.g. by setting to zero the rate of synthesis of the PER protein, oscillations disappear and the system evolves toward a stable steady state, as observed in mice for the double mutants *mPer1/mPer2* [2].
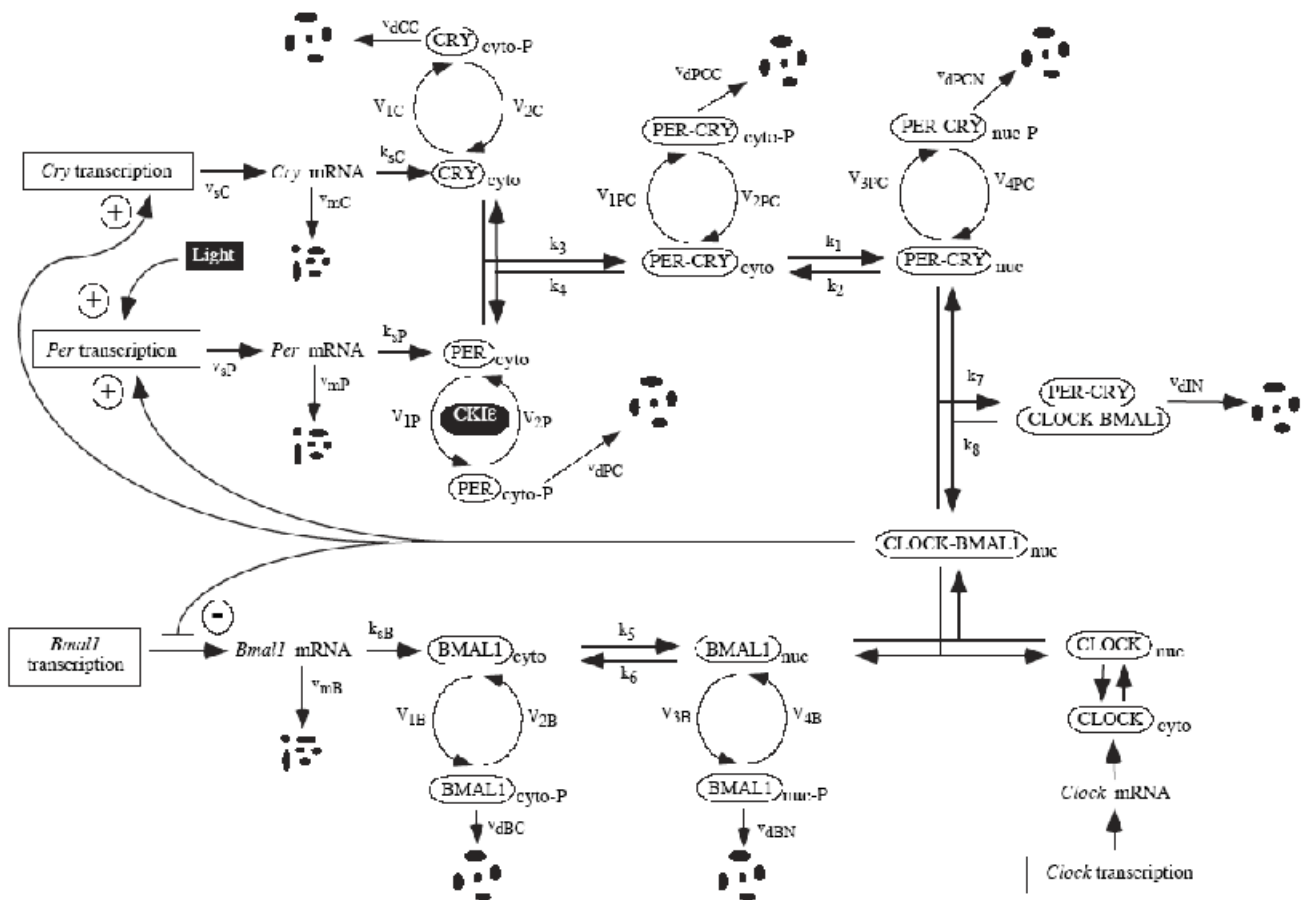
Figure 3: Model for circadian oscillations in mammals involving interlocked negative and positive regulations of *Per*, *Cry*, *Bmal1* and *Rev-Erbα* genes by their protein products [14]. We focus on the case where BMAL1 exerts a direct negative feedback on the expression of its gene. The role of the *Rev-Erbα* gene product in the indirect regulation of *Bmal1* expression by BMAL1 (indicated in grey) is considered in a second stage (the grey loop then replaces the direct negative feedback exerted by BMAL1). The kinetic equations governing the time evolution of the model are given in [15], together with the definition and values of the parameters.

It is noteworthy that the model also predicts the possibility of sustained oscillations in the absence of *Per* mRNA or PER protein. Indeed, the model indicates that the negative autoregulatory feedback exerted by CLOCK–BMAL1 on the expression of the *Bmal1* gene suffices to produce sustained oscillatory behaviour in all the other variables. At least two oscillators are thus coupled within the circadian control system. The first oscillator relies primarily on the indirect negative feedback exerted on the expression of *Per* and *Cry* through the binding of PER–CRY to the CLOCK-BMAL1 activating complex. The second mechanism capable of generating sustained oscillations is based on the negative feedback exerted by CLOCK-BMAL1, via REV-ERBα, on the expression of the *Bmal1* gene. The latter mechanism should become unmasked only when the other feedback becomes inoperative, provided that parameter values are such that they allow for sustained oscillations. Parameter values may indeed be such that in the absence of the PER–CRY feedback loop the system evolves, with or without damped oscillations, to a nonoscillatory state. It is also possible that the second oscillatory mechanism is only capable of producing damped oscillations, which could become sustained when entrained by a periodic signal such as temperature cycles. The possibility of a second oscillatory mechanism in mammals holds with the observation that in *mPer1/mPer2* deficient mice, rhythmicity can be restored for several days by an extended light pulse (K. Bae and D. Weaver, personal communication).

Of particular interest is the effect of the maximum rate of PER phosphorylation, V_{phos} because the phosphorylation status of PER has been related to disorders of the sleep-wake cycle in humans. Thus, a mutation of hPER2 that reduces its ability to be phosphorylated by casein kinase Iε has been linked with the familial advance sleep phase syndrome, FASPS [19]. In this syndrome, sleep onset and offset occur very early in a 24h LD cycle, and the phase of sleep is advanced by 3-4 h. This phenomenon can be accounted for by the model upon decreasing the value of the PER phosphorylation rate [15].

The behavior outside the range of entrainment also bears on physiological disorders of the sleep-wake cycle. Lack of entrainment corresponds to the non-24h sleep-wake syndrome in which the phase of the sleep-wake pattern constantly changes with respect to the LD cycle [20]. Such free-running circadian oscillations have been observed both in blind [21] and sighted subjects [22].

The theoretical model for the mammalian clock thus allows us to address not only the molecular mechanism of circadian rhythms but also the dynamical bases of physiological disorders related to perturbations of the circadian clock.

References

[1] A. Goldbeter. *Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour*. Cambridge University Press, Cambridge, UK, 1996.

[2] R. Stanewsky. Genetic analysis of the circadian system in *Drosophila melanogaster* and mammals. *J. Neurobiol.,* 54:111-47, 2003.

[3] J.J. Loros and J.C. Dunlap. Genetic and molecular analysis of circadian rhythms in *Neurospora. Annu. Rev. Physiol.,* 63:757-94, 2001.

[4] S.S. Golden. Timekeeping in bacteria: The cyanobacterial circadian clock. *Curr. Opin. Microbiol,* 6:535-40, 2003.

[5]   M.E. Eriksson and A.J. Millar, The circadian clock. A plant's best friend in a spinning world. *Plant. Physiol.,* 132:732-8, 2003.

[6]   J.-C. Leloup and A. Goldbeter. Modeling the molecular regulatory mechanism of circadian rhythms in *Drosophila. BioEssays,* 22:83-92, 2000.

[7]   M.E. Jewett and R.E. Kronauer. Refinement of a limit cycle oscillator model of the effects of light on the human circadian pacemaker. *J. Theor. Biol.,* 192:455-65, 1998.

[8]   B.C. Goodwin. Oscillatory behavior in enzymatic control processes. *Adv. Enzyme Regul.,* 3:425-438, 1965.

[9]   P. Ruoff, M. Vinsjevik, S. Mohsenzadeh and L. Rensing. The Goodwin model: simulating the effect of cycloheximide and heat shock on the sporulation rhythm of *Neurospora crassa. J. Theor. Biol.,* 196:483-94,1999.

[10] A. Goldbeter. A model for circadian oscillations in the *Drosophila period* (PER) protein. *Proc. R. Soc. Lond. B.,* 261:319-24, 1995.

[11] J.-C. Leloup and A. Goldbeter. A model for circadian rhythms in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins. *J. Biol. Rhythms,* 13:70-87, 1998.

[12] H.R. Ueda, M. Hagiwara and H. Kitano. Robust oscillations within the interlocked feedback model of *Drosophila* circadian rhythm. *J. Theor. Biol.,* 210:401-6, 2001.

[13] P. Smolen, D.A. Baxter and J.H. Byrne. Modeling circadian oscillations with interlocking positive and negative feedback loops. *J. Neurosci.,* 21:6644-56, 2001.

[14] J.-C. Leloup, D. Gonze and A. Goldbeter. Limit cycle models for circadian rhythms based on transcriptional regulation in *Drosophila* and *Neurospora*. *J. Biol. Rhythms,* 14:433-48, 1999.

[15] J.-C. Leloup and A. Goldbeter. Toward a detailed computational model for the mammalian circadian clock. *Proc. Natl Acad. Sci. USA,* 100:7051-6, 2003.

[16] D.B. Forger and C.S. Peskin. A detailed predictive model of the mammalian circadian clock. *Proc. Natl Acad. Sci. USA,* 100:14806-11, 2003.

[17] A.T. Winfree. Integrated view of resetting a circadian clock. *J. Theor. Biol.,* 28:327-74, 1970.

[18] J.-C. Leloup and A. Goldbeter. A molecular explanation for the long-term suppression of circadian rhythms by a single light pulse. *Am. J. Physiol. Regulatory Integrative Comp. Physiol.,* 280:R1206-12, 2001.

[19] K.L. Toh, C.R. Jones, Y. He, E.J. Eide, W.A. Hinz, D.M. Virshup, L.J. Ptacek and Y.-H. Fu. An *hPer2* phosphorylation site mutation in familial advanced sleep-phase syndrome. *Science* 291:1040-3,2001.

[20] G.S. Richardson and H.V. Malin. Circadian rhythm sleep disorders: pathophysiology and treatment. *J. Clin. Neurophysiol.,* 13:17-31, 1996.

[21] S.W. Lockley, D.J. Skene, K. James, K. Thapan, J. Wright and J. Arendt. Melatonin administration can entrain the free-running circadian system of blind subjects. *J. Endocrinol.,* 164:R1-6, 2000.

[22] A.J. McArthur, A.J. Lewy and R.L. Sack. Non-24-hour sleep-wake syndrome in a sighted man: circadian rhythm studies and efficacy of melatonin treatment. *Sleep,* 19:544-53, 1996.

Models for axes formation in higher organisms

Hans Meinhardt

Max-Planck-Institut für Entwicklungsbiologie, Spemannstr. 35, D- 76072 Tübingen, Germany.

A crucial step in the development of a higher organism is the generation of the primary body axes. How are these axes generated? How is their orthogonal orientation guarantied? How is the internal patterning along an axis accomplished? How are legs and wings inserted at the correct place, with the correct orientation in relation to the main body axes? How are segment formed? I have proposed molecular-realistic models for the steps mentioned above. These models reveal the minimum molecular requirements that have to be satisfied to accomplish these steps. Most of these models found direct support from more recent molecular-genetic observations.

1   An ancestral axis: the patterning of the freshwater polyp Hydra and of the brain region in higher organisms

Radial-symmetric animals such as Cnidarians are regarded to be close to the common ancestor before bilaterality emerged during evolution. Comparison of the expression patterns of homologous genes in Hydra and in higher organisms suggest the following scenario (Fig. 1):
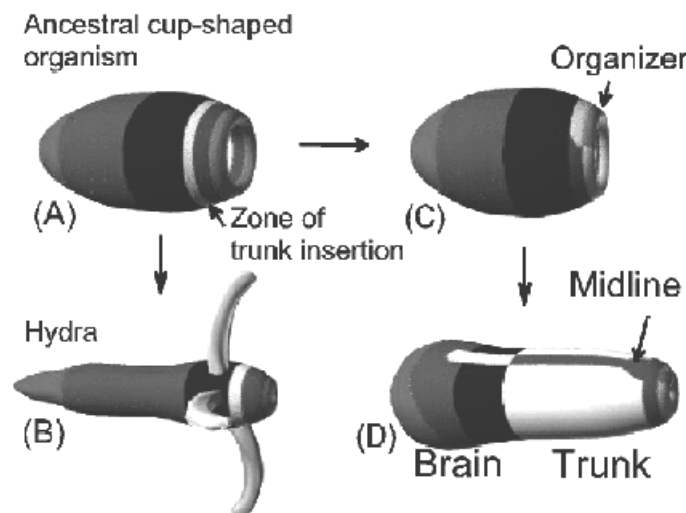
Figure 1: Relation of the axis in the freshwater polyp Hydra to those of higher organisms. A hypothetical common ancestor (A) of Hydra (B) and vertebrates (D) is assumed to have a cup-shaped radial-symmetric geometry. The bottom of the cup-shaped organisms (left) gave rise to the foot in Hydra and to the most anterior part of vertebrates, the forebrain and heart (*Nkx2.5* expression, light grey at left in each figure). In contrast, the opening of the gastric cavity gave rise to the so-called mouth opening in Hydra and to the anus of higher organisms (*Wnt*-expression). An essential intermediate step is assumed to be a widening of the blastoporus. (C) The Spemann-Organizer, initiated on the blastoporus, is a new organizing region that initiates and elongated the midline, which is responsible for the organization of the dorsoventral axis.

1. The hypostome with the gastric opening of the hydra (the so-called "head") corresponds to the most posterior pole, the anus of higher organisms (*Wnt*-expression).
2. The Hydra-foot corresponds to the most anterior part of higher organisms, the forebrain and heart. This is indicated by the *Nkx2.5* expression.
3. The tentacle/hypostome border corresponds to the Midbrain/Hindbrain border (posterior border of *Otx*-expression).
4. A narrow zone next to the gastric opening, the region of *Gsc/Brachyury* expression enfolded to form the trunk in higher organisms.

Thus, hydra can be regarded as a living fossil, telling us about evolutionary early axis formation before bilaterality and trunk formation was invented. In this view it is straightforward that the 3'-5' Hox gene pattern, typical for the trunk of higher organisms, is absent in hydra: in the common ancestor the trunk was not yet there. The system that were patterning once the body of the ancestor was later used to pattern essentially the fore- and midbrain of vertebrates. The trunk, the largest portion of higher animals was added during further evolution by an enfolding of a narrow zone between the tentacle zone and the gastric opening.

2 How to organize a single axis: pattern formation by autocatalysis and lateral inhibition

Some small specialized regions obviously play a decisive role for the overall organization of the developing organism. Such organizing regions direct pattern formation in the surrounding tissue. Pattern formation requires that originally equivalent cells become different from each other. We have proposed that biological pattern formation is based on local self-enhancement and long-range inhibition. It can be shown that the mechanism proposed by Turing (1952) is also based on this principle although Turing has not interpreted his mathematically abstract concept in this way.
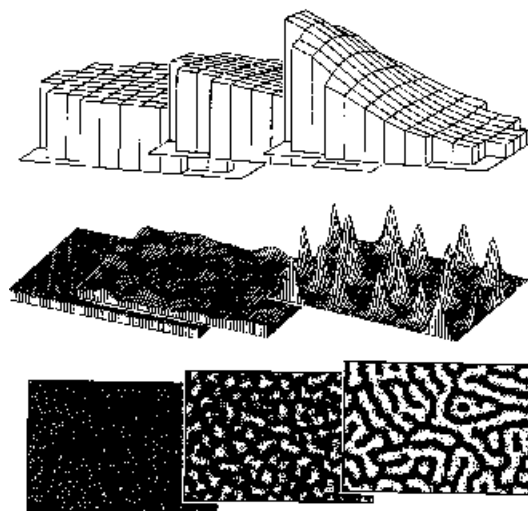
Figure 2: Stages in the generation of elementary patterns by local self-enhancement and long ranging inhibition. Shown are the initial, an intermediate and the final activator distribution. Top: Monotonic gradients are formed if the range of the activator is comparable to the size of the field. The pattern orients itself along the longest extension of the field. This is appropriate to supply the tissue with positional information. Centre: a more or less regular arrangement of peaks results in fields that are large compared to the range of the inhibitor. Bottom: stripe-like distributions result if the autocatalysis saturates (Meinhardt, 1989).

A molecular realization of this concept requires a substance (or a cascade of substances) that has an autocatalytic feedback on its own synthesis. A simple molecular interaction with pattern-forming capability could consist of an 'activator' *a(x)* whose autocatalysis is slowed down by a long ranging 'inhibitor', *h(x)*. The following equation provides an example:

$$\frac{\partial a}{\partial t} \quad = \quad \rho\,\frac{a^2}{h} - \mu_a\,a + D_a\,\frac{\partial^2 a}{\partial x^2} + \rho_0$$

$$\frac{\partial h}{\partial t} \quad = \quad \rho\,a^2 - \mu_h\,h + D_h\,\frac{\partial^2 h}{\partial x^2}$$

These equations describe the change of the activator and inhibitor concentration per time unit. The concentration change of the activator is proportional to an autocatalytic production term (a^2), and that the autocatalysis is slowed down by the action of the inhibitor ($1/h$). As any other biological substance, the activator molecules become degraded. It is natural to assume that the number of disappearing activator molecules is proportional to the number of activator molecules present. The autocatalysis must be non-linear since the production rate must overcome the disappearance by the linear decay. The simulations in Fig. 2 shows that this interaction is able to generate basic pattern observed in development, gradients, periodic patterns and stripes.

3   Gene activation: a pattern formation in the 'gene space'

The generation of signals by the exchange of molecules via diffusion works only in small fields. In larger fields the time required to exchange information by randomly moving molecules would be much too long. Therefore, the signals generated at small scales have to be translated into more permanent cell states that can be maintained upon further growth. The obvious means is a stable concentration- (and thus space-) dependent activation of genes. The choice of a particular pathway under the influence of a morphogenetic signal requires the activation of particular genes and the suppression of alternative genes. This situation has formal similarities with pattern formation. Pattern formation in space requires an activation at a particular position and the inhibition in the remaining part. Analogously, the selection of a particular pathway requires the activation of a particular gene and the suppression of the alternative genes. Thus, cell determination can be regarded as a pattern formation among alternative genes. Based on this similarity, I have proposed that gene activation requires a direct or indirect feedback of a gene product on the activation of its own gene and their mutual competition such that only one of the alternative genes can remain active in a particular cell. (Fig. 3). Meanwhile many such autoregulatory genes have been found.

Based on ligation experiments with early insect embryos I have proposed that starting from a default gene activity, other genes become activated in a step-wise manner. Each further step requires a higher morphogen concentration. This process comes to rest if the actually activated gene corresponds to the local morphogen concentration (Fig. 3). The patterning of the hindbrain under the control of retinoic acid follows this mode of regulation. To require higher and higher morphogen concentrations for such a promotion, the system has to be less and less sensitive for the morphogen. In order that these less-sensitive feedback loops are able to dominate over the more sensitive loops, the 'higher' genes must be stronger in the autoregulation. Due to the positive feedback in the gene activation together with the unidirectional promotion, the process is essentially irreversible. A reduction of the signalling molecules is without effect since the morphogen is not required for the maintenance of the gene activity.
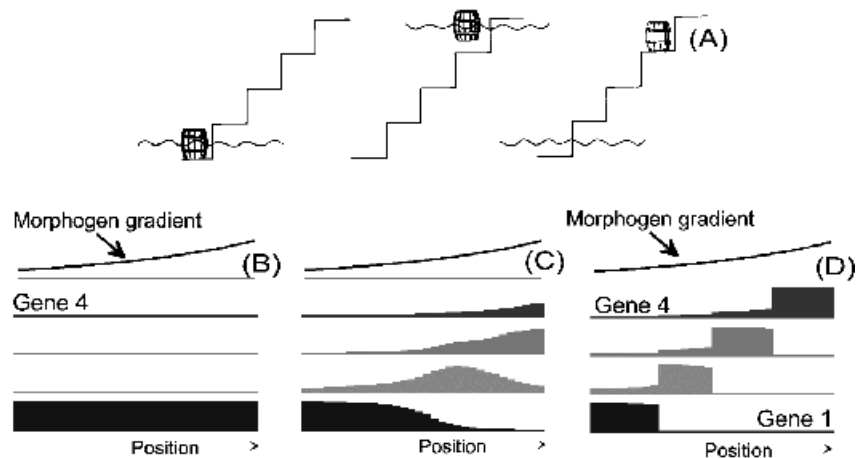
Figure 3: Space-dependent gene activation by a morphogen gradient. (A) An analogy: a flood can displace a barrel to a higher level on a staircase. After lowering of the water level, the barrel remains on a particular level that corresponds to the highest level the flood has obtained. (B-D) A set of genes 1, 2... is assumed whose gene products feed back on the activation of the corresponding genes. In addition, all genes compete with each other for activity. This has the consequence that only one of the genes can be active within one cell. Depending on the local level of the morphogen, a promotion from gene 1 to gene 2 and so on takes place until the local morphogen concentration is insufficient for a further step. Sharply confined regions of gene activities emerge despite the fact that the initiating signal is graded. The once achieved gene activation is maintained even if the promoting signal is no longer available.

4 Somite formation: sequential conversion of a periodic pattern in time into a periodic pattern in space

The formation of somites (Fig. 4 a, b) is a crucial step in the primary anteroposterior patterning of vertebrates. Somites are the precursors of many essential structures including the vertebrae and the body musculature. During such a metamerization two patterns are laid down. A periodic pattern consisting of a serial repetition of homologous structures. Superimposed is a sequential pattern, which makes the repetitive subunits different from each other. For instances, only the thoracic somites form vertebrae that bear rips.

To be compatible with classical observations I proposed in 1982 that somites are generated by a stepwise conversion of a periodic pattern in time into a periodic pattern in space (Fig. 4). Although somites are separated from the presomitic mesoderm in an anterior-to-posterior sequence, the counter-intuitive prediction was made that the specification of anterior and posterior half-somites occur by wave-like processes that are initiated at the posterior end of the presomitic mesoderm and move toward anterior until they come to rest at the correct distance from the last formed half-somite. The prediction has been confirmed by the observation of the *c-hairy1* gene in the chicken (Palmeirim et al., 1997) that behaves as expected for a signal generating the posterior half-somite. (Fig. 4 c, d). Also the prediction that this oscillation is used to activate genes that specify the individual character of the somites has found meanwhile support.
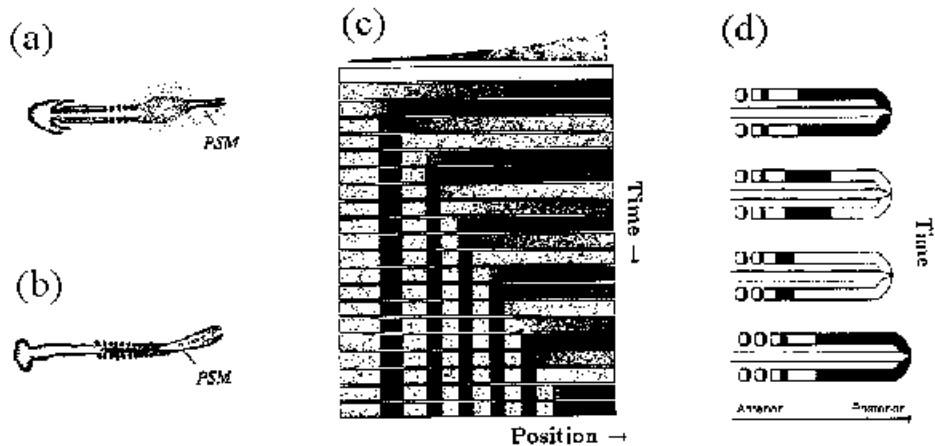
Figure 4 : Somite formation. (a, b) In chicken, somites separate from the non-segmented presomitic mesoderm (PSM) in an anterior-to-posterior sequence. At 25hr of incubation about 5 somites are visible. Ten hours later, about 12 somites are present. (c) Model (1982): Under the influence of positional information (top), cells oscillate between two states, A (white) and P (black). Only cells above a threshold can participate. This leads to a first A/P border. At such a border, the two states stabilize each other mutually. Each full further cycle leads to an additional pair of half-somites. In the course of time a transition from a periodic pattern in time to a periodic pattern in space occurs. The model predicted that waves originate at the posterior pole, move towards anterior and come to rest were the next somite has to be formed. This scheme found full support by the observation of the oscillating expression of *c-hairy* in the presomitic mesoderm of the chick by Palmeirim et al. (1997).

Further readings:

- Many simulations in an animated form and references can be found on our website http://www.eb.tuebingen.mpg.de/meinhardt

- For the modelling axes formation in vertebrates see:
    - Models for organizer and notochord formation. *Comptes Rendus Biol*., 323: p23-p30, 2000.
    - Organizer and axes formation as a self-organizing process. *Int. J. Dev. Biol*., 45: p177-p188, 2001.
    - The radial-symmetric hydra and the evolution of the bilateral body plan: an old body became a young brain. *BioEssays,* 24: p185-p191.

- The elementary mechanisms of biological pattern formation are treated in the book "Models of Biological Pattern formation" (1982; Academic Press). A remake of this book is available for download at: http://www.eb.tuebingen.mpg.de/dept4/meinhardt/82-book/Bur82.htm

- A book in which the beautiful patterns on tropical seashells have been used as a natural picture book to illustrate principles of pattern formation: "The Algorithmic Beauty of Sea Shells". 2003, 3rd Edition; Springer, Heidelberg, New York:
    http://www.springer.de/cgi/svcat/search_book.pl?isbn=3-540-44010-0
    This book contains a CD with animated simulations and programs that run on a PC. In these programs, parameter can be changed and the effect on the pattern formation can be inspected. New interactions can be introduced if the programs are recompiled (corresponding BASIC compilers are freely available from the web).
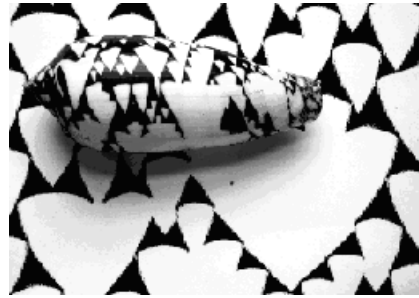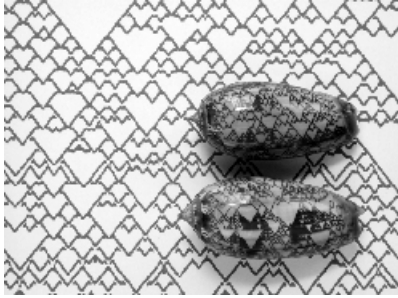
Since a mollusc can enlarge its shell only at the growing edge, the patterns are time records of a one-dimensional patterning process. This allows to decode that underlying patterning process. Above are two examples. The oblique lines in *Oliva prophyria* result from travelling waves of pigment production in the shell-forming zone. Remarkable are the branches, which result from a spontaneous trigger of backwards waves – an unusual behaviour for waves in excitable media. The triangles on *Conus marmoreus* (right) result from two travelling waves, a slow one that switches pigmentation on and a fast one that switches it off.

Microtubule self-organisation as an example of the development of order in living systems

James Tabony [1], Nicolas Glade [1,2], Cyril Papaseit [1] & Jacques Demongeot [2]

[1] Commissariat à l'Energie Atomique, Département Réponse et Dynamique Cellulaires, Laboratoire d'Immunochimie, INSERM U 548, D.S.V, C.E.A. Grenoble, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France.

[2] Institut d'Informatique et Mathématiques Appliquées de Grenoble, Laboratoire des Techniques de l'Imagerie, de la Modélisation et de la Cognition, Faculté de Médecine, Domaine de la Merci, 38706 La Tronche Cedex, France.

Abstract

We address the physical chemical processes underlying biological self-organisation by which a solution of reacting chemicals spontaneously self-organises. Theoreticians have predicted that macroscopic self-organisation can arise from a non-linear coupling of reactive processes with molecular diffusion. In addition, the presence of an external symmetry-breaking factor, such as gravity, can determine the morphology that subsequently develops. We have found that the formation *in vitro* of microtubules, a major element of the cellular skeleton, from tubulin, shows this type of behaviour. These preparations spontaneously self-organise by way of reaction and diffusion, and the morphology that develops depends upon the presence of a weak external factor, such as gravity or a magnetic field, at a critical bifurcation time early in the process. Once assembled from tubulin, microtubules grow and shrink from opposite ends by reactive processes involving the addition and loss of free tubulin. The shrinking end of a microtubule leaves behind itself a chemical trail of high tubulin concentration. Neighbouring microtubules preferentially grow into these regions, whilst avoiding regions of low tubulin concentration. The chemical trails produced by individual microtubules thus activate and inhibit the formation of neighbouring microtubules and this progressively leads to self-organisation in a manner that show analogies with the way that ants self-organise. Numerical simulations of the reaction-diffusion process based on the chemical dynamics of a population of microtubules successfully predict the main features of the experimental behaviour. These simulations provide insight as to how self-organisation occurs at a microscopic level and how weak external factors trigger this process. Evidence is presented that processes of this type occur *in vivo* during both embryogenesis and in the course of the cell cycle.

1   Introduction

The mechanisms by which biological self-organisation occur from a largely unstructured state, such as is initially present in a developing egg or seed, remain poorly understood. Although over the last two decades, studies in molecular biology have started to reveal some of the genes involved, our understanding of the physical-chemical laws underlying these phenomena remain uncertain. There are two possible physical-chemical approaches that might account for biological self-organisation and pattern formation; one is based on static interactions and statistical physics; the other is based upon non-linear chemical dynamics and co-operative phenomena. This article is wholly concerned with the latter approach.

Normally, solutions of reacting chemicals do not self-organise. On the contrary, in living organisms, order and form develop spontaneously by way of biochemical processes from a state that initially is largely devoid of these properties. For example, during the early stages of embryonic

development, patterns of certain gene products come about that determine the morphology and body plan of the organism that subsequently develops. Identifying and putting a name to the molecules that form these patterns does not as such tell us anything about the physical chemical process of how the pattern comes about. In addition, the early stages of biological development is characterised by the fact that beyond a certain critical stage, cells of identical genetic content take different developmental pathways so as to become differentiated from one another.

One way of describing these phenomena is to identify all the molecules involved and establish how biological functions are related to them. This is the classical biochemical approach. However, another approach is slowly developing. Some scientists are asking whether a certain number of global biological properties might not be described in terms of what are now known as 'complex' systems and 'emergent' phenomena. A 'complex' system [1-3] is comprised of a population of units that mutually effect one another and show non-linear dynamics. In such systems, a certain number of new 'emergent' phenomena arise. These 'emergent' properties are not properties intrinsic to the individual elements present in the system, but on the contrary come about from the way that the individual elements 'talk to one another' and behave as a collective unit. In many 'complex' systems, self-organisation is an important 'emergent' phenomenon. A particular feature of some of these systems is that self-organisation can be strongly affected by the presence of weak external factors that break the symmetry of the system and so modify its collective behaviour. Striped arrangements often arise; when they do, they are nearly always the result of an outside external perturbation that induces a directional bias on the actions of the individual.

Living systems consume energy in one form or another and this implies that they are out of thermodynamic equilibrium. Systems that are far-from-equilibrium are often characterised by non-linear dynamics and consequently living systems provide many examples of self-organisation by collective processes [3, 4]. A well-studied example of this type of behaviour is the formation of ant colonies. A moving ant leaves behind itself trails of chemicals, known as pheromones, which can attract or repulse other ants. An ant encountering a trail of an attractive pheromone will change its direction to follow the trail. This ant, will, in its turn, deposit more pheromone on the trail thus reinforcing it [5]. The self-amplification of attractive and repulsive pheromone trails leads to the self-organisation of the ant population. It turns out that there are many similarities between the collective behaviour of ants and the manner by which microtubules self-organise.

One of the advantages of this type of process is that ants can rapidly establish the shortest route between a food supply and their nest [3, 6, 7]. Consider a situation where there are two identical food supplies close to an ant colony. One source is slightly closer to the nest than the other. As ants return to the nest with food, they leave chemical trails that are followed by other ants. These ants reinforce the trails, and so more and more ants follow the paths to the two food supplies. However, from the closer of the two sources it takes less time for an ant to return to the nest. This leads to a larger number of ants taking this path, thus reinforcing the strength of its chemical trail, compared with the longer path. Hence, progressively more and more ants take the shorter path to the closer food supply until they all follow this route. If the food supplies are at exactly equal distance from the nest, then a weak external factor will suffice to favour one pathway over the other and hence determine which of the two food sources the ants consume. It is easy to see that the choice as to which route develops is determined at an early stage of the process before the reinforcement of the pathway has gone very far. As pathway reinforcement progresses, then it takes the application of an increasingly stronger external factor to induce a change whereby the alternate pathway develops. In the absence of such an effect, the deciding external factor need only be present for a critical period at the beginning of the process. Once pathway reinforcement has started, it will continue until the food source is completely consumed. At the beginning of this process, the distribution of ants is unstable to weak external factors. This is a simple example of a bifurcation in a non-linear dynamic system.

The behaviour described above is a typical example of 'emergent' properties in a 'complex' system. A question that arises is how processes of this type might come about at a molecular level by way of biochemical reactions within biological objects such as an egg or a cell. Although solutions of reacting chemicals or biochemicals do not normally self-organise, nevertheless, since the 1930's some theoreticians have predicted that certain types of chemical reaction, when sufficiently far-from-equilibrium, might show 'complex' chemical dynamics which gives rise self-organisation [8-17]. At a molecular level, self-organisation results from a combination of reaction and diffusion. Chemical energy is consumed, and starting from an initially homogenous solution, a stationary chemical pattern made of periodic variations in the concentration of some of the reactants, develops. We are all familiar with the fact that diffusion leads to mixing and the loss of order. The idea, that the contrary might arise, i.e. a suitable coupling of diffusion with reaction can bring about a partial separation of chemical products is not at all intuitive and constitutes a change of paradigm.

The first scientists to have outlined this possibility seem to have been Kolmogorov [8] in 1937 and Rashevsky [9] in 1940. In 1952, Turing published a fully developed theoretical model of self-organisation by reaction and diffusion [10]. Over the next 30 years, starting from considerations of out-of-equilibrium thermodynamics [11], Prigogine, Glansdorf and Nicolis made very substantial developments to the subject [12-15]. Self-organised structures of this type are often called reaction-diffusion or Turing-like structures. They also go under the name of dissipative structures. The latter term was widely used by Prigogine and co-workers because a dissipation of chemical energy is required to drive and maintain the system sufficiently far-from-chemical equilibrium that self-organisation occurs. Even though such terms were not used at the time, what these theoreticians predicted was that biological self-organisation could arise as an 'emergent' phenomenon in a 'complex' system by molecular processes of reaction and diffusion. Rashevsky, Turing, Prigogine and others, all postulated that such mechanisms might provide an underlying basis for biological morphogenesis and self-organisation.

In addition to self-organisation, such reaction-diffusion systems can also show bifurcation properties [2, 14]. For a bifurcation of the bistable type, there exists a critical moment at which the initial state is unstable [18]. At this moment, early in the self-organising process, the presence of a weak external field, such as gravity or a magnetic field, can determine the morphology of the state that subsequently develops. For certain reaction-diffusion systems, Kondepudi and Prigogine [19, 20] explicitly calculated that terrestrial gravity could have this effect. Once the bifurcation has occurred, the system evolves progressively along the selected pathway to the pre-determined morphology. It behaves as though it retained a memory of the conditions prevailing at the bifurcation.

These concepts, although a subject of interest and debate for many years, are only now beginning to be progressively accepted by biologists. There are several reasons for this long delay. The first is conceptual; there has been (and still is) a great reluctance to accept that pattern and form can arise from non-linear chemical dynamics and not from static interactions. However, another more practical reason is that until relatively recently there were no examples of chemical and biochemical systems accepted as behaving in this way. For example, in chemistry, although the Belousov-Zhabotinsky reaction, discovered in 1951 [23, 24], was cited by Glansdorff and Prigogine [13] as an example of a dissipative structure, it was not until 1990 that a similar type of reaction, first discovered in 1917, was at last accepted by the majority of workers in the field as the first example of a Turing-like structure [21, 22]. The same situation has prevailed in biology. Many authors have compared the morphologies that occur in biological organisms with the mathematical predictions of reaction-diffusion theories. There is a substantial body of literature in this area [25-28]. Workers [29-31] have demonstrated that the patterns of calcium waves that come about *in vivo* in the cytosol arise from reaction-diffusion processes. However, one of the factors that has been lacking is the

concrete example of a molecular biochemical system behaving in the general manner predicted by Rashevsky, Turing, Prigogine, and others. Until the work reported here on the self-organisation of microtubules, first published in 1990 [32], there were no *in vitro* examples of a simple biochemical system in a test tube, known to self–organise in the general manner predicted by Kolmogorov, Rashevsky, Turing, Prigogine, and others.

Under appropriate conditions, *in vitro* preparations of microtubules, a major component of the cytoskeleton, show this type of behaviour [32-49]. These microtubule preparations spontaneously self-organise by a combination of reaction and diffusion, and the morphology of the state that forms depends upon external factors such as gravity at a critical bifurcation time early in the process.

2   Microtubules

The interior of the cell is organised by the cytoskeleton. The latter is composed of three filamentary components; microtubules, actin and intermediate filaments. Microtubules are long tubular shaped objects, with inner and outer diameters of about 16 nm and 24 nm respectively [50, 51]. They arise from the self-assembly of a protein, tubulin, by way of reactions involving the hydrolysis of a nucleotide, guanosine triphosphate (GTP), to guanosine diphosphate (GDP). Their length is variable; but often they are several microns long.

Microtubules have two major roles; they organise and control the structure of the cytoskeleton, and they permit and control the directional movement of intracellular particles and organelles from one part of the cell to another. They participate in many fundamental cellular functions including, the maintenance of shape, motility, signal transmission, and play a determining role in the organisational changes that occur during the early stages of embryogenesis. Microtubules are a significant component of brain neuron cells and they make up the mitotic spindles that separate chromosomes during cell division.

Tubulin has a monomer molecular weight of about 50 KDa, a diameter of about 4 nm, and occurs as a dimer of the alpha and beta monomer forms. Microtubules can be formed *in vitro* by warming a solution of purified tubulin in the presence of GTP, from about 7 to 35°. During this process, a series of chemical reactions occurs and GTP is hydrolysed to GDP. Once microtubules are formed, chemical activity continues through processes whereby tubulin is added and lost from opposing ends of individual microtubules by reactions involving GTP hydrolysis. There is hence a continual consumption or dissipation of chemical energy through the system.

One of the particularities of microtubules is that due to differences in reactivity at opposing ends, they frequently grow from one end whilst shrinking from the other. Since the rates of growth and shrinkage are often comparable, individual microtubules change position and appear to move at speeds of several microns per minute. Tubulin is incorporated at the growing end as the complex tubulin-GTP and liberated at the shrinking end as tubulin-GDP. The shrinking end of a microtubule leaves behind itself a chemical trail of high local concentration in tubulin-GDP. Whilst progressively diffusing out into the rest of the solution, this tubulin-GDP is progressively reconverted to tubulin-GTP by the excess GTP present, at which point it is once again available to be incorporated at the growing end of neighbouring microtubule. Likewise, the growing end creates a region depleted in tubulin-GTP. Neighbouring microtubules will preferentially grow into regions of high tubulin-GTP concentration whilst avoiding the regions of low concentration. The chemical trails produced by individual microtubules activate and inhibit the formation of their neighbours. Thus, neighbouring microtubules "talk to each other" by depleting and accentuating the local concentration of active chemicals in a manner that has analogies with the self-organisation of ants. Under appropriate conditions, the population of microtubules can behave as a 'complex' system and

the coupling of reaction with diffusion progressively leads to macroscopic variations in the concentration and orientation of the microtubules.

3   *Microtubule self-organisation* in vitro

When microtubules are assembled as described above, then under appropriate conditions spontaneous macroscopic ordering occurs [32]. Following assembly in spectrophotometer cells measuring 4 cm by 1 cm by 0.1cm, a series of periodic horizontal stripes of about 1 mm separation progressively develop in the sample over about 5 hours. Once formed, the striped pattern remains stationary for between 48 to 72 hours, after which the system runs out of reactants. The preparations are of high optical birefringence and this indicates that the microtubules are strongly aligned with respect to one another. The microtubules in each striped band are oriented at either 45° or 135°, but adjacent stripes differ in having alternating orientations. This pattern of variations in orientation can be observed by placing the sample between crossed linear polars with a wavelength retardation plate placed at 45° between them (Fig. 1). The retardation plate produces a uniform background having a mauve interference colour. Microtubule orientations, such that their birefringence adds to the birefringence of the wavelength plate, produce a blue wavelength shift, whereas orientations that subtract cause a yellow shift. Sample regions made up of microtubule orientations that are either acute or obtuse, differ by producing yellow or blue interference colours respectively and this give rise to the alternating blue and yellow bands illustrated in Fig. 1. In addition to these changes in orientation, periodic variations in microtubule concentration, of about 30% of the mean, also occur from stripe to stripe [37]. The concentration pattern coincides with the pattern of variations in microtubule orientation.

The 0.5-mm stripes also contain within them another series of stripes of about 100 µm separation. These, in their turn contain other sets of stripes of about 20 µm, 5 µm and 1 µm separation. Some of these structures are shown in Fig. 2. In samples made up in larger sample containers, an additional level of ordering several mm in separation arises. These large stripes in turn contain the lower levels of organisation already mentioned. Hence, similar types of pattern spontaneously develop encompassing distances ranging from a few microns up to several centimetres. The range of dimension over which these microtubule structures occur is typical of those found in many types of higher organisms. Cells are about 20 µm in size, eggs are often about a mm, and a developing mammalian embryo is several centimetres long. When microtubules are assembled in small sample containers about 100 µm in size, self-organisation still occurs but with periodicity's that are consistent with the size of the sample. This observation demonstrates that microtubule self-organisation by the processes described here also occurs in containers of the dimensions of a cell or an embryo.

Striped morphologies occur when the microtubules are prepared in upright sample containers, but a different pattern (Fig. 1) arises when they are prepared in the same containers lying flat [34, 35]. This behaviour is attributed to the determining role of the gravity direction during s   tructure formation. So as to test this hypothesis, microtubules were assembled in flat horizontal containers fixed to the turntable of a record player, and with the long axis of the sample oriented along the direction of the centrifugal field (0.14 g). A s triped morphology once again forms, and the direction of the stripes is perpendicular to that of the applied centrifugal field [34].

Once formed, the structures are independent of their orientation with respect to gravity. To establish at what moment the sample morphology depends upon the gravity direction, the following experiment was carried out [35]. Microtubule formation was simultaneously instigated in twenty identical rectangular sample cells. Initially all the cells were upright. Consecutive cells w   ere then turned from vertical to horizontal at intervals of one minute, and the samples examined 12 hours

later, after the structures had formed. Twenty minutes after instigating microtubule formation, when the last sample was turned from vertical to horiz   ontal, there are no obvious signs of a striped structure. Since the structure forms while the cells are flat, one might expect that they might all form the horizontal pattern. This is the case for samples inverted during the first few minutes. However, sam ples those were upright for six minutes or more, all formed striped morphologies identical to preparations that remained vertical all the time. The final sample morphology depends upon whether the sample was horizontal or vertical at a critical period, six   minutes after instigating assembly, and at an early stage in the formation of the self -organised structure. The process can be described as a bifurcation between pathways leading to two different morphological states, and in which the direction of the sam   ple with respect to gravity determines the morphology that subsequently forms.

To establish whether the self-organising process is directly dependent on gravity, microtubules were assembled under conditions of weightlessness [39]. The experiment was carried out during the space flight of a sounding rocket of the European Space Agency which provided approximately 13 minutes of weightlessness before the payload fell back to earth and was recovered. Since on the ground, the sample morphology is determined by the orientation with respect to gravity 6 minutes after instigating microtubule assembly, 13 minutes of low gravity should suffice to investigate the effect of weightlessness on the self-organising process. Flight samples were contained in an experimental module divided into two compartments, a "low gravity" compartment and a "1 g on-board" centrifuge compartment. We found that microtubule samples formed in the "centrifuge" part of the module formed stripes when the centrifugal field was parallel to the long axis of the cell, and the circular morphology when it was perpendicular. These morphologies resemble those that arise in the laboratory and show that the self-organising process is unaffected by payload re-entry and recovery. In contrast, the samples formed in the "low gravity" compartment did not self-organise (Fig. 1). Ground based experiments that produce effects close to weightlessness, either by rotating the sample around the horizontal axis or counterbalancing gravity with a strong magnetic field gradient, produce a similar behaviour. It needs to be stressed in these experiments that weightlessness effects microtubule self-organisation but not microtubule self-assembly. Microtubules assemble to the same extent and show the same assembly kinetics under conditions of weightlessness as at 1 g and regardless as to whether weightlessness arises by ground based methods or by space flight.

Experimental problems can occur in space experiments due to air bubbles. In this case, although care was taken to prevent it, small air bubbles formed in the neck of some of the sample cells. During re-entry, when the sample is subject to high centrifugal fields, the air bubble may be pushed through the sample. In one sample this process was filmed. A line of high birefringence formed along the trajectory of the air bubble [41] indicating that the bubble oriented the microtubules along its trajectory. Subsequently, striped regions limited in extent, developed perpendicular to this trajectory. Hence, orienting the microtubules at an early stage in the process can also trigger self-organisation.
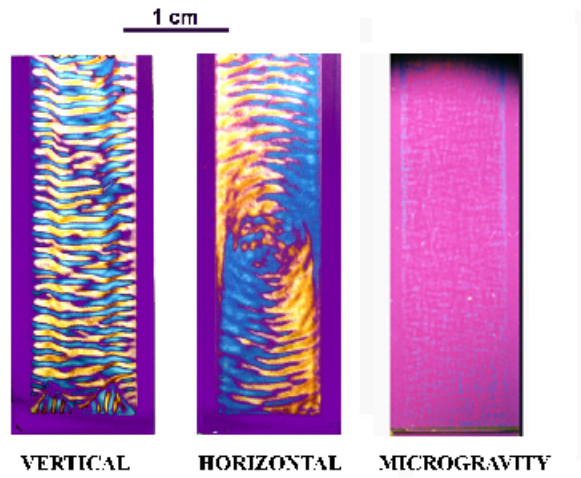
Figure 1: Microtubule structures formed in spectrophotometer cells 4 cm by 1cm by 0.1 cm. In the presence of gravity, the solution spontaneously self-organises. Stripes form when the sample container is upright but circles arise when the microtubules are assembled with the container flat. The structures, which take about 5 hours to form, are stationary and independent of the orientation of the cell with respect to gravity. The morphology that forms, depends on the orientation of the container at a critical moment at an early stage in the self-organising process and before any pattern has developed. No self-organisation occurs when microtubules are assembled in the absence of gravity for the first 13 minutes. Samples were photographed through linear cross polars with a wavelength retardation plate. The blue interference colour arises from microtubules oriented at about 45°, and the yellow interference colour from those at 135°. Periodic changes in microtubule concentration coincide with the changes in orientation.
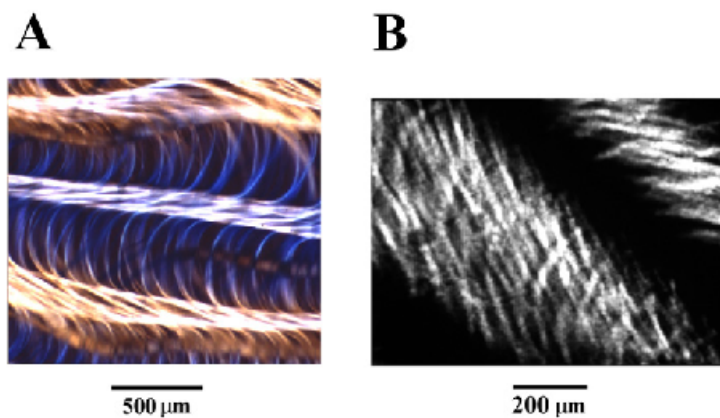
Figure 2: The striped structure, as shown in Fig. 1a b, is itself comprised of stripes of smaller spacing (100 µm, 20 µm and 5µm and 1µm). Photographs A) and B) show images of the preparation at higher magnification.

A different manner of orienting the microtubules is to apply a strong uniform magnetic field. When microtubules were assembled in flat horizontal cells, exposed to a strong horizontal magnetic field for the first 15 minutes of the self-organising process, then a striped 'vertical' morphology formed instead of the circular ' flat' morphology. This demonstrates that orienting the microtubules at the bifurcation time by way of a magnetic field modifies the self-organised morphology in a way similar to a change in the gravity direction. These experiments show that any effect that at the

bifurcation time leads to a partial orientation of microtubules will trigger self-organisation. Another factor that strongly affects self-organisation is sample shape. This is implicit in the fact that for the rectangular shaped samples considered so far, different morphologies arise when the microtubules are assembled with the sample placed flat down or upright.

4   *Self-organisation results from chemical reactions and not static interactions*

An important feature to be established is whether self-organisation results from reactive processes or from static interactions related to the liquid crystalline properties of the microtubule solution. A simple way to test for this is the following. Form the self-organised structure as described above. Then destroy it by mixing and wait and see if the structure reforms [38]. After mixing, the solution contains microtubules at the same concentration and temperature as before. However, the consumption of chemical energy is much less than when the microtubules initially form. If the self-organised structure arises from static interactions, such as occur in some liquid crystals, then the structure will reform after mixing. This is not the case. Microtubules can be disassembled by cooling the solution to 4°C. When the preparation is again warmed to 36°C, microtubules once again form and GTP is hydrolysed to GDP. In this case, the striped self-organised structure also reforms [37, 39]. These simple experiments show that the striped structure arises via chemical processes associated with microtubule formation and maintenance, and not from static interactions between the microtubules.

A further strong argument against static interactions is the dependence of the self-organisation on gravity at a critical moment early in the self-organising process. Static interactions, such as may occur in liquid crystals, are equally present under conditions of weightlessness (0g) as at terrestrial gravity (1g). The absence of microtubule self-organisation under low gravity conditions at the bifurcation time, is a clear demonstration that self-organisation does not arise from static interactions. If the structure involved phase separation under the action of gravity, then although the structure might not form under low gravity conditions, it would form, if a brief period of low gravity were followed by 1g conditions. The fact that this does not occur eliminates this possibility.

Static interactions in liquid crystals, although they can give rise to macroscopic variations in orientational order, do not lead to macroscopic variations in concentration. On the contrary, the central prediction of reaction-diffusion theories is the formation of macroscopic variations in the concentration of reacting species. Both neutron small angle scattering and fluorescence imaging (Fig. 3) demonstrate that substantial macroscopic microtubule concentration variations are present in the self-organised preparations. The microtubule concentration is approximately constant in neighbouring stripes, but drops by about 25% between the stripes, where the orientation flips from acute to obtuse.

Another possibility, which has also been considered and rejected, is that the pattern might in some way involve a coupling of reactive processes with flow due to thermal gradients during the warming process. The microtubule preparations are gels of high viscosity (≈5000 poise) [52]. This renders thermal convection difficult. In addition, samples prepared in a hot room at 35°C, in which there was no thermal convection, gave patterns the same as samples prepared under conditions where the bottom of the sample was warmer than the top by 5°C. It is not necessary to form microtubules by warming a premixed solution of tubulin and GTP from 4°C to 35°C. They may also be formed by mixing together separate solutions of tubulin and GTP, pre-warmed to 35°C. In this case, the self-organised structure that develops is the same as that obtained by mixing tubulin and GTP in the cold and then warming. Hence thermal convection appears to play no part in the self-organising process.
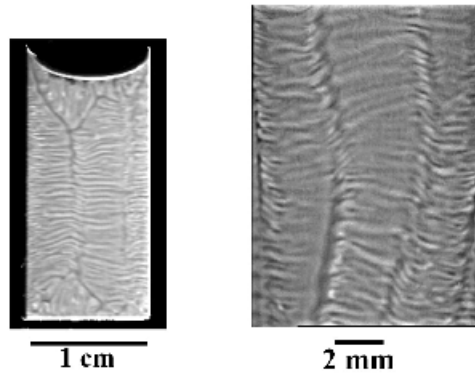
Figure 3: Microtubule concentration patterns as shown by fluorescent imaging. See [38] for details.

An important parameter in reaction-diffusion systems is the rate of energy dissipation. This will be strongly dependent upon experimental variables, such as concentration and temperature. In microtubule preparations, the rate of hydrolysis of GTP to GDP has been determined using P[31] NMR spectroscopy [32, 33, 36, 39, 40]. In a reaction-diffusion system of the Turing type, the periodicity or wavelength of the structure corresponds to approximately the distance over which groups of molecules diffuse before reacting. The periodicity, L, is related to the reaction rate, R, and diffusion constant, D, by terms involving $L^2=R/D$ [27,35]. In agreement with this, we found in the self-organised microtubule preparations, that increasing the reaction rate by a factor of 2 decreased the stripe separation by 1.4, i.e. the square root of the increase in the reaction rate [27].

The rate of diffusion is not varied as readily as the reaction rate. One simple approach is to examine the effect on self-organisation of the addition to the initial reaction mixture of small quantities of gelling agents. Increasing quantities of gelling agent increase the viscosity of the preparation, and inhibit diffusion. When microtubule assembly was carried out in gels of increasing agarose concentration, the effect was to perturb and eventually inhibit self-organisation [36, 37]. This, and other observations concord with a diffusive contribution to the self-organising process.

During the initial stages of self-organisation, the left and right hand sides of the sample show either yellow or blue birefringent interference colours corresponding to either obtuse or acute microtubule orientations. The striped structure subsequently develops by blue zones forming in the yellow region, and yellow zones forming in the blue region. In the zones where there is no colour change, the microtubules retain their initial orientation. Whereas, in those regions where there is a colour change occurs, the microtubule orientation flips from acute to obtuse, or vice versa. In neutron small angle scattering measurements, that were limited to a horizontal band of dimensions of an individual stripe [35], this process is manifested as a change in the direction of the microtubule scattering on the detector, from an acute to an obtuse arc. Simultaneous with this orientational re-ordering, the intensity of the microtubule scattering, decreases, then rises before declining again. Orientational re-ordering, which is itself the stripe forming process, is hence concurrent with a chemical wave, involving different concentrations of microtubules and free tubulin, crossing the sample area under investigation. In other words the stationary pattern arises because microtubules disassemble and reassemble with different orientations and concentrations in alternating parts of the sample. This neutron scattering experiment clearly shows that self-organisation is associated with the reaction dynamics of the microtubules.

The conclusion agrees with what many biologists have long known. In many cellular processes involving microtubules, changes in microtubules organisation occur not by the rearrangement of an existing structure, but by disassembly followed by reassembly to form a different arrangement

within the cell. Biologists are fully aware that much of the intriguing behaviour of microtubules comes about from the fact that they are transient dynamic entities in which structural rearrangements and self-organisation are associated with their reaction dynamics. The question that arises is to establish the exact molecular basis by which microtubule reaction dynamics result in self-organisation, and investigate whether this is the same *in vitro* and *in vivo*?

5   *Molecular basis of self-organisation by reactive processes*

The kinetics of formation of the self-organised microtubule preparations shows an 'over-shoot' about 6 minutes after instigating microtubule formation [35]. The bifurcation point in any out-of-equilibrium system, and at which point the system is sensitive to weak external factors, coincides with a condition of instability in the homogenous state [18]. In the microtubule system, the 'overshoot' describes a chemical instability in the relative proportions of free tubulin and microtubules. This occurs at the bifurcation time, at the moment that the system is gravity dependent. The microtubules assemble, and then partially disassemble by about 20% before approaching a constant level after about 60 minutes.

Microtubules are continually growing from one end and shrinking from the other. We postulate the following (Fig. 4) [40]. The shrinking end of a microtubule leaves behind it a chemical trail of high tubulin concentration. Likewise the growing end produces regions depleted in tubulin. A neighbouring microtubule will preferentially grow into a region of high tubulin concentration whilst avoiding the regions of low concentration. The chemical trails produced by a given microtubule, will modify and determine the direction of growth of its neighbours. Thus neighbouring microtubules "talk to each other" by depleting and accentuating the local concentration of active chemicals. The coupling of reaction with diffusion progressively leads to macroscopic variations in the orientation and concentration of the microtubules.

When the microtubules first form from the tubulin solution they are still in a growing phase. There is almost no disassembly from their ends and the microtubules are distributed uniformly through the solution in an isotropic manner. However, the rapid initial growth of the microtubules results in a reduction of the free tubulin concentration in the solution. The depletion of tubulin by the growing microtubules, causes unfavourable reactive conditions, and triggers their own partial disassembly and this manifests itself as the 'overshoot' shown in the assembly kinetics. When disassembly does start to occur, just prior to the bifurcation time, it leads to the formation of the chemical trails outlined above. The isotropic arrangement of microtubules is now unstable. At this point in time (the bifurcation time), orienting just a few microtubules will induce their neighbours to grow along the same orientation. Once started, the process mutually reinforces itself with time and leads to self-organisation. Hence, in agreement with experiments, any small effect that partially orients microtubules over the entire sample, or leads to a privileged direction of microtubule growth, will trigger self-organisation.
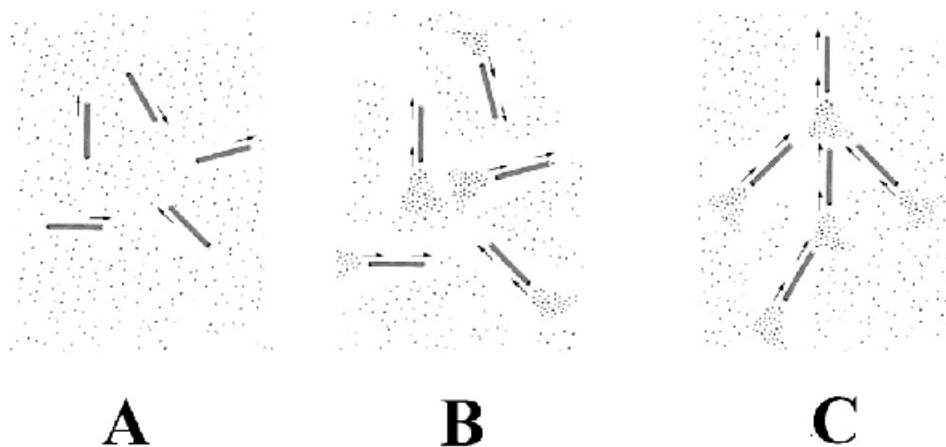
Figure 4: A possible mechanism for the formation of the self-organised structure. Microtubules are chemically anisotropic, growing and shrinking along the direction of their long axis. This leads to the formation of chemical trails, comprised of regions of high and low local tubulin concentration from their shrinking and growing ends respectively. These concentration trails (density fluctuations) are oriented along the direction of the microtubule. Neighbouring microtubules will preferentially grow into regions where the local concentration of tubulin is highest. In A), microtubules have just formed from the tubulin solution. They are still in a growing phase and have an isotropic arrangement. In B), microtubule disassembly has started to occur at the bifurcation time. This produces trails of high tubulin concentration from the shrinking ends of the microtubules. In C), microtubules are growing and forming preferentially into these tubulin trails. The isotropic arrangement shown in B) is unstable. Once a few microtubules start to take up a preferred orientation then neighbouring microtubules will also grow into the same orientation. Once started, the process mutually reinforces itself with time and leads to self-organisation. Any small effect that leads to a slight directional bias, such as slightly different rates of molecular transport in the up-down and left-right directions, will trigger self–organisation. Gravity acts by way of its directional interaction with the macroscopic density fluctuations present in the solution.

6   Numerical simulations of self-organisation.

To test this hypothesis, we carried out numerical simulations of a population of growing and shrinking microtubules [42, 46]. These simulations incorporated as parameters, experimentally realistic microtubule reaction dynamics and the experimentally determined tubulin diffusion constant. Simulations involving a few microtubules demonstrate both the formation of the tubulin trails outlined above and promotion of the growth of neighbouring microtubules along the same orientation. When the simulations were extended to a population of about 10^4 microtubules on a two-dimensional reaction space, 100 µm by 100 µm, then after 2-3 hours of reaction time, a self-organised structure comprised of regular bands of about 5 µm separation developed. This structure is comparable with the experimental self-organised structure that arises over a similar distance scale (Fig. 5).

In the course of these simulations, we noticed that the direction of the stripes was always along the diagonal of the reaction space and this suggested that a directional bias had been unwittingly built into the algorithm. This turned out to be the case and arose from a small asymmetry in the way that the tubulin diffusion was digitised. When the asymmetry in diffusion was removed,

macroscopic self-organisation did not come about. A bias was then reintroduced into the algorithm that broke the symmetry of the system by either of two ways. Either by orienting some of the microtubules at the bifurcation time (this would resemble the manner that magnetic fields act on the system) or by making tubulin diffusion anisotropic at the bifurcation time (this would resemble the manner that gravity acts on the system). Gravity triggers self-organisation in the following way. At the bifurcation time, gravity interacts with the strong density fluctuations produced by the partial disassembly of the microtubules. This interaction causes a 'drift' term that breaks the symmetry of the transport processes. By promoting microtubule growth along a specific direction over the entire sample, self-organisation is triggered. Sample shapes effect self-organisation because microtubule reaction dynamics are modified in the region close to the boundary in such a way that microtubule growth and hence orientation parallel to the boundary is favoured. When the shape of the boundary is strongly asymmetric this effect can suffice to trigger self-organisation without any assistance from gravity. This is more fully described by Glade et al [42, 46].
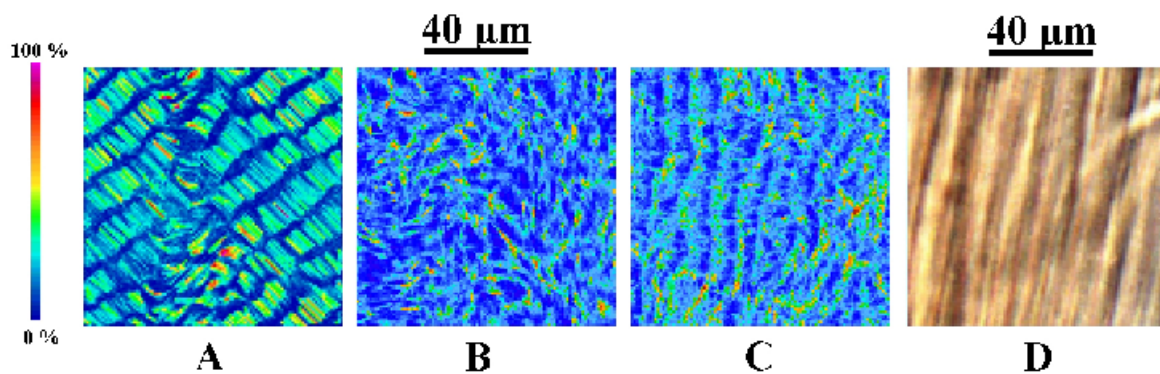
Figure 5: Reaction-diffusion simulations for a population of microtubules. A) is a simulation on a reaction space, 100 µm by 100 µm, containing $4\ 10^{4}$ microtubules. The diagonal stripes are triggered by a small asymmetry in the part of the algorithm describing diffusion. B) is a simulation in which this asymmetry is eliminated. The reaction space is 100 µm by 100 µm and contains the same number of microtubules. Although concentration inhomogeneities are present there is no macroscopic self-organisation. In C) the simulation is identical to B) except that diffusion is now twice as fast along the y-axis as along the x-axis. In this case, self-organised stripes develop in which the microtubules are perpendicular to the direction of the stripes. D) Experimentally observed self-organised structure over the same distance scale.

7 *Do microtubule reaction diffusion processes occur* in vivo*?*

Rashevsky, Turing and Prigogine first developed their theories as a possible underlying physical-chemical explanation for biological self-organisation during embryogenesis. They predicted a way by which macroscopic chemical patterns could spontaneously develop from an initially unstructured egg. The results obtained on the *in vitro* microtubule system demonstrate that reaction-diffusion processes involving biochemical reactions can result in self-organisation. The question obviously arises as to whether these processes also occur *in vivo*; in particular do they occur during embryogenesis and the cell cycle? To find out whether or not this is so, it is paramount to devise and carry out experiments on *in vivo* systems that distinguish between microtubule self-organisation by reaction-diffusion and that from other causes. One of the characteristic properties of the microtubule reaction-diffusion process is that self-organisation can be either triggered or modified by changes in external factors such as gravity, magnetic fields and sample geometry. Hence, experiments showing that microtubule organisation in cells or embryos is modified by upon the

presence of these factors at a critical time, in a manner resembling either their *in vitro* behaviour or the predictions of numerical simulations, is strong evidence in favour of such a process. One of the major effects that our *in vitro* studies suggest, is that microtubule organisation could be strongly modified when either cells or embryos are placed under conditions of weightlessness or in a modified gravity environment.

For several decades, experiments in space [53, 54] have been furnishing an increasing body of evidence that various cellular processes, such as growth rates, signalling pathways and gene expression are substantially modified when various cell types are placed under conditions of weightlessness [55-58]. A substantial number of experiments point to an involvement of the cytoskeleton [58-63]. Recently, researchers have observed substantial modifications in the organisation of the microtubules under conditions of weightlessness. Human lymphocyte (Jurkat) cells cultured in space show a disorganised microtubule network compared to ground control experiments [60]. Likewise, for human breast cancer cells (MCF-7) cultured under conditions of weightlessness [61], the author's report that many cells exhibit a strongly disorganised microtubule network (Fig. 6). Glial cells when placed under conditions of reduced gravity [62] show a disorganised microtubule network and weightlessness has been reported to modify microtubule organisation in rat utricular hair cells [63]. These observations, which are consistent with the in vitro behaviour discussed, suggest that microtubule reaction-diffusion processes might well be occurring in many cell types under normal 1 g conditions.
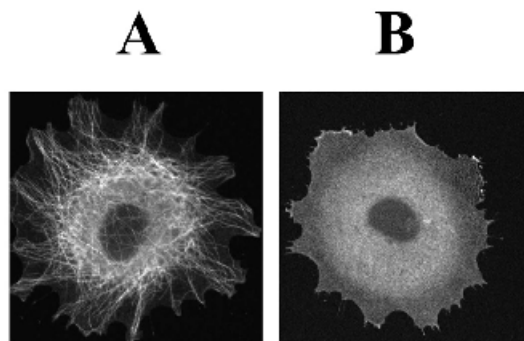
Figure 6: Microtubule organisation in human breast cancer (MCF 7) cells. A) at 1g ; B) at 0 g. Microtubules do not organise under low gravity conditions. Reproduced from Vassy et al [61] by permission of FASEB J.

In certain types of egg, it has long been known that gravity is involved in the early developmental stages [64, 65]. For example, serious malformations result, when *xenopus* embryos are rotated through 90° at a critical time when the so-called 'grey crescent' forms [66]. In *xenopus* eggs, grey crescent formation and its gravity dependence are an essential factor in determining the body plan of the organism. The grey crescent is comprised of a macroscopic array of aligned microtubules [67] and it is quite plausible that both its formation and gravity dependence result from the type of reaction-diffusion processes outlined above.

Another example where microtubule reaction-diffusion processes may arise during embryogenesis, is the formation of striped patterns in *drosophila* fruit fly eggs. In these eggs, the early stages of development occur by consecutive nuclear divisions in a non-compartmentalised space [68] and organisation of the cytoplasm by microtubules is known to play a major role in the morphogenetic processes that occur. Between nuclear divisions 10 to 14, cells progressively form at the surface of the embryo. These cells remain open towards the inside of the egg until the end of the 14th nuclear division. Just prior to this, for around 5 minutes when the ventral and cephalic furrows

appear at gastrulation, the distribution of microtubules in the egg displays a striped arrangement (Fig. 7) [69, 38, 45]. Although the contrast is low, twelve stripes can be counted, and there may be 2 additional stripes of lower intensity. The stripes occur in the central part of the egg only; the end regions are not striped. This pattern coincides with and arises at the same time, as the pattern formed by the segmentation gene product *engrailed* and which plays a major role in determining the segmentation pattern of the larvae that develop [70].
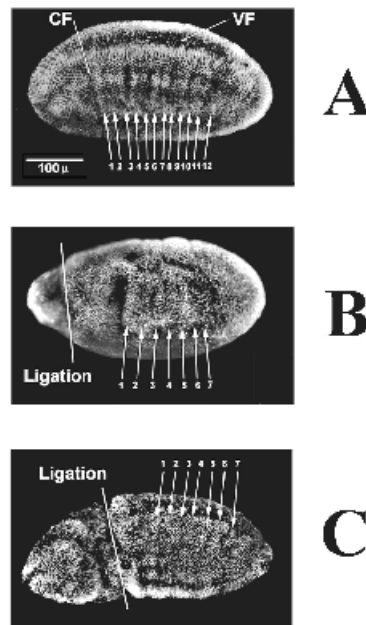
Figure 7: Microtubules patterns observed by immunofluorescence, in A) whole, and B, C,) ligated *drosophila* eggs. The position of the ventral and cephalic furrows, indicating gastrulation, are shown in A). Ligation divides the egg into two unconnected fragments, and development continues in one, or both, of the fragments.
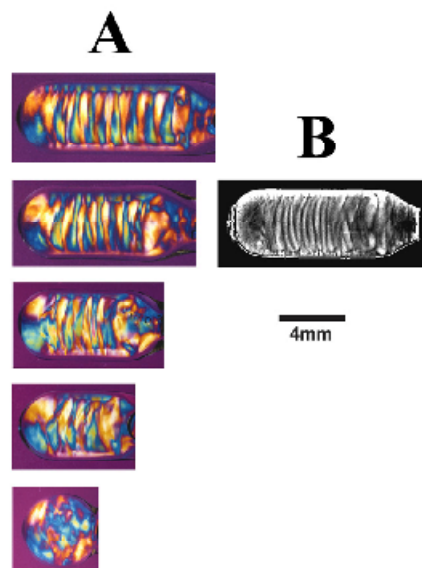
Figure 8: Effect of sample length on microtubule patterns. A) Birefringent patterns formed by microtubules assembled in cylindrical 'egg' shaped sample containers of different length. B) Microtubule concentration variations as detected by fluorescence imaging. The overall morphology resembles the microtubule pattern in *drosophila* eggs.

As mentioned above, microtubule self-organisation *in vitro* by reaction-diffusion processes is affected by sample shape. When self-organised structures are prepared in cylindrical containers, whose shape mimic that of a *drosophila* egg, the morphology shown in Fig. 8 arises [38, 45]. As for the microtubule pattern in the *drosophila* egg, there are two stripe-free regions at each end of the sample, separated by a striped central zone. The exact morphology of the *in vitro* pattern depends on the length of the sample. Below a certain critical length, the striped central region does not form. For longer samples, the end stripe-free zones remain of the same length and the number of stripes in the central region increase with sample length. For samples of appropriate length, morphologies containing 7 blue and yellow birefringent stripes arise. When observed by fluorescence there are fourteen stripes in the microtubule concentration (Fig. 8). This pattern closely resembles the microtubule pattern in *drosophila* eggs.

Drosophila eggs can be shortened, shortly after they are laid, by ligation. Within ten minutes a membrane forms that separates the egg into two unconnected fragments. Development can occur in either one, or sometimes both, parts of the egg. The dependence of the microtubule pattern on sample length is a feature of the *in vitro* pattern. We examined microtubule patterns in ligated *drosophila* eggs as a function of egg fragment length [38, 45] (Fig. 7). Although it is not easy to count the exact number of stripes, nevertheless, as for the *in vitro* microtubule preparations, the microtubule pattern is clearly comprised of two stripe-free regions separated by a striped central region. Also, as for the *in vitro* microtubule patterns, the length of the end zones is independent of fragment length and approximately the same as in unligated eggs. For the case of the ligated egg shown in Figure 7c, both fragments have continued to develop. However, the shorter fragment does not show any stripes. This is consistent with the fact that in this case the fragment length is below the critical length necessary for stripe formation.

The behaviour of the microtubule pattern in the *drosophila* eggs resembles that observed for *in vitro* microtubule self-organisation in samples of different length, and this suggests that similar reaction-diffusion processes might be occurring in both cases.

In addition to the cases outlined above, there is also evidence that microtubule organisation by reaction and diffusion occurs in peripheral nerve cells and during plant cell development [36]. The sum of these observations, and which are consistent with the *in vitro* behaviour, provides an increasing body of evidence that microtubule reaction-diffusion processes occur *in vivo* both during embryogenesis and the cell cycle.

8 Conclusions

The results summarised in this article demonstrate that microtubule preparations self-organise by way of reaction and diffusion. Their behaviour is that of a chemically dissipative system in the general framework advanced by Kolmogorov [8], Rashevsky [9], Turing [10], and Prigogine, Glansdorff and Nicolis [11-15]. They demonstrate how a very simple biological system comprised initially of just a protein and a nucleotide as a source of energy, and without DNA, can a show a complex behaviour reminiscent of certain aspects of living systems. These complex phenomena which are of considerable biological importance, appear as a consequence of chemically non-linear dynamics.

The molecular basis of self-organisation is the elongation and contraction of individual microtubules by the addition and loss of tubulin. Under appropriate reactive conditions, microtubule contraction leads to the formation of chemical trails of free tubulin that can then be incorporated, into the growing end of neighbouring microtubules. Microtubule elongation produces furrows depleted in tubulin. These in their turn activate and inhibit the formation of neighbouring microtubules. Thus neighbouring microtubules "talk to each other" by depleting and accentuating

the local concentration of active chemicals. The coupling of reaction with diffusion progressively leads to macroscopic variations in the concentration and orientation of the microtubules. There are many analogies between microtubule self-organisation and that of ants.

In this system, there are at least two significant differences from the type of reaction-diffusion scheme originally proposed by Turing. In the Turing system, the molecules communicate with one another across the sample by way of diffusion (fast diffusion of the inhibitor and slow diffusion of the activator). In the microtubule system, the molecules communicate by way of the reactivity of the microtubules. The fact that the microtubules are growing from one end at the rate of the order of a few μm per minute whilst shrinking by approximately the same, results in a displacement of the microtubule at the same speed. At the same time they leave behind a trail having a high concentration in inactive tubulin and create in front of them a furrow depleted in active tubulin. These chemical trails determine the pathways taken by surrounding microtubules. It is a reaction-diffusion system for without tubulin diffusion at the appropriate rate, self-organisation would not occur.

The second major difference with a normal Turing system is the reactive anisotropy of the microtubule system. In a normal reaction-diffusion scheme, the reaction has no inherent anisotropy. This is not the case with microtubules, which can obviously only react at their ends. A microtubule can move in only one direction and this will equally lead to anisotropic reactive tubulin trails along the same direction. The system has an in-built propensity for symmetry breaking under the effect of a weak external factor.

At this stage, it is not clear whether these processes are widespread in biology, or if they are limited to microtubules. It is quite possible that in living organisms other reaction-diffusion processes involving different molecular systems occur. Under appropriate conditions, actin will probably show a similar behaviour. It may be that the specific type of reaction-diffusion mechanism encountered here, based on reactive growth and shortening of tubes or rods, is a mechanism particularly suited to biological self-organisation.

The overall phenomenological behaviour of the microtubule preparations shows a qualitative resemblance to some aspects of living organisms in that a number of global properties emerge that frequently arise in living systems; namely self-organisation at different distance levels, a dependence on weak effects, and a primitive form of memory. Firstly, macroscopic ordering appears spontaneously from an initially homogenous state. Secondly, the morphology that forms depends on small differences in conditions at a critical moment early in the process. In the present case, external factors such as gravity trigger self-organisation. The direction of the external factor, by effecting the actions of the individual components, breaks the symmetry of the homogenous state and leads to the emergence of form and pattern. Such processes might hence act as a mechanism for biological signal transduction, and subsequently lock the system into a developmental pathway. Processes of this type could form a general class mechanism by which weak environmental factors are transduced into biological systems and could have played a role in the development of life on earth.

The attraction of reaction-diffusion dissipative processes as a possible explanation for biological self-organisation and development arises not only from their self-organising properties. It is also due to the fact that their bifurcation behaviour could account for some of the global aspects of biological development whereby cells of identical genetic content take different developmental pathways so as to become differentiated from one another. Just after bifurcating, a non-linear system could be described in biological vocabulary as being 'determined but not yet differentiated'.

The application of theoretical concepts of 'complex' systems to biology has to some extent been retarded by the absence of simple biochemical examples. The *in-vitro* microtubule system described above will be of value as an experimental model and will permit comparison with theory. Understanding these processes *in-vitro* is a preliminary to detecting and investigating their role and function in living organisms. The most important question now to be answered, is whether processes of this type also arise *in-vivo*, and in particular do they play an active role during embryogenesis and cell division? At this stage, although preliminary results suggest that this might well be the case, it is too early to give a definite answer. What we can say is that the approach of out-of equilibrium chemical dynamics can in principle account for biological self-organisation, and that an important cellular component, microtubules, behaves this way in a test tube

The general approach of non-linear dynamics may have considerable long-term effects on biology. Because the properties that arise in 'complex' systems are not those of the isolated molecule, studying such molecular properties may be of only limited value. Reducing a biological system down to the behaviour of individual molecules may serve little purpose to further our understanding in the case where the phenomena of interest arise from the collective behaviour of a population of individuals. When 'emergent' phenomena are present in a biological system, a complete molecular description is unlikely to be of much use on its own. One might know everything about the molecules and the phenomena will still not make much sense. Unfortunately, it is likely to take many years before a scientific community forms which does not view biology uniquely through the glasses of molecular reductionism.

Self-organised chemical systems constitute one type of 'complex' system. Many other types are known. Examples in biology are the gene products network, the immune system and the neural network. Some workers have proposed that the development of the metabolism as a whole is an 'emergent' phenomenon in a 'complex' system of reacting chemicals, and that appearance of life itself may be considered as the development of an 'emergent' phenomenon in a complex system [10]. From this standpoint, the self-organising stage in the development of life might be considered as part of a sequence of events, going from the formation of galaxies to the development of complex living organisms and societies, all based on the laws of non-linear dynamics in 'complex' systems.

References

[1] P. Coveney and R. Highfield. *Frontiers of Complexity.* Random House, New York, 1995.

[2] G. Nicolis and I. Prigogine. *Exploring Complexity.* Freeman, New York, 1989.

[3] S.Camazine and al. *Self-organisation in biological systems.* Princeton University Press, 2001.

[4] J. K. Parrish and L.Edelstein-Keshet. *Science,* 284: p99, 1999.

[5] B. Hölldobler and E. O.Wilson. *The Ants.* Springer-Verlag, Berlin, 1991.

[6] S.Goss, S. Aron, J. L. Deneubourg and J. M. Pasteels. *Naturwissenschaften,* 76: p579,1989.

[7] S. C. Nicolis and J. L.Deneubourg. *J. theor. Biol*. 198: p575, 1999.

[8] Kolmogorov, I. Petrovsky and N. Piskunov. Bull. *Univ Moskou Ser. Int Ser,* A16: p1, 1937.

[9] N. Rashevsky. *Bull. Math. Biophys.* 2: p15, 1940.

[10] A. M. Turing. *Phil. Trans. Roy. Soc*. 237: p37, 1952.

[11] P. Glansdorff and I. Prigogine. *Physica* 20: p773, 1954.

[12] I. Prigogine, R. Lefever, A. Goldbeter and M. Herschkowitz-Kaufman. *Nature,* 223: p913, 1969.

[13] P. Glandsdorff and I. Prigogine. *Thermodynamic theory of structure, stability and fluctuations.* Wiley, New-York, 1971.

[14] G. Nicolis and I. Prigogine. *Self-organisation in non-equilibrium system.* Wiley, New York, 1977.

[15] I. Prigogine and I. Stengers. *Order out of chaos.* Heinemann, London, 1984.

[16] A. Babloyantz. *Molecules, dynamics and life.* Wiley, New York, 1986.

[17] P. Coveney and R. Highfield. *The arrow of time,* W. H. Allen, London, 1990.

[18] J. M. T. Thompson. *Instabilities and catastrophes in science and engineering.* Wiley, New York, 1982.

[19] D. Kondepudi and I. Prigogine. *Physica,* 107A: p1, 1981.

[20] D. Kondepudi. *Physica,* 115A: p552, 1982.

[21] V. Castets, E. Dulos, J. Boissonade and P. de Kepper. *Phys. Rev. Lett,*. 64: p2953, 1990.

[22] Q. Ouyang and H. Swinney. *Nature,* 352: p610, 1991.

[23] B.P. Belousov, 1951, reprinted in "*Oscillations and travelling waves in chemical systems*". Eds. R. Field and M. Burger. Wiley, New-York, p605-p613, 1985. B.P. Belousov. *Sb. Ref. Radiats. Med. Medgiz, Moscow*, p145-p147, 1958.

[24] A.M. Zhabotinsky. *Biofizika,* 9: p306, 1964.

[25] H. Meinhardt. *Models of biological pattern formation.* Academic Press, London, 1982.

[26] J.D. Murray. *Mathematical Biology.* Springer-Verlag, Berlin, 1989.

[27] L.G. Harrison. *Kinetic theory of living pattern.* Cambridge University Press, 1993.

[28] E. Cox, E. *Curr. Opin. Gen. Dev.,* 2: p647, 1992.

[29] J. Lechleiter, S.Girard, E. Peralta and D. Clapham. *Science,* 252: p509, 1991.

[30] G. Dupont and A. Goldbeter. *Bioessays*, 14: p485, 1992.

[31] A. Goldbeter. *Biochemical Oscillations and Cellular Rhythms.* University Press Cambridge, 1996.

[32] J. Tabony and D. Job. *Nature,* 346: p448, 1990.

[33] J. Tabony and D. Job. *Nanobiology,* 1: p131, 1992

[34] J. Tabony and D. Job. *Proc. Natl. Acad. Sci. USA*, 89: p6948, 1992.

[35] J. Tabony. *Science*, 264: p245, 1994.

[36] J. Tabony. *Nanobiology*, 4: p117, 1996.

[37] J. Tabony and C. Papaseit. *Advances in Structural Biology*, 5: p43, 1998.

[38] C. Papaseit, L. Vuillard and J. Tabony. *Biophys. Chem.*, 79: p33, 1999.

[39] J. Tabony, L. Vuillard and C. Papaseit. *Adv. Complex Systems*, 2: p221, 2000.

[40] C. Papaseit, N. Pochon and J. Tabony. *Proc. Nat'l Acad. Sci. USA*, 97: p8364, 2000.

[41] J. Tabony, N. Pochon and C. Papaseit. *Adv. Space Res*. 28: p529, 2001.

[42] N. Glade, J. Demongeot and J. Tabony. *C. R. Biologies*, 325: p283, 2002.

[43] J. Tabony, N. Glade, C. Papaseit and J. Demongeot. *J. Phys IV France,* 11 Pr 6 : p239, 2001.

[44] N. Glade and J. Tabony. *J. Phys IV France,* 11 Pr 6 : p255, 2001.

[45] J. Tabony, N. Glade, J. Demongeot and C. Papaseit. *Langmuir,* 18: p7196, 2002.

[46] N. Glade, J. Demongeot and J. Tabony. *Acta Biotheoretica,* 50: p39-p268. 2002

[47] J. Tabony, N. Glade, J. Demongeot and C. Papaseit. *Adv. Space Bio. Med.*, 8: p19, 2002.

[48] J. Tabony, N. Glade, J. Demongeot and C. Papaseit. *Rec. Res. Devel. Biophys. Chem.,* 3: p11, 2002.

[49] J. Tabony, N. Glade, C. Papaseit and J. Demongeot. *J. Grav. Phys.* 9, p245-p248, 2002.

[50] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts and J. Watson. *Molecular biology of the cell.* Garland, New York, 1983.

[51] P. Dustin. *Microtubules.* Springer, Berlin, 1984.

[52] R.E. Buxbaum, T. Dennerll, S. Weiss and S. Heidemann. *Science*, 235: p1511, 1987.

[53] Various authors. *Cell and molecular biology in space.* FASEB J., Supplement 13, 1999.

[54] A. Cogoli and F. Gmünder. *Adv. Space Bio. Med.*, 1: p183, 1991.

[55] T.G. Hammond and al. *Nature Medicine* 5: p359, 1999.

[56] M. Hughes-Fulford, M. L. Lewis. *Exp. Cell Res.* 224: p103, 1996.

[57] M. Hughes-Fulford. *Adv. Space Biol. Med* 28: p129, 2002.

[58] M.L. Lewis. *Adv. Space Biol. Med.* 28: p77, 2002.

[59] M. Cogoli-Greuter, A. Spano, L. Sciola, P. Pippia and A. Cogoli. *J. J. Aerospace Env. Me*, 35: p27, 1998.

[60] M.L. Lewis, J.L. Reynolds, L.A. Cubano, J.P. Hatton, B. Lawless and E.H. Piepmeier. FASEB J, 12: p1007, 1998.

[61] J. Vassy, S. Portet, M. Beil, G. Millot, F. Fauvel-Lafeve, A. Karniguian, G. Gasset and P. Irin. *FASEB J.,* 15: p1104, 2001.

[62] B.M. Uva and al. *Brain Research,* 934: p132, 2002.

[63] S. Gaboyard and al. *Neuroreport,* 13: p2139, 2002.

[64] T. Morgan. *Anat. Anz.,* 25: p94, 1904.

[65] P. Ancel and P. Vintemberger. *Bull. Biol. Fr. Belg. (suppl.)*, 31: p1, 1948.

[66] J. Cook. *Nature,* 319: p60, 1986.

[67] N. Zisckind and R. Elinson. *Develop. Growth & Differ.*, 32: p575, 1990.

[68] B. Alberts and V. Foe. *J. Cell Science,* 61: p31, 1983.

[69] G. Calliani. *Development,* 107: p35, 1989.

[70] D. St. Johnson and C. Nusslein-Volhard. *Cell,* 68: p201, 1992.

Biologically-Inspired Cellular Computing Machines

Christof Teuscher[1,2]

[1]University of California San Diego (UCSD), Department of Cognitive Science, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA.
[2]Swiss Federal Institute of Technology Lausanne (EPFL), Logic Systems Laboratory, EPFL-IC-LSL, CH-1015 Lausanne, Switzerland. e-mail: `christof@teuscher.ch`, `http://www.teuscher.ch`

Abstract

The 21st century promises to be the century of bio- and nanotechnology. New materials and technologies such as self-assembling systems, organic and molecular electronics, hybrid electronic-biological machines, and the ever-increasing complexity and miniaturization of actual systems will require to fundamentally rethink the way we build computers, the way we organize, train, and program computers, and the way we interact with computers.

The main goal of this short contribution is to give an overview on some opponents of cellular-based biologically-inspired machines and aproaches.

1   Biologically-Inspired Computing

1.1   "Traditional" Artificial Intelligence

If we believe artificial intelligence (AI) guru Marvin Minsky, who co-founded the MIT Artificial Intelligence Laboratory in 1959 with John McCarthy, "AI has been brain-dead since the 1970s"[1]. In the same talk, Minsky also accused researchers of giving up on the immense challenge of building a fully autonomous, thinking machine. So-called "expert systems", which emulated human expertise within tightly defined subject areas like law and medicine, could match users' queries to relevant diagnoses, papers and abstracts, yet they could not learn concepts that most children know by the time they are three years old.

Marvin Minsky in not alone with his point of view: many a researcher is convinced that the quest for artificial intelligence has turned out to be a failure. Well-known philosophers such as John Searle [22] or Roger Penrose [21] argue that the intrinsic properties of the brain may not be modeled by any computer and that the human brain involves other mechanisms. On the other hand, other eminent scientists claim that human-like machine intelligence is only two steps away and that machines will soon become dominant on earth.

In traditional AI, it has been tried to specify a problem, i.e., its domain and the conditions, in the most generic and declarative way and to solve it by using a *General Problem Solver (GPS)* [5, 18]. The problem domain is usually modeled as a certain state space with operations leading from one state to another. Today, it seems obvious that the goal of a GPS — i.e., to create a single technique or engine for all problems — cannot be reached. The well-known problems of classical AI are the *Framing Problem* [7] and the *Symbol Grounding Problem* [11] (the domain knowledge has proven to be extremely important in designing efficient problem solving techniques).

However, the goal of modern AI is rather to build special purpose problem solvers that are able to incorporate as early as possible domain-specific knowledge and problem solving methods from other research communities.

It seems likely that neither the purely pessimistic nor the purely optimistic view of AI will become true. Although current AI does not offer human-like intelligence at all, machines outperform

---

[1]Mentioned during a recent speech at Boston University, see also:
`http://www.wired.com/news/technology/0,1282,58714,00.html.`

humans in many domains [13]. Machines are usually more powerful in domains where humans have difficulties (e.g., computation, etc.), but they cannot cope with many tasks that humans are able to perform almost unconsciously (e.g., pattern recognition, voice and image recognition, etc.).

It therefore seems evident to take a closer look at Nature and to find out how it solves such tasks.

1.2   Inspiration from Nature

Biologically-inspired computing can be seen as the visionary part of the general intellectual adventure of seeking for further progress in computer science. This can lead to the creation of new machines, endowed with properties usually associated with the living world: adaptation, evolution, growth and development, fault-tolerance, self-replication or cloning, reproduction, etc. The quest for novel machines and concepts is essentially motivated by the observation — as we have seen in the previous section — that fundamental progress in machine intelligence fairly often seems to stagnate. One of the keys to machine intelligence, for example, is computers that can learn, and we're still just scratching the surface of this problem.

It is evident that biological organisms operate on completely different principles from those with which most engineers are familiar. That is why most approaches do not try to faithfully model, duplicate, or copy biological systems but rather try to mimic them and to draw some inspiration from. But many biologically-inspired approaches have been labeled as failures for not having lived up to grandiose promises. "At the heart of this disappointment lies the fact that neither AI nor artificial life (Alife) has produced artefacts that could be confused with a living organism for more than an instant" [4], says Rodney Brooks. Something must be wrong! But what? He proposes different possibilities: (1) we might just be getting a few parameters wrong; (2) we might building models that are below some complexity threshold; (3) perhaps we still lack computing power; or (4) we might be missing something fundamental and currently unimagined in our models of biology.

Another possibility for new and unimaginable discoveries might be some kind of new mathematics. "We may simply not be seeing some fundamental mathematical description of what is going on in living systems and so be leaving it out of our AI and Alife models" [4]. However, none of the mathematical candidate models such as dynamical systems, chaos theories, etc., have so far revealed undiscovered fundamental descriptions.

Natural computation[2] is the study of computational systems that use ideas and get inspiration from natural systems, including biological, ecological and physical systems. It is an emerging interdisciplinary area in which a range of techniques and methods are studied for dealing with large, complex, and dynamical problems.

For example, the development of tomorrow's billion-transistor circuits and of electronic components of atomic dimensions [8, 12, 17, 37] is very likely to confront computer scientists and engineers with a challenge that John von Neumann already tried to face in the fifties: *building perfect systems out of imperfect components*. This challenge will have to be taken up by "alternative" approaches, such as by looking at Nature and by imitating some of its most interesting traits to provide architectural and organizational principles that will allow to build perfect systems from billions of partially imperfect components.

2   Biologically-Inspired Hardware

Once a machine has been imagined and "built" on paper, an algorithm has been figured out in mind, the question whether and how to test the functioning of the idea naturally arises. This might

---

[2]The terms *biologically-inspired computation* (see for example [16, 23]) or *soft-computing* [32] are also frequently used.

be done by a simulation or by building the machine. Of course, the usual way is to first simulate the system before a possible implementation is envisaged.

The question of whether to simulate computers in software or to implement them in (specialized) hardware is not at all a new one (see for example [9]). With the advent of *Field Programmable Gate Arrays (FPGAs)* [36], however, this question took somehow a back seat since it suddenly became easy and inexpensive to rapidly build (or rather *configure*) specialized hardware.

In the following, three projects shall be presented as examples of biologically inspired hardware where small and simple basic elements, i.e., the cells or molecule, are the basic building blocks. This decomposition offers several advantages and drawbacks. It is most often non-trivial to program a large set of parallel-working elements, on the other hand, its is usually much easier to implement redundant functionnality.

2.1 Example: The Embryonics Project

The *Embryonics* project (embryonic electronics) [14–16, 19, 28, 34], is a project developed in our lab by Prof. Daniel Mange and his colleagues since 1993. It is inspired by the development of multicellular organisms and by the cellular division and differentiation of living organisms and their stem cells. The final goal of this approach was to develop extremely robust integrated circuits able to self-repair and to self-replicate. The new hardware paradigm, called *autonomous reconfigurable tissue*, is based on a homogeneous, infinitely expandable tissue structured in three layers: an input, an output, and a logic layer. A molecule — the tissue's basic element — consists of a reconfigurable digital circuit. A finite set of molecules makes up a cell, a small processor with an associated memory. The cell is capable of *autonomous self-replication* and can thus create the finite set of cells making up an organism, an application-specific multiprocessor system. The final organism can itself replicate, giving rise to a population of identical organisms, i.e., clones.

The final architecture of the Embryonics project is based on four hierarchical levels of organization which, described from the bottom up, are the following (Figure 1):

- The basic primitive of our system is the *molecule*, the element of our new FPGA consisting essentially of a multiplexer associated with a programmable connection network. The multiplexer is duplicated to allow the detection of faults. The logic function of each molecule is defined by its molecular code or *MOLCODE*.

- A finite set of molecules makes up a cell, essentially a processor with an associated memory. In a first programming step of the FPGA, the *polymerase genome PG* defines the topology of the cell, that is, its width, height, and the presence and positions of columns of spare molecules. In a second step, the *ribosomic genome RG* defines the logic function of each molecule by assigning its molecular code or MOLCODE.

- A finite set of cells makes up an *organism*, an application-specific multiprocessor system. In a third and last programming step, the *operative genome OG* is copied into the memory of each cell to define the particular application, e.g., the BioWatch, executed by the organism.

- The organism can itself self-replicate, given rise to a *population* of identical organisms, the highest level of our hierarchy.

To the best of my knowledge, the Embryonics project is the only approach that implements quasi-biological growth processes in real hardware. A weak point of the Embryonics project is that there exists no methodology and no tools allowing to construct automatically artificial organisms. Thus, all organisms are coded by hand so far. Furthermore, the model is not flexible enough to allow for a direct implementation of learning and evolutionary mechanisms.
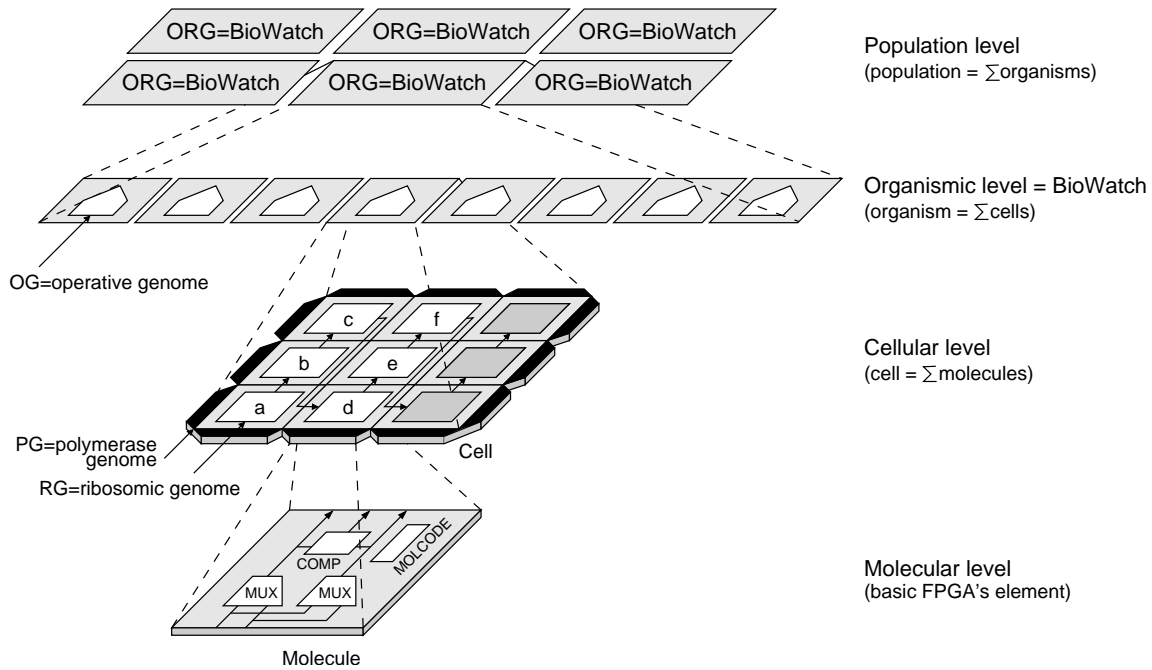
Figure 1: The Embryonics landscape of the BioWatch example: a four-level hierarchy.

2.2 Example: The BioWall

The main idea behind the construction of the BioWall[3] [26, 27, 29, 31] is the realization of Embryonic machines (see previous Section and [26, 27, 29]), however, many other applications have been implemented so far. The structure of such machines is hierarchical: organisms (application-specific systems) are realized by the parallel operation of a number of cells (small processors), and each cell is implemented as an array of molecules (programmable logic elements).

The BioWall is structured as a two-dimensional tissue composed of units (each unit corresponds to a molecule), where each unit (Figure 2) consists of an input element (a touch- sensitive membrane), an output element (an array of $8 \times 8 = 64$ two-color LEDs), and a programmable computing element (a Spartan XCS10XL Xilinx FPGA). The circuits are mounted on double boards where the logic board contains $5 \times 5 = 25$ Xilinx FPGAs, while the display board is made up of the displays and the membranes. The two boards are rigidly bound together and connected by a bus to allow two-way communication between the logic and the display (a dedicated circuit on the logic board automatically distributes the signals to and from the displays).

For a description of the major applications running on the BioWall see for example [31].

2.3 Example: The POEtic Project

The goal of the *POEtic* project[4] [30, 35] is the development of a novel digital electronic circuit, a flexible computational substrate or artificial tissue, capable of integrating the three biological models of self-organization (see [24] for a detailed description): phylogenesis (P), ontogenesis (O), and epigenesis (E). This tissue will be the essential substrate for the creation of POE-based machines, capable of evolution, growth, self-repair, self-replication, and learning. The POEtic tissue is a cellular surface composed of a variable number of elements, or cells. Each cell will contain the entire description (genome) for the whole tissue and will have the ability to communicate with the environment (through sensors and actuators) and with neighboring cells and accordingly executing

---

[3]For up-to-date information see also http://lslwww.epfl.ch/biowall.

[4]For further information see: http://www.poetictissue.org
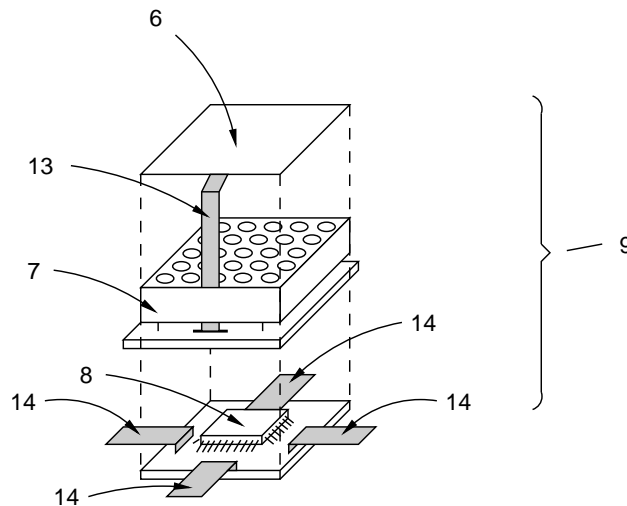
Figure 2: An illustration of the BioWall's basic building block (9). (6) touch-sensitive membrane, (7) bi-color 8×8-dot LED display, (8) Spartan XCS10XL Xilinx FPGA, (13) and (14) connections.

a function. During the project, the tissue will be validated on different kinds of applications that can benefit from its life-like properties, interacting with and adapting to their own environment.

2.4 Example: Amorphous Membrane Blending

The von Neumann architecture was first expressed in 1945 and has largely dominated in many variants and refinements computer science for more than half a century. Alternative architectures always occupied a marginal place only, despite a growing need for new concepts and paradigms in computer science. Most "classical" biologically-inspired approaches are based on on well established theories such as artificial neural networks, evolutionary algorithms, and cellular automata.

Amorphous Membrane Blending (as first proposed in [33]) takes an alternative path and is mainly motivated by the insight that tomorrow's computational substrates and environments might be very different from what we know today. Some of tomorrow's computers might be embedded in the paint that covers your desk or printed on a sheet of paper by means of a special ink. Most of such pervasive computing concepts have some common elements:

1. the computer's basic elements are very simple, identical, and available in a huge number,

2. the interactions between the elements are purely local,

3. the elements as well as the interconnections are unreliable, and

4. there is no global control mechanism available.

As its name suggest, Amorphous Membrane Blending is principally based on the unification of the following three domains of research:

1. amorphous computing [1],

2. membrane systems [20], and

3. blending [6].

An *Amorphous Computer* is a massively parallel machine made up of myriads of simple, unreliable, and identical elements, distributed randomly on a surface and interconnected locally by
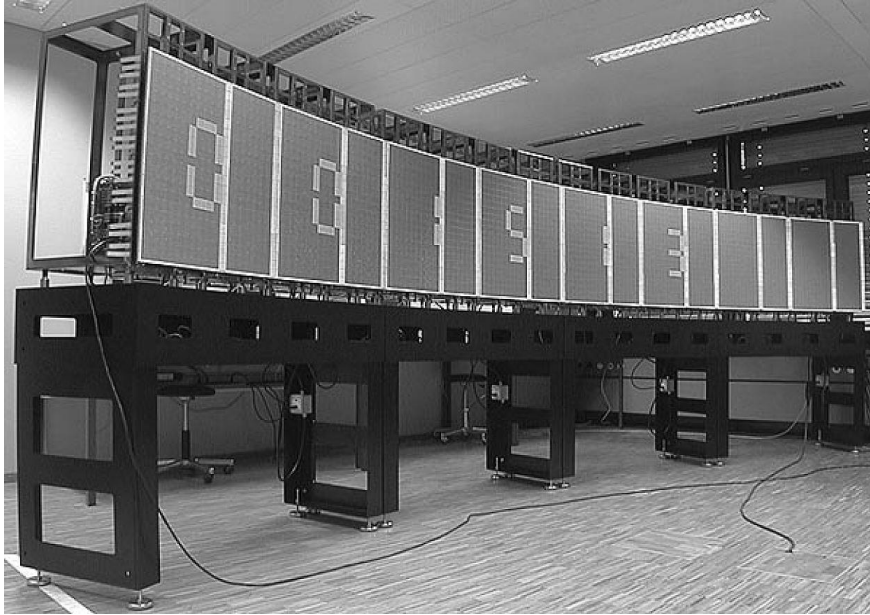
Figure 3: The BioWatch implemented on the real BioWall. Photo: André Badertscher.

unreliable connections. In a 1996 white paper, Abelson *et al.* first described the philosophy of Amorphous Computing[5] [1].

Membrane computing [20] is a paradigm that tries to imitate the way Nature "computes" at the cellular level. It makes use of a hierarchical membrane structure of cells that is similar to the structure used in the chemical abstract machine as proposed by Berry and Boudol [3]. The evolution rules within each cell that transform the multisets are basically inspired by the Γ systems proposed by Banâtre [2].

Finally *Blending* [6] is a framework of cognitive science which tries to explain how we deal with mental concepts and how creative thinking emerges.

From a bird's eye view the Amorphous Membrane Blending landscape can be decomposed into four principal hierarchical levels as depicted in Figure 4: (1) reactor level, (2) amorphon level, (3) cellular level, and (4) supercellular levels.

The combination of the above mentioned three domains of reasearch is mainly motivated by the following considerations:

- Membrane systems allow to build hierarchical structures and to divide complexity. They are also a mean to divide a task into several sub-problems and to provide a mode of parallelism.

- Artificial chemistries as implemented in membrane systems (by means of molecules an reaction rules) are an ideal mean to compute in uncertain environments. In addition, they have been identified as potentially very promising for the perpetual creation of novelty (see for example [10]).

- An amorphous computer is a promising paradigm for future computational substrates and technologies such as molecular electronics or printable digital circuits.

- Irregular and imperfect cellular machines such as amorphous computers might play a crucial role for new computational substrates and environments. It is also often hypothesized that nanotechnology will need approaches which are able to compute on the basis of an irregular and imperfect medium.

---

[5]AC website: `http://www.swiss.ai.mit.edu/projects/amorphous`
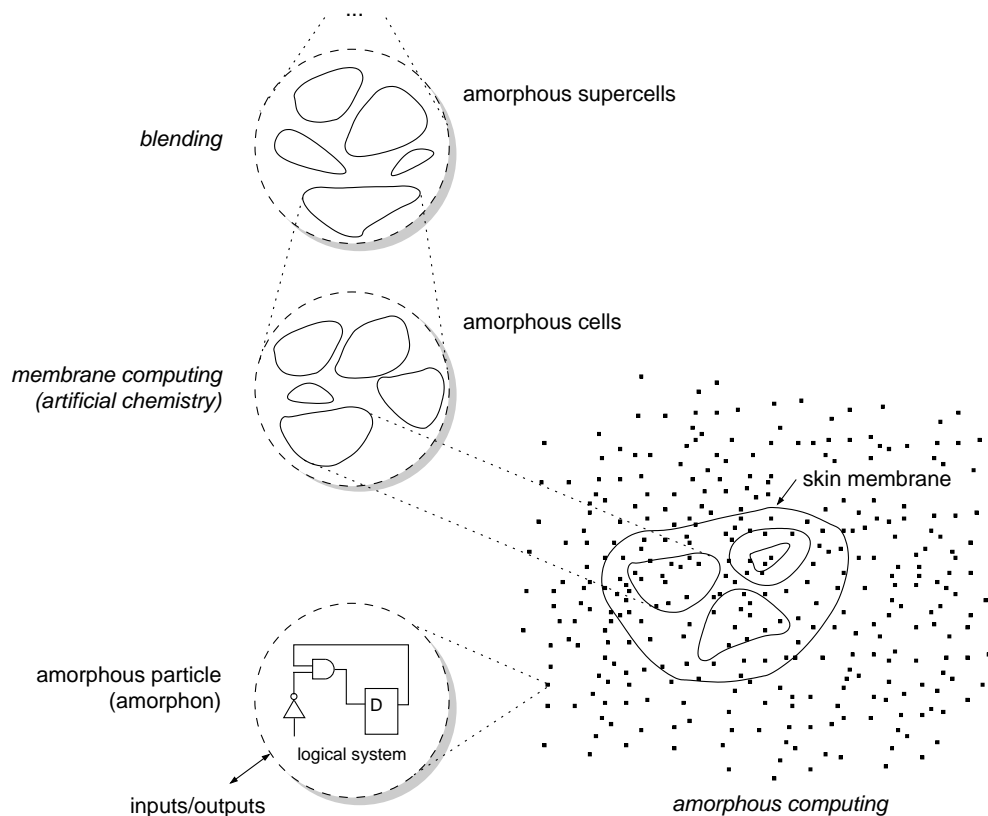
Figure 4: The Amorphous Membrane Blending landscape can be composed into four principal hierarchical levels, although cells can form hierarchies of almost any complexity.

- Blending (inspired from conceptual integration) appeared to be a potentially interesting way to create new cells from two existing cells.

Unifying membrane systems and amorphous computers would offer several interesting characteristics. One of the main reasons is that such an implementation might fully make use of the massive parallel and fine-grained architecture of the amorphous computer and the membrane system. P systems were usually simulated on a single-processor machine and their inherent parallelism could therefore not be fully exploited. Paun writes, "[...] As long as we do not have genuinely parallel hardware on which the parallelism [...] of membrane systems could be realized, what we obtain cannot be more than simulations, thus losing the main, good features of membrane systems" [20, p. 379].

Finally, putting together membrane systems on an amorphous computer together with a blending-inspired approach basically allows the system to adapt to a new environment.

References

[1] H. Abelson, D. Allen, D. Coore, C. Hanson, E. Rauch, G. J. Sussman, and R. Weiss. Amorphous computing. *Communications of the ACM*, 43(5):74–82, May 2000.

[2] J. P. Banâtre, A. Coutant, and D. Le Metayer. A parallel machine for multiset transformation and its programming style. *Future Generation of Computer Systems*, 4:133–144, 1988.

[3] G. Berry and G. Boudol. The chemical abstract machine. *Theoretical Computer Science*, 96:217–248, 1992.

[4] R. Brooks. The relationship between matter and life. *Nature*, 409:409–411, January 18 2001.

[5] G. Ernst and A. Newell. *GPS: A Case Study in Generality and Problem Solving*. Academic Press, New York, 1969.

[6] G. Fauconnier and M. Turner. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books, 2002.

[7] K. M. Ford and P. J. Hayes. *Reasoning Agents in a Dynamic World: The Frame Problem*, volume 1 of *Advances in Human and Machine Cognition*. JAI Press, Greenwich, Connecticut, 1991. Originally presented at the First International Workshop on Human & Machine Cognition, Pensacola, Florida, May 11–13, 1989.

[8] S. De Franceschi and L. Kouwenhoven. Electronics and the single atom. *Nature*, 417:701–702, June 13 2002.

[9] L. Garber and D. Sims. In pursuit of hardware-software codesign. *IEEE Computer*, 31(6):12–14, June 1998.

[10] D. Gross and McMullin B. The creation of novelty in artificial chemistries. In Standish et al. [25], pages 400–408.

[11] S. Harnad. The symbol grounding problem. *Physica D*, 42:353–346, 1990.

[12] Jenny Hogan. Quantum bits and silicon chips. *Nature*, 424(6948):484–486, July 31 2003.

[13] A. A. Hopgood. Artificial intelligence: Hype or reality? *Computer*, 36(5):24–28, May 2003.

[14] D. Mange, M. Sipper, and P. Marchal. Embryonic electronics. *Biosystems*, 51(3):145–152, September 1999.

[15] D. Mange, M. Sipper, A. Stauffer, and G. Tempesti. Toward robust integrated circuits: The embryonics approach. *Proceedings of the IEEE*, 88(4):516–540, April 2000.

[16] D. Mange and M. Tomassini, editors. *Bio-Inspired Computing Machines: Towards Novel Computational Architectures*. Presses Polytechniques et Universitaires Romandes, Lausanne, Switzerland, 1998.

[17] N. Mathur. Beyond the silicon roadmap. *Nature*, 419(6907):573–575, October 10 2002.

[18] A. Newell and H. Simon. *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, NJ, 1972.

[19] C. A. Ortega-Sánchez. *Embryonics: A Bio-Inspired Fault-Tolerant Multicellular System*. PhD thesis, The University of York, Department of Electronics, May 2000.

[20] G. Paun. *Membrane Computing*. Springer-Verlag, Berlin, Heidelberg, Germany, 2002.

[21] R. Penrose. *The Emperor's New Mind*. Oxford University Press, 1989.

[22] J. R. Searle. *The Rediscovery of the Mind*. MIT Press, Cambridge, MA, 1992.

[23] M. Sipper. *Machine Nature: The Coming Age of Bio-Inspired Computing*. McGraw-Hill, New York, 2002.

[24] M. Sipper, E. Sanchez, D. Mange, M. Tomassini, A. Pérez-Uribe, and A. Stauffer. A phyloge-netic, ontogenetic, and epigenetic view of bio-inspired hardware systems. *IEEE Transactions on Evolutionary Computation*, 1(1):83–97, April 1997.

[25] R. K. Standish, M. A. Bedau, and H. A. Abbass, editors. *Artificial Life VIII. Proceedings of the Eight International Conference on Artificial Life.* Complex Adaptive Systems Series. A Bradford Book, MIT Press, Cambridge, MA, 2003.

[26] A. Stauffer, D. Mange, G. Tempesti, and C. Teuscher. BioWatch: A giant electronic bio-inspired watch. In D. Keymeulen, A. Stoica, J. Lohn, and R. S. Zebulum, editors, *Proceedings of the Third NASA/DoD Workshop on Evolvable Hardware, EH-2001*, pages 185–192, Los Alamitos, CA, 2001. IEEE Computer Society.

[27] A. Stauffer, D. Mange, G. Tempesti, and C. Teuscher. A self-repairing and self-healing elec-tronic watch: The BioWatch. In Y. Liu, K. Tanaka, M. Iwata, T. Higuchi, and M. Yasunaga, editors, *Evolvable Systems: From Biology to Hardware. Proceedings of the 4th International Conference on Evolvable Systems, ICES2001, Tokyo, October 3-5, 2001*, volume 2210 of *Lecture Notes in Computer Science*, pages 112–127, Berlin, Heidelberg, 2001. Springer-Verlag.

[28] G. Tempesti. *A Self-Repairing Multiplexer-Based FPGA Inspired by Biological Processes*. PhD thesis, Swiss Federal Institute of Technology Lausanne, CH-1015 Lausanne, 1998. The-sis No 1827.

[29] G. Tempesti, D. Mange, A. Stauffer, and C. Teuscher. The Biowall: An electronic tissue for prototyping bio-inspired systems. In A. Stoica, J. Lohn, R. Katz, D. Keymeulen, and R. S. Zebulum, editors, *Proceedings of the 2002 NASA/DoD Conference on Evolvable Hardware*, pages 221–230, Los Alamitos, CA, 2002. IEEE Computer Society.

[30] G. Tempesti, D. Roggen, E. Sanchez, Y. Thoma, R. Canham, A. Tyrrell, and J.-M. Moreno. A POEtic architecture for bio-inspired hardware. In Standish et al. [25].

[31] G. Tempesti and C. Teuscher. Biology goes digital: An array of 5700 Spartan FPGAs brings the BioWall to "life.". *Xilinx XCell*, 47:40–45, Fall 2003.

[32] A. Tettamanzi and M. Tomassini. *Soft Computing: Integrating Evolutionary, Neural, and Fuzzy Systems*. Springer-Verlag, Berlin, Heiderlberg, 2001.

[33] C. Teuscher. *Amorphous Membrane Blending: From Regular to Irregular Cellular Comput-ing Machines*. PhD thesis, Swiss Federal Institute of Technology (EPFL), Lausanne, Switer-land, 2004. Thesis No 2925.

[34] C. Teuscher, D. Mange, A. Stauffer, and G. Tempesti. Bio-inspired computing tissues: To-wards machines that evolve, grow, and learn. *BioSystems*, 68(2–3):235–244, February–March 2003.

[35] A. Tyrrell, E. Sanchez, D. Floreano, G. Tempesti, D. Mange, J.-M. Moreno, J. Rosenberg, and Alessandro E. P. Villa. Poetic tissue: An integrated architecture for bio-inspired hardware. In A. M. Tyrrell, P. C. Haddow, and J. Torresen, editors, *Evolvable Systems: From Biology to Hardware. Proceedings of the 5th International Conference (ICES2003)*, volume 2606 of *Lecture Notes in Computer Science*, pages 129–140. Springer-Verlag, Berlin, Heidelberg, 2003.

[36] J. Villasenor and W. H. Mangione-Smith. Configurable computing. *Scientific American*, 276(6):54–59, June 1997.

[37] V. V. Zhirnov and D. J. C. Herr. New frontiers: Self-assembly in nanoelectronics. *IEEE Computer*, pages 34–43, January 2001.

Emergence in Biology

Marc H.V. Van Regenmortel

Ecole Supérieure de Biotechnologie de Strasbourg, CNRS UMR 7100, section 21, Parc d'Innovation, Boulevard Sébastien Brandt, 67400 Illkirch, France.

1   Introduction

Scientists tend to subscribe to a world view known as materialism or physicalism, according to which nothing exists but the physical. Since biological systems are solely composed of atoms and molecules, biologists when they are committed to the Unity of Science idea, tend to believe that it should be possible to reduce biology to chemistry and physics. Many neuroscientists, for instance, accept that consciousness and mental phenomena are reducible to chemical reactions occurring in the brain [4].

In recent years, an increasing number of biologists have become critical of the view that complex biological systems can be fully explained, in a reductionist manner, by the physico-chemical properties of their constituent parts. One of the reasons for this new attitude is that it is now generally accepted that biological systems can only be understood in terms of their evolutionary history on Earth. Biology is regarded as an autonomous discipline requiring its own explanatory concepts, not found in chemistry and physics and its unique character as an historical science has given rise to what philosophers call the Disunity of Science [8, 23]. According to this viewpoint, reductionism is seriously inadequate for explaining the complexity of biological systems. One of the reasons is that the vocabulary of chemistry and physics is not adapted to the type of functional language that is required for describing the role played by cellular constituents in meeting the needs of an organism and keeping it alive [12, 31].

2   The limits of reductionism

The limits of reductionism for providing adequate biological explanations have been discussed at several recent international meetings [6, 33] and special issues of scientific journals have been devoted to this question [7, 35].

It is important to stress that the value of reductionism as a research strategy for dissecting and analyzing the constituents of complex systems has never been questioned. The achievements of molecular biology during the past fifty years in unraveling the structural and chemical basis of living processes is indeed a clear testimony of the value of the reductionist approach.

The debate between reductionists and antireductionists actually centers on the nature of explanation and causality. The protagonists disagree about the validity of reductive explanations for providing an understanding of what is causally relevant in bringing about certain biological phenomena [5]. Antireductionists deny that descriptions of biomolecules and their interactions can yield an understanding of life processes and they point to the failure of genetic reductionism in explaining phenotypic and behavioural traits [18, 16, 17].

Reductionists, on the other hand, believe that the behaviour of a complex biological system can, in principle, be explained by the properties of its constituent parts and by an analysis of the multiple connections and interactions that exist between these parts.

They assume that the activity of the whole can be inferred, deduced, calculated and predicted from the properties of the parts. Antireductionists disagree and maintain that the properties of the whole cannot be deduced from the properties of the parts because interactions between parts as well as certain inputs from the environment will give rise to unpredictable novel features that are absent in the parts taken in isolation [11].

They stress that biological systems possess emergent (also called relational) properties that do not exist in the isolated components and that it is these properties that allow autonomous organisms to be directively organized in a self-regulated and integrated manner.

3   Emergence appears when reduction fails

Emergence is a concept that is complementary to reduction since emergence can be said to occur when reduction fails. Emergent properties are properties that cannot be reduced, even in principle, to properties of the constituent parts of the system and they resist any attempt at being predicted or deduced by explicit calculation or other means. For instance, the saltiness of sodium chloride is not reducible to the properties of sodium and chlorine gas and the appeal of a melody is not present in the individual notes but arises from the melody as a whole. In a similar way, life is said to emerge from non-living constituents and the mind is said to have emerged in living organisms as a matter of contingence during the evolution of the biosphere. Emergence acquired an explanatory force when it was accepted that higher level properties could possess a causal efficacy of their own and causality was no longer restricted to lower level phenomena. Mental states, for instance, are believed to be able to cause psychosomatic disease. It is also accepted that the experience of pain may alter human behaviour and that it is not necessary, in order to explain this, to invoke the causal efficacy of lower level chemical reactions occurring in neurons.

4   Levels of analysis

A biological phenomenon can be studied at different levels dependent on the analytical methods, instruments and approaches that are used. The quantum physical level is deemed to be irrelevant, while the biochemical, physiological, psychological or ecological level may be considered appropriate, depending on the question being asked. It is important to realize that descriptions at these different levels may in fact pertain to the same phenomenon. An example may make this clear. As pointed out by Rose (1998), a biochemical reaction such as the interaction between actin and myosin is not the cause of a physiological event such as muscle contraction. The biochemical process does not precede muscle contraction and it cannot, therefore, cause it: it simply describes the physiological event at the lower, chemical level. Biochemical and physiological processes occur simultaneously and the claim that the one causes the other overlooks the fact that both are descriptions of the same event (Rose 1997). According to Achinstein's classification of scientific explanations, the biochemical event is not a causal explanation of the physiological event but it represents a so-called identity explanation [1]. Instead of reducing a phenomenon described at a higher level to one at a lower level, this type of explanation simply replaces one reaction by another. Schaffner (1993) has called this type of pseudo-reduction the general reduction replacement model. Instead of reducing biology to chemistry, such explanations actually amount to a shift in subject matter whereby the attempt at describing a biological phenomenon has been abandoned. It could be argued that a considerable part of what is today described

as research in molecular biology does not really belong to biology at all since it is aimed at providing answers to chemical rather than to biological questions. When the cellular or organismic context is left out of the picture, the biological enquiry tends to become a purely chemical investigation whose relevance to biology may be questioned. The nature of an enquiry is in fact defined by the type of question that is being asked.

5   *Causal explanations versus functional explanations*

A causal explanation is reductive in the sense that one factor is singled out for attention and is given undue explanatory weight on its own. In biological systems, any observed effect always results from a complex network of interactions and an explanation in terms of a single cause is never satisfactory. Instead of invoking causes, it is more appropriate to refer to the many factors that simultaneously influence the behaviour of a biological system [33, p50]. There is also no unique causal relation between the structure and the activity of a biomolecule. A single chemical structure or protein fold can have a multiplicity of activities and a single activity can be generated by a variety of structures. It is also impossible to deduce binding activities from the structure of an interacting molecule unless a particular relationship with a specific partner has first been identified. This is because a binding site in a protein is a relational entity defined by the interacting partner and not by intrinsic structural features of the protein identifiable independently of the relational nexus with a particular ligand [31].

The appeal to causes as an explanation of events and behaviour is reinforced by the human predisposition to look for the reasons (often called causes) of human actions. However, reasons are not the causes of events and initial states of a thing are not causes of its subsequent states. Since things are not causes, DNA sequences and genes cannot be causes of phenotypic traits and of behaviour patterns [14, p39].

It is not possible to describe biological systems without referring constantly to the roles that molecules, organelles, cells and organs play in keeping an organism alive. To a human observer, organisms appear to have goals and purposes and it is tempting to try to explain their behaviour in terms of such end-directed features. However, when biologists attribute a purpose to a cellular process and explain it by referring to its goal or function, they do not imply the type of human intentionality that is used to explain purposive human action.

The occurrence of biochemical and cellular processes is explained by the fact that they contribute to the reproductive fitness of the organism in which they are found. The performance of a function must confer some good to the biological system as a whole for instance by contributing to its health, performance, survival or reproduction. Biologists tend to favour functional explanations for a currently observed structure or cellular process and they emphasize the selective advantage that these features conferred to the organism during its evolutionary history. Functions have to pass through the sieve of natural selection. Evolution is seen to operate on the DNA sequence through feedback from its gene product effects, similar to a pattern of backward causation [23].

All biological functions in a cell or organism are interdependent and internally regulated. Since their occurrence is context-dependent they cannot be understood in isolation. Functional explanations are thus more useful for understanding complex biological systems that exhibit many coupled interactions than are causal explanations that focus on a single factor.

6 Molecular, cellular and organismic functions

The term biological function is highly ambiguous. Biochemists tend to regard the function of a protein simply as what the molecule does, i.e. its functioning or activity at the molecular level. The primary activity that is considered is mostly binding activity and function is then taken to be synonymous with binding. Additional functions, however, have also been studied at the molecular level, for instance catalysis or signaling.

Functions become meaningful only when the biological context is taken into account and this implies that they must be analyzed at the level of the cell and of the organism as a whole. The simplest autonomous system that can be said to be alive is a cell and it should be remembered that during the first two billion years of terrestrial evolution, life existed only in the form of unicellular organisms. When integrated in a living cell, molecular constituents such as DNA, plasmids, proteins and macromolecular assemblies take part in the emergent self-organizing phenomenon known as life, but none of these constituents on their own can be said to be alive.

Analyzing functions at the lower molecular level without considering the cellular context amounts in fact to a purely physico-chemical analysis of little biological relevance. The catalytic activity of an enzyme, for instance, acquires a biological significance only when it is integrated in a cellular process, such as a biosynthetic pathway. The activity of an enzyme such as trypsin can be described purely at the chemical level, i.e. outside of the biological context, for instance in terms of which peptide bonds are catalysed. However, the biological function of trypsin appears only at the cellular level in the form of protein degradation or at the physiological level through its participation in the digestion process.

At the organismic level, protein functions have been classified in three broad classes corresponding to energy-, information- and communication- associated proteins [27]. However, the link between these biological roles and the structure of the many participating proteins is rather indirect and is even more difficult to ascertain than the relationship between the structure and binding activity of a single protein. Difficulties in analyzing correlations between protein structure and function have been discussed elsewhere [29, 32] and they are compounded by the fact that a single protein usually possesses several binding domains.

Functions possess a biological connotation only when they are integrated at the cellular level and are playing a role in ensuring the self-maintenance of a living organism. When biological functions are analysed at the chemical level, the ensuing reduction turns them into molecular functions and any biological role is then left out of the picture. Although the term function then has a different meaning, this is not taken into account in the existing classification schemes of protein functions which lump them all together in what has been called a mixed bag of "apples and oranges" [19, 20].

7 Emergence and complexity

Complex systems have been defined as systems that possess emergent properties and which, therefore, cannot be fully explained by the properties of their component parts. Since the constituents of a complex system interact in a non-linear manner, the behaviour of the system cannot be predicted by classical, analytical and mathematical methods that do not incorporate non-additivity and cooperative effects. Another important feature of complex biological systems is that they are open systems which exchange matter and energy with the environment and which, therefore, are not in thermodynamic equilibrium.

The recent advent of bioinformatics and powerful computers, however, has given the science of complexity the tools it needs to simulate and predict the behaviour of complex biological systems [35].

8 Modeling complexity

Computer modeling makes it possible to analyse the effect of changing one factor in a complex network of interactions and it will calculate the probable end state of the system using non-linear mathematical relations. If the behaviour predicted by modeling agrees with the observed behaviour of the system, it may seem as if some "understanding" has been achieved, even if the complexity of the system does not allow the identification of all the causal, mechanistic relations that are involved [3].

One of the Herculean tasks that lie ahead for biologists in the post-genomic era is to elucidate the functions of all the gene products in sequenced genomes. This functional analysis of proteins is usually called functional genomics although it does not really analyse the functions of genes and would be better described as functional proteomics [30].

Information on complex biological systems needs to be organized in such a way that software tools can be used to retrieve and analyse the wealth of structural and functional data that are being accumulated at an ever increasing rate in the post-genomic era. An important step in this direction was the development of so-called gene ontologies which, by using a standardized vocabulary and catalogue of terms, ensure uniform annotation and facilitate the sharing of information [2, 13].

In order to be useful for modeling the intricate networks of functional interactions that occur in different cellular compartments, databases must provide information on the localization and constitution of the various molecular assemblies such as ribosomes, proteasomes, nuclear membranes, the Golgi apparatus etc. The interactions between various cellular constituents are catalogued in terms of their role in cellular processes and pathways, for instance metabolic and biosynthetic pathways, gene regulation, transcriptional regulation, assembly and disassembly of multimolecular complexes, signal transduction, transport etc. In the aMAZE database (http://www.ebi.ac.uk/research/pfbp) the cellular constituents as well as the processes and pathways are considered to be objects in their own right [28]. Each process has its boundaries defined by the compounds that are its inputs and outputs, or alternatively by the feedback control of the end product of an earlier step in the process. Such databases that incorporate information on biological functions are a prerequisite for the computational manipulations that will be needed for modeling complex biological systems.

9 New insights

9.1 The failure of genetic reductionism

The earlier belief that Mendelian genes were responsible for unique phenotypic traits was abandoned when it was discovered that genes were made of DNA. The simple relation that had been assumed to exist between gene and trait was replaced by two relations. The first one which linked a gene to a protein became the focus of intensive study by molecular biologists and was found to be exceedingly complex [15]. The second relation which linked a gene product to a phenotypic character remained, however, clouded in mystery.

When the tools of genetic engineering became available, attempts were made to characterise all the genes involved in particular biological functions. This led to the discovery that in many cases the

molecular components involved in a complex function were molecules that had been described previously to participate in various biochemical pathways and networks. For instance, genes affecting memory formation in the fruit fly were shown to code for proteins and enzymes known to participate in the cAMP signaling pathway and to be involved in mechanisms of hormone action. Many hormones bind to membrane receptors which then induce an intracellular second messenger such as cAMP which in turn activates an intracellular target such as a protein kinase. The specificity of hormone action is thus not residing in the specificity of the molecular components that mediate its action since the same messengers are used in many different functional processes.

Although a particular gene product can be involved in various biological functions at the integrated level of the organism, at the molecular level the protein always has the same elementary activity namely binding to a specific ligand. It is the particular cellular compartment and environment in which the second messenger is released which allows the protein to have a unique effect.

Two other features explain why genetic determinism is inadequate for explaining biological functions, i.e. gene pleiotropy and gene redundancy. It is known that one genotype can lead to many different phenotypes and that one phenotype can arise from different genotypes. Gene redundancy, on the other hand, is responsible for the common finding that gene knockout experiments very often do not throw light on the role played by the gene that has been removed. The knockout frequently has no effect whatsoever despite the fact that the gene codes for a protein thought to be essential. In other cases, the knockout has an effect completely different from the one that is expected. Such results are due to the fact that the absence of one gene product is often compensated by the presence and activity of another protein [16].

Another reason why genetic reductionism is not a useful framework for explaining phenotypic traits is that genes never act in isolation. Traits always result from the activity of many different gene products. Genetic explanations of human behaviour, for instance, have led to many bogus claims, eagerly picked up by the popular press, that new genes are being discovered for alcoholism, violence, depression, rape, criminal behaviour, etc. [18, 22, 33].

Genes are studied today not only because of the role they play in the inheritance of traits from parents to offspring but also because they are involved in the genetic programme that unfolds during development and embryogenesis and because they are useful phylogenetic markers for reconstructing the process of evolution. The rise of epigenetics in recent years is a further indication that simplistic genetic reductionism has been abandoned as an acceptable explanatory strategy in biology.

9.2   The shortcomings of the somatic mutation theory of carcinogenesis

The somatic mutation theory (SMT) has been for many years the prevailing paradigm in the field of carcinogenesis. According to this theory, cancer arises when a single somatic cell becomes a neoplastic cell through the accumulation of multiple mutations in genes that control cell proliferation and the cell cycle. The SMT encouraged biologists to search for oncogenes, i.e. genes believed to produce cancer by causing excessive cell proliferation. The assumption that oncogenes constitute a particular kind of genes presumably selected for during evolution has been abandoned since cancer in fact results from the deregulation of normal, cellular processes. Although more than one hundred oncogenes have been identified, they are no longer considered to belong to a distinct gene category whose function would be to cause carcinogenesis. The effect of an oncogene depends on the type of cell in which it is expressed and overexpression of a given oncogene may enhance growth in one cell type but inhibit growth or

induce apoptosis in another [34]. In reality, the presumed association between a pattern of mutated oncogenes and cancer type has never been found.

In cancer research, established cell lines tend to be studied <u>in vitro</u> on the assumption that it is legitimate to reduce phenomena occurring at the level of tissues or whole organisms to purely cellular phenomena. Cancers were in fact reduced to transformed cells and carcinogenesis was reduced to enhanced proliferation of cells in a culture dish. This unwarranted reduction has been criticized by Sonnenschein & Soto (2000) who suggested that cancer is neither a genetic nor a cellular problem. These authors proposed a tissue organization field theory of neoplasia which postulates that carcinogens act by disrupting the normal interactions that take place between the cells of certains tissues and organs, in particular epithelium and stroma tissues. They suggested therefore that carcinogenesis should be studied at the level where it is identified, i.e. at the tissue level of biological organization rather than at the cellular or molecular levels [17].

It is now recognized that differentiation is brought about by morphogen gradients resulting from the diffusion of certain gene products away from a given source. Such gradients have been shown for instance to lead to the development of the anteroposterior and ventral-dorsal axes in Drosophila and in higher organisms. It can be said, metaphorically, that the morphogens produce positional information that allows cells to know where they are [10].

Evidence for the theory of Sonnenschein and Soto (1999, 2000) is found in the observations that liver carcinoma cells can revert to normal cells when they are injected into normal livers, whereas normal cells can be made to behave like carcinoma cells when they are repositioned. Furthermore, the neoplastic phenotype can be normalized at a frequency much higher than that found when a mutant reverts to wild type. Instead of explaining cancer by the action of mutated genes operating at the level of single cells, the cancer phenotype may thus be analysed as an emergent phenomenon of altered gene expression arising at the tissue level of biological organization.

9.3 Proteomics and drug discovery

Proteomics consists in the analysis of all the proteins expressed in a cell and involves not only their identification and quantification but also a study of their structure, localization, modifications, interactions, activities and functions [9].

It is expected that by using high throughput chips based on DNA and protein microarrays, it will be possible to identify protein targets useful in the drug discovery process. A commonly used strategy is to compare cell extracts derived from healthy and diseased organisms in order to identify which proteins are up or down regulated and therefore potentially useful as markers of the diseased state. The limited value of this approach becomes obvious when it is realized that any difference in the level of gene expression may actually be a side effect of the diseased phenotype that is not causally linked to the pathogenesis. Differences in the concentration of a protein in healthy and diseased cells may thus be of little value for identifying potential targets for drug discovery. It seems unlikely; therefore, that proteomics will make it possible to discover new drugs by a shortcut approach that would bypass the need to elucidate the complex functional networks of protein interactions that are currently under investigation.

References

[1]   P. Achinstein. *The Nature of Explanation.* Oxford University Press, New York, 1983.

[2]   M. Ashburner, and al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25: p25-p29, 2000.

[3]   J. Berger. Understanding science: why causes are not enough. *Philos. Sci.* 65:p306-p332, 1998.

[4]   J. Bickle. *Philosophy and Neuroscience: A Ruthlessly Reductive Account.* Kluwer Academic Publishing, Dordrecht, 2003.

[5]   H. Byerly. Reductionism: analysis and synthesis in biological explanations. *Quat. Rev. Biol.* 78: p336-p342, 2003.

[6]   G. Bock and J. Goode. *The Limits of Reductionism in Biology.* Novartis Foundation Symposium N° 213. Wiley, Chichester, 1998.

[7]   R. Callagher and T. Appenzeller. Beyond reductionism. *Science* 284: p79-p109, 1999.

[8]   J. Dupré. *The Disorder of Things. Metaphysical Foundation of the Disunity of Science.* Harvard University Press, Cambridge, Mass., 1993.

[9]   S. Fields. Proteomics in genomeland. *Science* 291: p1221-p1229, 2001.

[10] G. Fox-Keller. *Making Sense of Life.* Harvard University Press, Cambridge, Mass., 2002.

[11] J.H. Holland. *Emergence. Perseus Books*, Reading, 1994.

[12] D Hull and M. Ruse. *The Philosophy of Science*. Oxford University Press, 1998.

[13] N. Lan and al. Ontologies for proteomics: towards a systematic definition of structure and function that scales to the genome level. *Curr. Opin. Chem. Biol.* 7: p44-p54, 2003.

[14] M. Mahner and M. Bunge. *Foundations of Biophilosophy*. Springer, Berlin, 1997.

[15] M. Morange. *A History of Molecular Biology*. Harvard University Press, Cambridge, Mass., 1998.

[16] M. Morange. *The Misunderstood Gene*. Harvard University Press, Cambridge, Mass., 2001.

[17] L. Moss. *What Genes Cannot Do*? MIT Press, Cambridge, Mass., 2003.

[18] D. Nelkin and M.S. Lindee. *The DNA Mystique: The Gene as a Cultural Icon*. Freeman, New York, 1995.

[19] M. Riley. Systems for categorizing functions of gene products. *Curr. Opin. Struct. Biol.* 8: p388-p392, 1998.

[20] S.C.G. Rison and al. Comparison of functional annotation schemes for genomes. *Funct. Integr. Genomics* 1: p56-p69, 2000.

[21] S. Rose. *Lifelines. Biology Beyond Reductionism.* Penguin, London, 1997.

[22] S. Rose. What is wrong with reductionist explanations of behaviour. In: *The Limits of Reductionism in Biology.* Novartis Foundation Symposium N°213, Wiley, Chichester, 1998.

[23] A. Rosenberg. *Instrumental Biology or the Disunity of Science.* Chicago University Press, Chicago, 1994.

[24] K. Schaffner. *Discovery and Explanation in Biology and Medicine.* Chicago University Press, Chicago, 1993.

[25] C. Sonnenschein and A.M. Soto. *The Society of Cells. Cancer and Control of Cell Proliferation.* Springer, New York, 1999.

[26] C. Sonnenschein and A.M. Soto. The somatic mutation theory of carcinogenesis: Why it should be dropped and replaced. *Mol. Carcinogen* 29: p1-p7, 2000.

[27] J. Tamames et al. Genomes with distinct function composition. *FEBS Lett.* 389: p96-p101, 1996.

[28] J. Van Helden and al. Representing and analysing molecular and cellular function using the computer. *Biol. Chem.* 381: p921-p935, 2000.

[29] M.H.V. Van Regenmortel. Analysing structure-function relationship with biosensors. *Cell. Mol. Life Sci.* 58: p794-p800, 2001a.

[30] M.H.V. Van Regenmortel. Proteomics versus genomics. What type of structure-function relationship are we looking for? J. *Mol. Recogn.* 14: p321-p322, 2001 b

[31] M.H.V. Van Regenmortel. Reductionism and the search for structure-function relationship in antibody molecules. *J. Mol. Recogn.* 15: p240-p247, 2002 a.

[32] M.H.V. Van Regenmortel. A paradigm shift is needed in proteomics: "structure determines function" should be replaced by "binding determines function". *J. Mol. Recogn.* 15: p349-p351, 2002 b.

[33] M.H.V. Van Regenmortel and D. Hull. *Promises and Limits of Reductionism in the Biomedical Sciences.* Wiley, Chichester, 2002.

[34] I.B. Weinstein. Addiction to oncogenes. The Achilles heel of cancer. *Science* 297 : p63-p64, 2002.

[35] H. Zwirn. La complexité, Science du XXIe siècle. *Pour la Science* N°314, 2003.

Part 4

LECTURES / ABSTRACTS

Modelling self-organizing systems from the bottom up

Eric Bonabeau

Icosystem Corporation, 10 Fawcett Street, Cambridge, MA 02138, USA.
e-mail: eric@icosystem.com

Abstract

Agent-based modeling (ABM), also known as bottom-up modeling or micro-simulation, is a generic approach to modeling systems from the bottom up, starting from the system's constituent units and their interactions and working all the way up to the system's aggregate level behavior [3]. This approach is particularly suitable to modeling self-organizing systems, that is, systems that exhibit emergent, aggregate-level properties [1,4].

The characteristics of ABM include:

- Natural description: behavior of individuals rather than transition rates; describing complex individual behavior with equations is often intractable; ABM describes activities rather than processes.
- ABM captures emergent phenomena from interactions among constituent units.
- Scalability: adding more agents, tuning the complexity of agents (rules of interaction, degree of rationality, ability to learn, adapt and evolve).
- Aggregation and de-aggregation: tunable level of description: family of agents, sub-family, single agents; the right level of description depends on the model's purpose: sometimes microscopic level is necessary, sometimes an aggregate level of description is fine, sometimes a "mesoscopic" level of description must be used; different levels of description can coexist in a given model, with some parts of the model being detailed and others being higher level.
- When averages won't do the job: differential equations tend to smooth out fluctuations, not agent-based modeling, which is important as under certain conditions, fluctuations can be amplified –the system may be linearly stable but is unstable to larger perturbations.
- Heterogeneous interactions, network effects: flow equations usually assume global, homogeneous mixing. In a marketing campaign, advertising is a global source of information, but viral marketing on social networks can also be important.
- Nonlinear behavior at the individual level and in interactions: if-then rules, while describing nonlinearity and even discontinuity in individual behavior is difficult with differential equations.
- Memory and history, non-markovian behavior, temporal correlations.
- Time can be continuous or discrete.
- ABM is compatible with other types of modeling approaches. In most cases, classical approaches do just fine. Parts of a simulation model can be agent-based while other parts are classical flow equations.
- Validation and calibration can paradoxically be easier.
- Increased ownership by domain experts.
- Stochasticity can be easily included into ABM: sources of randomness can be easily applied to the right places (as opposed to a noise term added to an equation).

A useful illustration of ABM is pattern formation in social insects [4], particularly nest construction [5,6] and cemetery formation in ants [2,7].

For example, Theraulaz and Bonabeau [5] have introduced an agent-based model, inspired by the building behavior of wasps [4,6] which illuminates how coordination may emerge in the collective construction of wasp nests. Wasp nests range from very simple to highly organized. Nests are made of plant fibers that are chewed and cemented together with oral secretion. The resulting carton is then shaped by the wasps to build the various parts of the nest (pedicel, walls of cells, or external envelope). Sixty different types of wasp nests have been categorized, with many intermediates between extreme forms. A mature nest can have from a few cells up to a million cells packed in stacked combs, the latter being generally built by highly social species. Modularity is another widespread feature of nest architectures: a basic structure is repeated.

Previous studies showed that individual building algorithms consist of a series of if-then decision loops. The first act usually consists, with only a few exceptions, of attaching the future nest to a substrate with a stalk-like pedicel of wood pulp. Then, wasps initiate the nest by building two cells on either side of a flat extension of the pedicel. Subsequent cells are added to the outer circumference of the combs between two previously constructed cells. As more cells are added to the evolving structure, they eventually form closely packed parallel rows of cells and the nest generally has radial or bilateral symmetry around these initial cells. A wasp tends to finish a row of cells before initiating a new row, and rows are initiated by the construction of a centrally located cell first.

The number of potential sites where a new cell can be added increases significantly as construction proceeds: Several building actions can in principle be made in parallel—and this is certainly an important step in the emergence of complex architectures in the course of evolution. Parallelism, however, could deorganize the building process by introducing the possibility of conflicting actions being performed simultaneously. But the architecture seems to provide enough constraints to canalize the building activity. It can be seen, with a careful study of the dynamics of local configurations of cells during nest development in the primitively eusocial wasp *Polistes dominulus*, that there are not equal numbers of sites with one, two, or three adjacent walls. The great majority of them are composed of sites with two adjacent walls. Cells are not added randomly to the existing structure: Wasps have a greater probability to add new cells to a corner area where three adjacent walls are present than to initiate a new row by adding a cell on the side of an existing row. Therefore, obviously, wasps are influenced by previous construction, and building decisions seem to be made locally on the basis of perceived configurations in a way that possibly constrains the building dynamics.

Theraulaz and Bonabeau [5] introduced a class of agent-based algorithms that could hardly be made simpler: asynchronous automata that move in a three-dimensional discrete space and take actions on a pure stimulus-response basis, relying on information that is local in space and time. The deposit of an elementary building block (hereafter called a brick) by an agent depends on the local configuration of bricks in the cells surrounding the cell occupied by the agent. Two types of bricks can be deposited. No brick can be removed once it has been deposited. All simulations start with a single initial brick. A micro-rule is defined as the association of a stimulating configuration with a brick to be deposited, and we call algorithm any collection of compatible micro-rules. Two micro-rules are not compatible if they correspond to the same stimulating configuration but lead to the deposition of different bricks. An algorithm can be characterized by its micro-rule table, basically a lookup table, comprised of all its micro-rules, that is, all stimulating configurations and associated actions. A single agent in this model is able to complete an architecture. In that respect,

building is a matter of purely individual behavior. But the individual building behavior, determined by the local configurations that trigger building actions, has to be organized in such a way that a group of agents can produce the same architecture as well. Some natural wasp species face the same problem since nest construction is generally first achieved by one female, the founder, and is then taken over by a group of newly born workers. The group of agents has to be able to build the architecture without the combined actions of the different agents interfering and possibly destroying the whole activity of the swarm. The most significant result of the agent-based model is that it is possible to produce complex shapes (Figure 1), some of them strikingly similar to those observed in nature, with extremely simple stimulus-response algorithms.
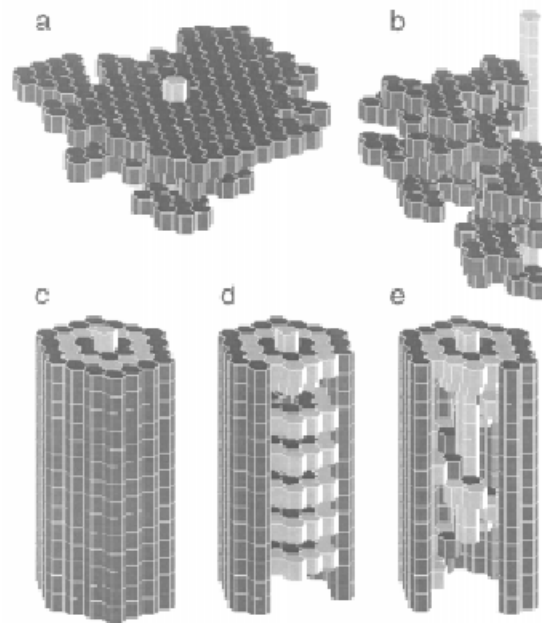
Figure 1 : Simulations of collective building on a three-dimensional hexagonal lattice. Simulations were made on a 20x20x20 lattice with 10 agents. (a) Nest-like architecture (*Vespa*) obtained after 20,000 steps. (b) Nest-like architecture (*Parachartergus*) obtained after 20,000 steps. (c,d) Nest-like architecture (*Chatergus*) obtained after 100,000 steps. A portion of the front envelope has been cut away in d. (e) Lattice architecture including an external envelope and a long-range internal helix. A portion of the front envelope has been cut away.

It was also found that structured shapes can be built only with special algorithms, coordinated algorithms, characterized by emergent coordination: Stimulating configurations corresponding to different building stages must not overlap, so as to avoid the deorganization of the building activity. This feature creates implicit handshakes and interlocks at every stage, that is, constraints upon which structures can develop consistently and robustly. This coordination is essential, since parallelism, which has a clear adaptive advantage by allowing a real colony to build at several locations at the same time, introduces the possibility of conflicts. This approach therefore shows how the nest itself can provide the constraints that canalize a stigmergic building behavior into producing specific shapes.

In another example, the formation of spatial clusters of corpses in ants [2,7], agent-based modeling also proved useful in showing how simple rules followed by workers to pick up and drop

corpses can lead to a clustering spatio-temporal dynamics that closely resembles that of laboratory experiments. An interesting extension of the agent-based model is an aggregate, density-based partial differential equation model [7] that is inspired by the agent-based model and lends itself to rigorous mathematical analysis. In other words, the agent-based model approach enabled researchers to discover the potential underlying mechanisms of cluster formation, while the top-down approach enabled a deeper mathematical understanding of the phenomenon. The two approaches are therefore not exclusive but rather complementary.

References

[1] E. Bonabeau, J.-L. Deneubourg, G. Theraulaz, A. Aron and S. Camazine, S. Self-organization in social insects. *Trends Ecol. Evol.*, 12:188-192, 1997.

[2] E. Bonabeau, G. Theraulaz, V. Fourcassié and J.-L. Deneubourg. Phase ordering kinetics of cemetery organization in ants. *Phys. Rev. E.,* 57:4568-4571, 1998.

[3] E. Bonabeau. Agent-based modeling: methods and techniques for simulating human systems. *Proc. Nat. Acad. Sci. USA,* 99:7280-7287, 2002.

[4] S. Camazine, J.-L. Deneubourg, N.R. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau. *Self-Organization in Biological Systems*, Princeton University Press, 2001.

[5] G. Theraulaz and E. Bonabeau. Coordination in distributed building. *Science* 269:686-688, 1995.

[6] G. Theraulaz, E. Bonabeau and J.-L. Deneubourg. The origin of nest complexity in social insects. *Complexity* 3:15-25, 1998.

[7] G. Theraulaz, E. Bonabeau, S.C. Nicolis, R.V. Solé, V. Fourcassié, S. Blanco, R. Fournier, J.-L. Joly, P. Fernández, A. Grimal, P. Dalle, and J.-L. Deneubourg. Spatial patterns in ant colonies. *Proc. Nat. Acad. Sci. USA,* 99:9645-9649, 2002.

Machine learning for gene networks modelling

Florence d'Alché-Buc

Maison Genopole de la Complexité, Programme de Biologie Moléculaire Intégrative, 93 rue Henri Rochefort, 91000 Evry, France. e-mail: `dalche@epigenomique.org`

Abstract

In molecular biology, functions are produced by a set of macromolecules that interact at different levels. Genes and their products, proteins, participate to regulatory networks that control the response of the cell to external or internal input signals. One of the most important challenge to biologists is undoubtedly to understand the mechanisms that govern this regulation, and to identify among a set of genes which play a regulator role and which are regulated. While the problem used to be approached by a gene to gene approach, this is changed significantly by the development of microarray technology. Expression of thousands of genes of a given organism or a given tissue can now be measured simultaneously on the same chip.

This revolution opens a large avenue for research on reconstruction of gene regulatory networks from experimental data. Several recent works have shown that this problem can be handled by a machine learning approach. However it is already difficult to make a difference among the various models and methods that have been proposed.

In order to clarify the debate, we first present fundamentals of machine learning and show how it provides both a formal setting for different important questions and a methodology to tackle the reverse modeling problems. We then discuss of existing works with respect to the proposed methodology. It appears that there is less difference between the existing approaches that one could predict. Conversely some crucial points such as sample complexity, modularity and heterogeneity of regulation are yet unsolved. This analysis tends to show that maybe there is a need for re-inventing the problem, enriching it with new perspectives and re-formulate it. Inspiration for drawing a roadmap can then be taken from biology and machine learning itself. We propose several working directions that range from building a set of benchmark problems to the conception of a complete modeling loop where learned models can be validated and sorted using the model-checking methods to suggest biologists new experimental settings.

The Combinatorics of Membrane Interactions

Vincent Danos

P.P.S., CNRS UMR 7126, Université Paris 7, 2 place Jussieu, 75251 Paris Cedex 05

Abstract

We introduce a "projective brane calculus" refining Cardelli's brane-calulus and meant to represent membrane interactions and the investigation of abstract scenarios for various viral invasion modes.

The refinement consists in introducing directed actions and brings the calculus closer to actual biological membranes, and also, and perhaps more importantly, obtains a pleasing invariance property of membrane interactions.

Elementary membrane interactions become stable under symmetries generated by shifting one's point of view. An associated structural congruence, termed the projective equivalence, is defined and shown to be preserved by sequences of interactions.

Technological developments for genetic network inference

Xavier Gidrol

Service de Génomique Fonctionnelle CEA, Genopole d'Evry, 92057 Evry cedex, France.

Abstract

To comprehend biology as a system, we must examine the structure and dynamics of cell constituents as modules rather than isolated part. Many progress in technological devices, analytical methods and biological models are required to decipher molecular networks and eventually analyze the cell as a system. In the mid nineties the development of DNA microarrays to measure the expression of thousands of genes simultaneously, was the first technological step toward systems biology. Clustering analysis of gene expression profiles generated by such a device provides insight into the "correlation" between genes and biological conditions. However it is yet restrictive as it does not reveal the causality of regulatory relationships. Besides it is very difficult to infer molecular networks from expression profiling only, as the only accessible information is the steady-state concentration of mRNA. This information is necessary to characterize the structure of genetic network and analyze its dynamic and functional properties, but it is not sufficient.

Modeling of molecular networks should take into account RNA and protein concentrations, phosphorylation state, *cis*-acting sequences, chromosome localization and so forth, since each molecular variable carries unique biological information. However due to limitation in accurate and highly parallel measuring technologies, these data are not routinely accessible. To gain as much predictive power as possible from modeling, one need to generate more data on molecular components of the cell.

We have developed innovative bioarrays to measure with sufficient accuracy, parallelism and throughput relevant data to infer genetic networks. We are manufacturing DNA array containing either intergenic region (yeast) or promoter region (human) to perform ChIP on chip analysis and localize for a given transcription factor, all putative binding sites onto the genome. Indeed genes located in theses regions are potential targets for the transcription factor under scrutiny. One can then return to DNA arrays to confirm hypothesis generated from ChIP on chip data. We are also developing cell microarrays to characterize, genome wide, upstream regulators for a given gene. Using this device we are able to transfect simultaneously thousands of genes cloned in expression vectors (allowing synthesis of gene products) into a cell line carrying the promotor region of the gene of interest cloned upstream of a reporter gene. All the genes products promoting synthesis of the reporter gene are potential regulators of the gene under investigation.

Technological breakthroughs in micro and nanotechnologies to generate comprehensive and relevant data are as critical as innovation in analytical methods for deciphering genetic networks and developing system biology.

Pour une biologie moléculaire darwinienne

Pierre Sonigo

Génétique des Virus, Institut Cochin, Paris, France.

Résumé

Pour la biologie moderne, l'origine et le fonctionnement du vivant sont attribués à l'ADN. Le programme génétique a tout prévu pour notre plus grand bénéfice et en détient les clefs : embryogenèse, fonctions, organes, régulations... Pour les premiers naturalistes, les grands équilibres naturels relevaient d'une conception analogue : le programme, celui de la Création cette fois, expliquait l'origine et le fonctionnement de la Nature, conçue de la même manière pour le plus grand bénéfice de l'Homme. La biologie moderne ne commet-elle pas, à l'intérieur de l'organisme, l'erreur que les Anciens commettaient à l'extérieur [1]?

Qu'est ce qu'un gène [1] ? Selon la définition initiale, le gène est une *abstraction* représentant le caractère héréditaire. Associé aux lois de Mendel, le gène permet de faire des « calculs » sur la transmission. Ensuite, les travaux de T. Morgan montrèrent que des *locus* chromosomiques peuvent être *associés* aux caractères [2]. L'association n'implique pas la causalité, mais permet aux marqueurs génétiques, version moderne des *locus*, d'être exploitables pour le diagnostic ou l'étude des populations. La voie était ouverte pour conférer une réalité matérielle et une valeur causale au gène abstrait des mendéliens. On aboutit au *déterminant* génétique de la biologie moléculaire : un segment d'ADN, plus ou moins facile à délimiter, qui code pour la synthèse d'une protéine. Malheureusement, sauf à assimiler le caractère héréditaire à une protéine, il est impossible de confondre le gène-ADN avec le gène-caractère. Le chemin est bien long et tortueux des protéines aux caractères héréditaires et ne dépend pas seulement des facteurs génétiques (le prion en est l'illustration extrême). Les diverses définitions du gène, tout à fait opérationnelles en leur champ propre (généalogie, diagnostic, synthèse des protéines …), ne sont pas cohérentes entre elles.

Cette incohérence accroît la difficulté d'unir les concepts de la génétique à ceux de la théorie darwinienne de l'évolution, ce qu'a tenté la théorie synthétique (ou néodarwinisme). Selon cette théorie, la sélection naturelle agit sur les phénotypes issus des variations aléatoires du génotype. Le génotype est donc sélectionné *indirectement* par l'intermédiaire du phénotype associé [3]. La théorie synthétique peut être décrite par le « gène égoïste » de R. Dawkins : c'est le gène qui est l'objet de la sélection naturelle, associé à ses « extensions » phénotypiques [4]. Le problème principal de la théorie synthétique, c'est que le chemin qui mène du génotype au phénotype est parcouru avant que la sélection ne puisse opérer. Or c'est précisément ce chemin qui préoccupe la biologie « post-génomique ». Avant d'arriver au phénotype, la vie échapperait-elle à la sélection naturelle ?

Il est possible de résoudre cette difficulté en s'appuyant sur la théorie de l'évolution *à tous les niveaux.* La sélection -correspondant par exemple à des avantages de stabilité, de croissance ou de reproduction- opérant à un niveau arbitrairement choisi produit une organisation qui peut être sélectionnée aux autres niveaux arbitrairement choisis. Schématiquement, la compétition entre les réactions chimiques[2] produit une organisation -la cellule- qui entre en compétition avec les autres

---

[1] Sur les difficultés de la définition du gène, voir le dossier « Qu'est ce qu'un gène ? » paru dans *La Recherche,* n°348, p. 50-60, Décembre 2001.

[2] Plus techniquement, cela correspond aux modèles d'auto-organisation de la chimie. Voir par exemple [5].

cellules, ce qui définit l'organisme, qui entre en compétition avec les autres organismes etc. Pour illustrer une telle *fractale darwinienne*, la métaphore de l'écosystème peut être utile : l'organisme est une forêt, habitée par des animaux libres et autonomes. Dans un écosystème, il n'y a pas de programme représentant l'objectif collectif des animaux et des plantes, et garantissant la sauvegarde de l'ensemble. La structure globale émerge des interactions individuelles. On peut aussi se référer à la ruche, construite par des abeilles qui n'ont aucun plan ni vision d'ensemble et obéissent à des règles purement locales. De manière analogue, nos cellules s'auto organiseraient autour de « chaînes alimentaires » fondées sur des avantages métaboliques locaux. Le globule blanc ne serait pas dévoué à la survie de sa collectivité. Il dévorerait les microbes parce qu'ils constituent la « nourriture » la plus accessible là où il se trouve. La cellule du foie est voisine de l'intestin. Elle se spécialiserait pour exploiter les ressources qui lui arrivent du tube digestif, non pour assurer que nous aurons bien un foie. La vie est une *conjonction d'intérêts* : l'organisme bénéficie certes de ces spécialisations cellulaires et les re-sélectionne à son niveau. Mais, le finalisme insidieux du programme génétique est évacué : la logique de l'organisme n'explique pas la mise en place des organes. Conséquence troublante dans ce cadre : l'ADN n'est plus le vecteur privilégié de l'hérédité. Il n'intervient pas plus que les autres composants qu'ils soient « internes » (nucléaires, cytoplasmiques, …) ou « externes » (locaux, environnementaux, …). Les facteurs externes portent donc l'hérédité tout autant que l'ADN.

Envisageons quelques perspectives ou ouvertures expérimentales.
La thérapie génique ne serait plus conçue comme une « reprogrammation », mais comme une introduction d'espèce dans un écosystème : en fonction des avantages sélectifs, la nouvelle espèce peut s'installer de manière stable et permettre un nouvel état de l'ensemble [6]. L'espèce introduite peut aussi s'éteindre, c'est le cas le plus fréquent en thérapie génique, ou au contraire proliférer et/ou provoquer un déséquilibre mortel. En écologie, la problématique de l'introduction ou de la préservation d'espèce dans un écosystème ne se résume pas à l'efficacité des vecteurs de transfert.
Le cancer ne serait plus conçu comme une violation du traité anti-prolifération imposé par la toute-puissance génétique. Il serait un état de moindre spécialisation tissulaire répondant à une nouvelle distribution des ressources dans l'organisme. Les relations entre le niveau de spécialisation des individus et le partage des ressources disponibles sont une problématique de l'écologie évolutionniste (voir par exemple [7]). Il serait passionnant d'importer ces outils et concepts et d'étudier sous cet angle les micro-environnements des cellules normales ou cancéreuses.
Le soi immunologique ne serait plus une catégorie particulière de structures, issues d'un apprentissage. Il correspondrait à un état stationnaire de la production et de la consommation des constituants tissulaires. La logique de l'auto-immunité serait recherchée au niveau des dynamiques de consommation des cellules entre elles, plutôt que dans les anomalies de la régulation postulée des défenses de l'organisme.
La plasticité des cellules souches traduirait leur *adaptabilité* à des ressources environnementales diverses. Selon la théorie de l'évolution, le potentiel adaptatif résulte du taux de variation et de la taille de la population considérée. Inutile de chercher une propriété intrinsèque ou un marqueur fixe de cellule souche si c'est la vitesse de variation et la taille de la population qu'il faut mesurer.
Les grandes fonctions indispensables à notre existence (respiration, digestion, reproduction, etc.) résultent des interactions moléculaires et cellulaires, sans en être la cause. Les molécules et les cellules sont libres. L'individu n'est pas au centre de son monde intérieur. Pour comprendre notre existence, il faut commencer par l'oublier.

Références

[1] Jean-Jacques Kupiec et Pierre Sonigo. *Ni Dieu ni gène, pour une autre théorie de l'hérédité*. Seuil/Science ouverte, 2000.

[2] A. Pichot. *Histoire de la notion de gène*. Champs Flammarion, 1999.

[3] John Maynard Smith. *La construction du vivant : gènes, embryons et évolution*. Cassini, coll. Le Sel et le Fer, 2001.

[4] R. Dawkins. *The extended phenotype: The long reach of the gene*. Oxford University Press, 1999.

[5] A. Pacault et J.-J. Perraud. *Rythmes et formes en chimie*. Que sais-je, Presses Universitaires de France, 1997.

[6] Alain Fisher et coll. Traitement du déficit immunitaire combiné sévère lié à l'X par transfert ex vivo du gène gamma c. *Médecine/Sciences* n°5, Mai 2000.

[7] H. Seligmann. Resource partition history and evolutionary specialization of subunits in complex systems. *Biosystems,* 51(1):31-39, 1999.

The Topology of Protein Interaction Networks

Alessandro Vespignani

Laboratoire de physique théorique, Bât 210, Université Paris-Sud, 91405 Orsay cedex, France.

Abstract

In the last few years, new experimental techniques have allowed the study of Protein Interaction Networks (PIN) on a scale comparable with that of the entire proteome of the simplest eukaryotic organisms, as Yeast. The analysis of these data has led to a better understanding of the architecture of these networks that show several complex topological features. These properties and their mathematical characterization will be reviewed along with their relevance in the formulation of evolutionary models of PINs. Finally, I will hint at the possibility to predict the functionality of uncharacterized proteins by taking advantage of the topological properties of the underlying interaction patterns.

Part 5
POSTERS

The Regulatory Gene Game
Application of Game Theory to Gene Networks Analysis

Franck Delaplace & Matthieu Manceny

Laboratoire de Méthodes Informatiques, UMR CNRS 8042, Université d'Evry-Val d'Essonne, 523 Place des Terrasses de l'agora, 91025 Evry Cedex, France.

Abstract

We model genetic networks by using game theory.

The functional analysis of these networks shows that the adaptation of a biological agent corresponds to find a maximum of a payoff function while considering that the other agents also maximise their payoff functions. To find the steady states of these agents is equivalent, in game theory, to compute Nash equilibria. Pure Nash equilibria correspond to regular states whereas Mixed Nash Equilibria are representative of singular states.

Elementary regulatory circuits have been modelised and Nash Equilibria have been used to compute their steady states.

Résumé

La théorie des jeux nous a permis de modéliser des réseaux génétiques.

L'analyse fonctionnelle de ces réseaux montre que l'adaptation d'un agent biologique revient à déterminer le maximum d'une fonction de gain attribuée à cet agent tout en considérant que les autres agents maximisent également leur propre fonction de gain. Rechercher les états stables d'un agent est équivalent, en théorie de jeux, au calcul d'équilibres de Nash. Les équilibres pures de Nash correspondent à des états réguliers, alors que les équilibres mixtes de Nash représentent les états singuliers.

Les circuits élémentaires de régulation ont été modélisés et leurs états stables calculés à l'aide des équilibres de Nash.

Design of genetic networks with specified functions by evolution *in silico*

Paul François & Vincent Hakim

Laboratoire de Physique Statistique, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris, France.

Abstract

Recent studies have provided insights into the modular structure of genetic regulatory networks and emphasized the interest of quantitative functional descriptions. Here, to provide *a priori* knowledge of the structure of functional modules, we describe an evolutionary procedure *in silico* that creates small gene networks performing basic tasks. We used it to create networks functioning as bistable switches or oscillators. The obtained circuits provide a variety of functional designs, demonstrate the crucial role of post-transcriptional interactions, and highlight design principles also found in known biological networks. The procedure should prove helpful as a way to understand and create small functional modules with diverse functions as well as to analyze large networks.

Reference

[1]   P. François, V. Hakim, PNAS, 101:580-585, 2004.

Nucleocytoplasmic Oscillations of the Transcriptional Activator Msn2

Cecilia Garmendia-Torres, Fabio De Gobbi, Georges Renault & Michel Jacquet

Laboratoire Information Génétique et Développement, Institut de Génétique et Microbiologie, Université Paris-Sud, 91405 Orsay Cedex, France.

Abstract

Msn2 is a transcriptional activator that mediates a general response to stress in yeast *Saccharomyces cerevisiae* by eliciting the expression of specific sets of genes. In response to stress or nutritional limitation, Msn2 migrates from the cytoplasm to the nucleus and shuttles repetitively into and out of the nucleus with a periodicity of a few minutes. This periodic behaviour, which can be induced by various types of stress, including light emitted by the microscope, is not dependent upon protein synthesis.

Two kinds of models could account for the oscillatory shuttling associated with the stress response: either an external biochemical oscillator drives Msn2 nucleocytoplasmic transitions or oscillatory shuttling results from an autoregulatory loop involving Msn2. To produce oscillations, nuclear Msn2 should elicit, after a delay, the process that eventually leads to its exit from the nucleus. This mechanism likely involves phosphorylation-dephosphorylation. Indeed, PKA-dependent phosphorylation prevents nuclear localization by inactivating an NLS domain and activating nuclear export, but also hyperphosphorylation by (an) other kinase(s) has been shown to correlate with activation and nuclear localization. A computational model capable of producing oscillations, based on a simple implementation of the delayed activation mechanism was presented in *Jacquet et al* 2003.

Here, we present the study of different domains of Msn2 and their involvement in the oscillatory behaviour of the protein. Using GFP-tagged constructs and high-resolution time-lapse video microscopy on single cells, we have determined the smallest part of Msn2 that is able to oscillate, which contains a NLS and a NES. And, at the same time, we show that the NLS domain which has 4 consensus sites supposedly recognized by the PKA is enough to make oscillate a GFP fused to a stress-independent NES

BioΨ: a new scheme for biological process description

Pierre Mazière, Claude Granier & Franck Molina

Centre de Pharmacologie et Biotechnologie pour la Santé, CNRS UMR 5160, Faculté de Pharmacie, 15 Avenue Charles Flahault, 34093 Montpellier Cedex 5, France.

Abstract

In current databases, knowledge of biological function is mainly in a state of textual annotation. Even though some methods obey strict standards, they have limited use for bioinformatic tools apart from clustering and crosslinking. Knowledge about biological processes is much more important than the image reflected by these databases. The main reason for this difference seems to be the lack of an integration method allowing the exhaustive description of this knowledge whose nature is much more complex than other types of biological data.

In this work, we present an original scheme that revisits the fuzzy concept of biological function through the use of elementary actions. BioΨ is based on four levels of biological process description, correponding to four well-known biological observation scales: functional motifs, functional domains, molecular entities and functional modules. The first level is built on the definition of about one hundred *Basic Elements of Action*(BEA). Associated with relevant parameters, the combination of these BEA should allow us to rebuilt the complexity of biological processes on the basis of our four levels of description.

Current biological knowledge, scattered among databases and scientific litterature, could be retranscribed through a language derived from BioΨ. This language would have the double advantage of using a standardized description and providing a direct information source for biological process simulation and modelization systems.

Elementary flux mode analysis:
application to mitochondria metabolism

Sabine Pérès, Marie Aimar & Jean-Pierre Mazat

Université Bordeaux 2, Laboratoire de Physiologie Mitochondriale, 146 rue Léo Saignat, 33076 Bordeaux cedex, France.

Abstract

Elementary flux modes of a metabolic network are defined as minimal sets of enzymes that can operate at steady states with all irreversible reactions proceeding in the appropriate direction. They define a unique set of pathways, which represents a set of generating vectors of the solution space of steady states. It can be determined from the stoichiometric matrix of the network only. The determination of elementary flux modes allows to test the robustness of a network confronted to gene deletion or addition. This approach does not involve the kinetic parameters of the individual reactions but only the architecture of the network.

Mitochondria are intra cellular organelles with their own metabolic network, involving about forty main reactions. Applied to the mitochondria metabolism, the elementary flux mode analysis can express specific pathways which could be paramountly represented in particular tissues and give them their specificity.

DS² models and Drosophila patterns

Jacques-Deric Rouault[1] & Daniel Lachaise[2]

[1]laboratoire NAMC, UMR 8620, Université Paris-Sud, 91400 Orsay, France.
[2]Laboratoire PGE, UPR 9034, CNRS, 91190 Gif-sur-Yvette, France.
e-mail : jacques.rouault@ibaic.u-psud.fr

Drosophila patterns

Among the about 4000 described species of Drosophila, most of them exhibit very characteristic patterns : spots and stripes on thorax and head, colored wings, sex-combs, bristles. The patterns are specific and present a high level of stability within a species or/and a population, showing a manifest genetic origin.

At a first step, we focus on the patterns of the abdominal tergites of the Leucophenga genus because they present a high level of variability and are well documented [1].

DS² models

DS² models design Dynamical Systems evolving in a Dynamical Structure [2]. For instance, Gierer-Meinhardt models [3,5] can be implemented in a two-dimensional cell structure which develop along the time. We present here some simple DS² models representing cells in development in a two-dimensional space.

First results

This first approach presented here shows that a model integrating a simple Gierer-Meinhardt process implemented in a few cells rectangle completed by a cell multiplication is able to produce patterns reconstructing observations performed in Drosophila: a black transverse stripe on the posterior edge, and some sparse spots. The inversion mechanism (light vs. dark) produced by this model is known in insect patterns.

References

[1] G. Bächli. Exploration du parc national de l'Upemba. Fascicule 71. *Leucophenga* und *Paraleucophenga* (Diptera Brachycera), 192 pp., Bruxelles, Belgique, 1971.

[2] J.-L. Giavitto. Topological Collections, Transformations and their Application to the Modeling and the Simulation of Dynamical Systems. *RTA'03, 14th International Conference on Rewriting Techniques and Applications*, Valencia, Spain, June 9-11, 2003.

[3] A. Gierer and H. Meinhardt. A theory of biological pattern formation. *Kibernetick,* 12:30-39, 1972.

[4] H. Meinhardt. The algorithmic beauty of sea shells. The virtual laboratory. Springer, Berlin. 236 pp., 3th edition, 1995-2003.

[5] H. Meinhardt and A. Gierer. Pattern formation by local self-activation and lateral inhibition. *Bioessays*, 22 :753-760, 2000.

Positive circuits and differentiation

Christophe Soulé

IHÉS, Le Bois-Marie, 35 route de Chartres, 91440 Bures-sur-Yvette, France.

Abstract

René Thomas has conjectured that for a gene network to lead to multistationarity, hence differentiation, it must contain a positive circuit. We prove this assertion for differentiable and piece-wise linear models. We show furthermore that a variant of it is valid when stochatic variations are taken into account.

Towards Modelling and Simulation
of Natural DNA Compaction with a Synthetic System

Alain R. Thierry

Laboratoire des Défenses Antivirales et Antitumorales, UMR 5124, Université Montpellier 2, Place E. Bataillon, 34095 Montpellier Cedex 5, France.
e-mail : thierrya@univ-montp2.fr

Abstract

We examined biophysical properties of the ultrastructure of complexes made from DNA and suitable lipids (lipoplex, Lx) designed for gene transfer. We observed by cryo-electron microscopy (cryoEM) a distinct concentric ring-like pattern with striated shells when using plasmid DNA. These spherical multilamellar particles have a mean diameter of 254 nm with repetitive spacing of 7.5 nm with striation of 5.3 nm width. Small angle x-ray scattering revealed repetitive ordering of 6.9 nm, suggesting a lamellar structure containing at least 12 layers. This concentric and lamellar structure with different packing regimes also was observed by cryoEM when using linear double-stranded DNA, single-stranded DNA, RNA and oligodeoxynucleotides. Such specific supramolecular organization is the result of thermodynamic forces, which cause compaction to occur through concentric winding of DNA in a liquid crystalline phase. CryoEM examination of T4 phage DNA packed either in T4 capsides or in lipidic particles showed strikingly similar patterns. Small angle x-ray scattering suggested an hexagonal phase in Lx-T4 DNA. Our results indicate that both lamellar and hexagonal phases may coexist in Lx preparation as well as for viral particle and that transition between both phases may depend on equilibrium influenced by type and length of the DNA used.

Ultrastructure of Lx appears as a complex matter and is not yet elucidated, nevertheless they could constitute a valuable model system for studying DNA organization in biological structures. Organization of such nucleotidic supramolecular assemblies is relevant for prebiotic chemistry. Cellular life might have begun with a membrane vesicle containing just the right mixture of polymers. In light of the numerous observations made on DNA packaging in nature or by various organic or inorganic condensing agents, our data corroborate the notion that a parallel between natural and synthetic DNA compaction can be drawn.

References

[1] D. Durand, Y. Palvadeau, M. Schmutz, A. Etienne, B. Lebleu, and A.R. Thierry. Biophysical properties of lipoplex formation. Submitted to *Biochemistry*.

[2] M. Schmutz, D. Durand, Y. Palvadeau, A. Debin, A. Etienne, and A.R. Thierry. DNA packing in stable lipid complexes designed for gene transfer imitates DNA compaction in bacteriophages. *Proc. Nat. Acad. Sci. USA* 96 :12293-12298, 1999.

Integrated Modelling of Renal Function

S. Randall Thomas

INSERM U 467, CHU Necker, 75730 Paris Cedex 15, France.

Abstract

Mathematical modeling has long played a key role in research on the physiology of renal transport systems. The necessity for such a role is dictated by the complex interactions among coupled flows at all levels of kidney organization and also by the very inaccessibility of the inner kidney structures to in vivo intervention. Thus, in vivo behavior of nephron segments and blood vessels "hidden" within the kidney interior must be inferred from micropuncture samples taken at accessible surface sites, from permeability measurements performed in vitro by such techniques as microperfusion, vesicle studies, patch clamp, immunoflourescence imaging, and others, and, finally, from the known anatomical/architectural features of the kidney. Theoretical analyses have thus been crucial to quantitative formulation of working hypotheses, have been necessary for the interpretation of experimental results, and have occasionally led directly to the development of new experimental techniques (e.g., in vitro microperfusion) in order to measure crucial parameter values whose importance was only appreciated thanks to modeling studies. Such modeling studies have helped to understand, for example: the process of glomerular filtration, the mechanism of (quasi)isotonic reabsorption in the proximal tubule, the roles of several nephron segments in acid-base regulation, the mechanisms involved in tubuloglomerular feedback and autoregulation, the importance of countercurrent flows in proximal tubule fluid reabsorption and in the many "cycles and separations" relationships among medullary nephron segments, collecting ducts and medullary blood vessels (Thomas 1998,2000,2001,2003), and details of medullary microcirculation in relation to solute and water recycling.

More recently, it has also become clear that this legacy of models of renal function (mostly related to transport of solutes and water at all levels of organisation) can be exploited for functional interpretation of genome-level experiments, such as 1) knock-outs or directed mutation studies of (often disease-related) membrane transporters, channels, and receptors, 2) modifications of cell metabolism impinging on transport functions or cell signaling, 3) flourescent tagging and imaging of significant transport proteins such as aquaporins and urea transporters, which have lately brought new details concerning not only the intimate anatomy of the inner medulla but also unsuspected and surprising changes of permeability along the length of inner medullary nephron segments.

We will describe efforts begun recently, with other renal modeling groups, to develop tools for 1) the translation of new and legacy models into a common context, via a markup language based on CellML and SBML, and 2) development of a knowledge base for quantitative data from the experimental literature on kidney and related epithelia. This work is aimed at expanding the range of problems to which modeling can be applied, by rendering more accessible previous work via a dynamic web interface. As far as possible, these tools are being developed using a generic approach, with a view toward easy adaptation to fields other than kidney physiology.

Part 6

ABSTRACTS

On "Network Motifs in Natural and Artificial Transcriptional Regulatory Networks" [1]

Franck Molina

[1] W. Banzhaf and P. Dwight Kuo, Network Motifs in Natural and Artificial Transcriptional Regulatory Networks, this volume, p. 25.

Networks are everywhere in our environment. To understand them we should investigate their similarities and differences, the mechanism for their generation and the interactions of their structure and dynamics. Few steps have been made thanks to global characterization (scale free / small world topology) or local characterization (network motifs). For the speaker emergence (the appearance of qualitative new phenomena) could be addressed through a networks approach: Nature's and Man's way to stabilize emergent phenomena is to anchor them in networks (Regulatory networks of development, Traffic Networks, Communication Networks). Thus, emergence which requires the interaction of entities (node=entities and edges=interactions in the network) could be considered as new entities at higher level.

One can consider networks as constructive systems given that new components may appear in a network and new interactions may be formed. However, networks are difficult to understand because the causes are dissolved meanwhile effect are accumulated. For the author, although networks are non-linear devices they remain our best bet to catch emergent phenomena.

Actually, networks offer a smooth way to grow complexity not only because new nodes or new interaction between existing nodes could be added, but as well it is possible to increase density of connections between nodes or to decrease density of connections between clusters of nodes. However, add or remove nodes or interactions is dangerous for a network behavior ! There is an asymmetry in the increase/decrease connection strength : if you want to change the behavior of a running system, don't remove connections or node (adding seems safer).

Relatively of networks studies so far, the main focus was on static appearance. Nowadays, the dynamics on networks and their relationships with their structure are of main interest. From the observation, dynamics of the system changes if structure has change.

In a second part the speaker presents a method to compare natural regulatory networks (E. Coli, S. Cervisiae) and artificial regulatory network. The artificial regulatory networks (32 genes) is generated by a duplication/mutation approach. The author extracts structural elements of the networks (86 elementary 3-nodes sub-graphs) which represent basic elements of complex networks. Very few motifs occur with significantly higher probability than in random networks. In addition, network motifs have been shown to be conserved over evolutionary time.

Assistance pointed out the fact that the way the network motifs were identified could be questionable for some aspects especially in comparison to existing results from former analyses of E. Coli and S. Cervisiae. The author hopes to use this approach to address the problem of reconstruction of bionetworks based on these few elementary network motifs.

The author was questioned on overall behavior observed on dynamics of networks; it seems that oscillations are a very frequent observed phenomena.

On "The Biochemical Abstract Machine BIOCHAM" [1]

Jean-Louis Giavitto

[1] N. Chabrier-Rivier, F. Fages, S. Soliman, The Biochemical Abstract Machine BIOCHAM, this volume, p.35.

This presentation is about the BioCham project, presented by François Fages. BioCham is a specification language using rules to give a precise semantics to the biological network. BioCham (Biochemical abstract machine) appears concretely as an environment integrating a rule based language dedicated to the specification of the molecules and their interactions, a interrogation language based on CTL (a temporal logic making it possible to simply express a large variety of biological questions about the behavior of the system described) and an interface to NuSMV, a tool for model-checking making it possible to answer CTL requests.

The syntax of BioCham makes it possible to define very simply molecules, proteins, their complexation and to annotate them with their functional field (or their site of connection). This syntax also corresponds to algebraic operators. For example, the operator "-" represent a connection between two species: e.g., Cdk1-CycB represents the connection of a kinase with a cyclin.

The definition of various reactions, transformations, associations/dissociations, complexations, etc, are represented by rules inspired by transition systems (an approach largely used to formalize the activity of a set of parallel processes in concurrent and distributed systems). In this approach, a rule indicates the transformation of its left part into its right part. For example, A + B = > A-b defines the complexation/liaison/... of A and B in the A-b complex.

The proposed syntax was validated by the development of large examples (a library of models is available on the web site of the project): the signaling MAPK pathway and the modeling of the cellular cycle are amongst the largest example. This last example, corresponds to 147 rules (parametric rules, which makes it possible to factorize many processes: the equivalent non-parametric description has 2733 rules) which relate to 165 genes and proteins and 500 variables. The system described has 2500 states.

Once the system described, one can raise questions about the behavior of the system in the form of a CTL formula. CTL enables the formulation of atteignability questions: "being given an initial state, the system can produce eventually (will produce necessarily) the protein P?" or many questions of stability: "is this state of the system a state of balance?" (but there is no CTL formula characterizing the whole set of stable states). To answer the CTL request, it BioChamrelies upon a dedicated tool: NuSMV.

The perspectives are numerous. To take into account the kinetics of the chemical reactions require an hybrid approach mixing numerical and logical constraints. Existing work are in hand, in particular in the CLP community (constraint logic programming). Another research direction is the use of automatic learning by logical induction to infer the evolution rules from examples of the behavior of the system.

Questions:

1) The size of the some examples is very large, especially if one compares it with what is accessible to other methods. Why?

   The model-checking on several hundreds of Boolean variables is an accessible problem. On the other hand, if numerical constraints are added or the rule are weighted (a limited form to handle the kinetic of the reactions), then practical tools are still missing.

2) Several formalism attack the same problems: the formalization of biological networks. Is there a better formalism and will eventually the tools converge to a common standard?

   The existing tools can be divided into three classes; the first corresponds to description formalisms enabling for example to exchange data and models. An example in this call is the SBML proposal. The second class gathers the languages dedicated to computer simulation: V-cell, multi-agent systems, etc. The third class, to which belongs BioCham, goes beyond simulation and allows the study of the properties exhibited by a system (i.e. a property exhibited by all the possible behaviors of a system). These tools cannot reduced to simulation environment. They have some strong points and also some weaknesses that depend on the used formalism. It is rather improbable that one converges towards a single and universal tool.

3) What is the readability of this type of approach for a biologist?

   The dedicated syntax is already well accepted. The results can be visualized by some diagrams and the graphical notations used by the community is a possible target for the tools input and output.

On "Modeling the Circadian Clock in Drosophila and Mammals" [1]

Gilles Bernot and Janine Guespin

[1] J.-C. Leloup, A. Goldbeter, Modelling the Circadian Clock in Drosophila and Mammals, this volume, p.79.

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

On "Microtubule Self-organization as an Example of the Development of Order in Living Systems" [1]

Georgia Barlovatz

[1] J. Tabony, N. Glade, C. Papaseit, J. Demongeot, Microtubule Self-organization as an Example of the Development of Order in Living Systems, this volume, p.93.

James Tabony addressed the physical chemical processes underlying microtubules self-organisation *in vitro*. His experimental work is interpreted in the framework of dissipative systems, for which theoreticians such as I. Prigogine (Nobel prize) have predicted that macroscopic self-organisation can arise from a non-linear coupling of reactive processes with molecular diffusion.

J. Tabony has found that the formation *in vitro* of microtubules from tubulin, shows this type of behaviour. These preparations spontaneously self-organise by way of reaction and diffusion, and the morphology that develops depends upon the presence of a weak external factor, such as gravity or a magnetic field, at a critical bifurcation time early in the process. Thus, the presence of an external symmetry-breaking factor, such as gravity, can determine the morphology that subsequently develops. Once assembled from tubulin, microtubules grow and shrink from opposite ends by reactive processes involving the addition and loss of free tubulin. The shrinking end of a microtubule leaves behind itself a chemical trail of high tubulin concentration. Neighbouring microtubules preferentially grow into these regions, thus progressively leading to self-organisation in a manner that shows analogies with the way that ants self-organise. Numerical simulations of the reaction-diffusion process based on the chemical dynamics of a population of microtubules successfully predict the main features of the experimental behaviour. Evidence is presented that processes of this type occur *in vivo* during embryogenesis of *Drosophyla* egg. Experiments have been shown shedding light on the influence of the egg length sample on microtubule patterns : birefringent patterns formed by microtubules assembled in cylindrical 'egg' shaped sample containers of different length, in agreement with the theoretical predictions derived from nonlinear bifurcation analysis of reaction-diffusion systems.

On "Biologically-inspired Cellular Computing Machines" [1]

Jean-Louis Giavitto

[1] C. Teuscher, Biologically-inspired Cellular Computing Machines, this volume, p.113.

Christoph Teusher presented examples of computer architectures inspired by biological principles. These new hardware architectures are justified: 1) by the needs of new applications (network of sensors, autonomous robotics, cognitive science, robust system w.r.t. the failure of its components...), 2) by technological progress (nano-technology, molecular electronics, intelligent dust), 3) by new approach of computations, approaches largely inspired by biological mechanisms and their properties (network of neuron, self-assembly, chemical, aqueous and other bio-computing model, evolutionary process...).

The first architecture presented is more specifically inspired by the biological development and by tissular processes. These processes exhibit properties of self-repair (cicatrization) and robustness (the function is assured even in the event of failure a component of the system). This computer is organized according to three levels: the molecule (an autonomous module of computation), the cell (a cell gathers a set of interacting molecules) and the organ (a set of cells). The program corresponds to an artificial genome stored identically into each cell. The position of the cell in the machine, determines the part of the program (genes) which will be carried out. The robustness achieved thanks to a redundancy of the material entities and replication. When a molecule (resp. a cell) is faulty, then the genome of the faulty entity propagates to free redundant resources and the machine reconfigures itself to take into account the new topology.

A prototype was built: BioWall, a machine of 2400 molecules organized into a grid-like cellular automata. The principal application, validating the self-repair and robustness mechanisms, is the biowatch, a program which implements a wall clock on top of the cells, only by local interaction of the molecules.

The other types of architecture presented are POEtic machines, i.e. inspired by Phylogenesis, Ontogenesis and Epigenesis mechanisms. The objective is to develop an hardware architecture being able to evolve, grow, self-repair, adapt and replicate them-self. The AMB project (Amorphous Blending Membrane) proposes to use a P systeon implemented on top of an amorphous computer to support a notion of blending. The blending is a fundamental operation suggested per G. Fauconnier for knowledge representation. This notion, originated in linguistics, generalizes the concept of semantic networks. A P system increases the concept of chemical computing (i.e. multiset rewriting) by the concept of membrane enabling the localization of computations. Lastly, an amorphous machine is made up of a great quantity of autonome, unreliable and asynchronous computing element, connected in unknown, irregular, and time-varying ways.

On "Epistemology of modeling: interdiscipline or indiscipline ?" [1,2]

Vic Norris

[1] M. H.V. Van Regenmortel, Emergence in Biology, this volume, p.123
[2] P. Sonigo, For a Darwinian Molecular Biology, this volume, p.145

Pierre Sonigo took on the central tenet of neo-Darwinism that the DNA in the form of a genetic program fully commands cells and organisms and that natural selection acts directly on this program. He used the analogy of the Robot, where there is a master program, and the Forest, where there is not, to highlight two different interpretations of biological systems. He pointed out serious problems with the Robot view. For example, the assumption that selection could act just at the level of the gene is hard to reconcile with the gulf of complexity between the gene and selectable phenotypes. In a forest, selection operates at every level and he argued in favour of the Forest view where molecules are in competition with other molecules, cells with other cells, organisms with other organisms etc. In this fractal Darwinism, the key factors leading to organisation are stability – how long an individual survives – and growth – how often an individual reproduces. The white blood cell devours bacteria because they constitute its foodset and not because it follows a master plan. Cytokines like glucose molecules can be seen as nutrients or as signals depending on the context. The nature of the individual cannot be divorced from the context. In the context of a level in an ecosystem, the mutual interests of interacting individuals dominate selection at each level. The global structure of the forest emerges from the interactions between individuals obeying local rules. And in the Forest view of the cell emerging from interactions between individual constituents, DNA loses its quasi-divine status. The Forest view has powerful implications for medicine and he illustrated these with reference to autoimmunity, cancer and differentiation and gave the example of gene therapy, where the introduction of a gene into a cell can be likened to the introduction of a new species into a particular ecosystem.

A reductionist separates his world into system and environment and in his talk Marc van Regenmortel returned to the theme of the nature of an individual object. Like Sonigo, van Regenmortel insisted on the importance of the context. Genes only provide a function in a context. We cannot predict function *de novo* from sequence. We are accustomed to a causality of the form A causes B but the behaviour of a biological system must be explained not in a single cause but in hundreds of causes. Hence the paradigm of cause leads to effect is not very useful for understanding cells. Instead, the *function* of a constituent with respect to its context becomes important where function is about contributing to the survival or reproduction of a biological system. The same protein structure can be associated with many different functions so hoping to predict 4-D dynamic structure from 1-D sequence is illusory. A new paradigm is needed to take into account the myriad interactions – and myriad types of interaction – at the different levels that exist within complex biological systems. The paradigm therefore involves *non-linear* interactions, *functions* and the *emergence* of new properties of such systems. Emergent properties are not displayed by constituent parts – emergence might be said to occur when reduction fails. It is possessed only by the whole and is a relational or systemic property. Van Regenmortel gave the example of a "specific" binding site on an immunoglobulin in which up to

15 residues out of 50 hypervariable ones interact directly with an individual epitope; however, he pointed out that the same antibody can interact "specifically" with many different epitopes. One of these interactions, the binding site, has a function in a relational nexus. Indeed, a binding site is a relational entity and can be considered as an emergent feature.

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

On "Modelling Self-organizing Systems from the bottom up" [1]

Gilles Bernot and Janine Guespin

[1] E. Bonabeau, Modelling Self-organizing Systems from the bottom up, this volume, p.135.

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

On "Machine Learning for gene networks modelling " [1]

Christine Froidevaux

[1] F. D'Alché-Buc, Machine Learning for Gene Networks Modelling, this volume, p.139.

The talk of Florence d'Alché-Buc aimed at showing the benefits of a Machine Learning approach for modelling gene regulatory networks. The development of microarray technology makes it possible to compare simultaneously expression of thousand of genes of a given organism (or tissue) in two conditions and offers new valuable pieces of information to construct gene regulatory networks. F. d'Alché-Buc introduced the task of constructing this network as a search of parameters for a model of gene interactions to be learned from experimental data. Then, she addressed the fundaments of Machine Learning and started with a survey of the problems and concepts of statistical learning. The central point was the formulation of the objective of the learning task as an optimization problem. Indeed, once the hypotheses space has been built, learning amounts to choosing the best hypothesis in it. Assuming that each hypothesis can be evaluated by a cost function, we have to select a hypothesis that minimizes the cost. The main issue under discussion on this part was whether it is justified to seek an optimal solution. Not only is the task most of the time unfeasible, but it is sometimes undesirable too. The validity of the parsimony principle with respect to biological data was also debated.

In the second part of her talk, F. d'Alché-Buc introduced the graphical models for reverse modelling. She claimed that they are closely analogous to interaction networks, what is especially true for Bayesian networks. First, she presented the work by Segal *et al.* who introduce module networks, a probabilistic method for identifying regulatory modules and their condition-specific regulators from gene expression data. Then she showed how dynamic Bayesian networks used by Perrin *et al.* for gene networks inference were especially well suited to tackle the stochastic nature of gene regulation and gene expression measurement. F. d'Alché-Buc concluded her talk by a listing of a number of perspectives, the most promising being likely the elaboration of a set of benchmark problems.

On "The Combinatorics of Membrane Interactions" [1]

Vincent Schächter and François Képès

[1] V. Danos, The Combinatorics of Membrane Interactions, this volume, p.141.

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

On "Technological Developments for Genetic Network Inference" [1]

Vincent Schächter and François Képès

[1] X. Gidrol, Technological Developments for Genetic Network Inference, this volume, p.143.

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

On "Complex Networks" [1]

Franck Molina

[1] A. Vespignani, The Topology of Protein Interaction Networks, this volume, p. 149.

In his introduction the author describes networks as a system that allows its abstract/mathematical representation as a graph. He argues on the fact that we find networks of various types (physic, cybernetic, social, biology, etc.) in various situations. These networks are interconnected to form complex networks. They could form layer or have interdependencies.

Compared to a regular lattice (with $N=10^4$) where the average distance is proportional to 10^2, small word properties (for a same $N=10^4$) related to d~Ln N. In opposite, scale free properties (i.e. random graph)are related with diverging fluctuations. When random networks follow a power-law distribution for the probability that a node has k links (*P(k)*), exponential networks follow a Poisson distribution.

By opposition to a complex system, a complicated system was described as many elements assembled following a pre-definite blueprint imposed from outside. Consequently, the system shows expected properties and performs pre-defined tasks. On the other hand, complex systems which are composed of many interacting units have dynamical evolution and the capacity of self-organization. It results a non-trivial architecture, unexpected emergent properties and cooperative phenomena.

The author describe the preferential attachment mechanism: networks expand by the addition of new nodes and nodes are wired with higher probability to highly connected nodes (i.e. WWW : links to well known web-pages).

The author defends the idea that in dynamical processes studies modeling starts from the understanding of the basic mechanisms underlying the networks' growth. A complex topology is spontaneously generated in the models (opposite to ad-hoc constructions). For this purpose we need a shift of focus from static construction to dynamical evolution. The problem could be described in two ways: the direct problem from evolution rules to emerging topology, and the inverse problem from a given topology to evolution rules.

The speaker took as an example protein-protein interaction networks (PIN) obtained from the two-hybrid technique. He observed that in such networks by looking at the connectivity distribution both small world and scale free distributions were present. Then there is the need for a more sophisticated characterization with the connectivity correlation (Pastor Satorras, Vazquez &Vespignani, PRL 87, 258701 (2001)). By following the average nearest neighbor degree it was observed that low degree proteins connect to high degree proteins and highly degree proteins connect to low degree proteins (S. Maslov and K. Sneppen, *Science* 296, 210, (2002)). This suggests a start-like shape.

Using the clustering coefficient (which follows the connected peers will likely know each other) one can detect the presence of sub-networks (Ravasz eta al. PRE 2002). Analyses of PIN suggest the presence of sub-networks: small groups of networks organized in larger groups which act as the modules at the next level and so on "ad libitum".

The speaker addressed as well the problem of lethality and centrality in protein networks (H. Jeong, S. P. Mason, A-L Barabasi and Z. Oltvai *Nature*, 411, 41, (2001)). For a PIN lethality upon removal of a protein strongly correlates with its connectivity.

Again, the speaker wanted to verify if the duplication and divergent model (Duplication and Divergence …the basic mechanism beyond Evolution S. Ohno, *Evolution by Gene Duplication,*

Springer, NY, (1970)) was consistent with what we observed on PIN. His conclusion was that correlations and sub-networks arise from duplication.

Finally, the speaker proposed to use of the information provided by the PIN to acquire knowledge about the functionality of the proteins. This approach based on a scoring function taking into account maximization of unclassified interacting proteins with shared functionality and minimization of interacting classified proteins with different functionality. The global minimization of such a score function is not very easy and leads to a functional classification prediction made by considering the functions that occurred more frequently in the optimal solutions. For the speaker, this approach may serve as a global method to obtain statistical prediction of protein function from their interaction network with reliability around 80% in highly interacting proteins.

On "Epistemology of Modelling"
Round Table discussion

Vic Norris

The reductionist paradigm is that the behaviour of the whole system can be explained entirely in terms of the properties of its parts. In the extreme DNA-centric case of reductionism, a genetic program exists that fully commands cells and organisms. What are the limits and weaknesses of this kind of reductionism? And how might they be complemented? Can a new complexity approach to biological systems be developed with its own set of explanatory concepts and its own paradigm? The concepts may include those of propagating organisations and autonomous agents that are subject to selection for growth and survival; in such selection the *function* of a constituent with respect to its context becomes important. This paradigm may involve the myriad interactions – and myriad types of interaction – at the different levels that exist within complex biological systems. The paradigm therefore involves *non-linear* interactions, *epigenetics* and the *emergence* of new properties of such systems. Emergent properties cannot be predicted just from study of isolated constituents of the system. However, some understanding of emergence can be achieved via, for example, computer modelling of the effects of changing one factor in the complex intracellular network of functional interactions occurring in a cell in a particular microenvironment. To this end, databases must provide information on the nature, location and interaction of cellular constituents at different levels. It could be argued that the new field of complexity cannot mature into a proper discipline until it has its own paradigm. In the case of complex biological systems, this paradigm can only be achieved by interdisciplinary collaborations. Until then, we shall stumble undisciplined through a no man's land. We hope that the discussion will help identify some of the landmarks and landmines.

On "Structures depending of their functioning"
Round Table discussion

Patrick Amar

One of the afternoon sessions of Thursday has been animated by Michel Thellier and Patrick Amar.

Michel Thellier has introduced the notion of *functioning-dependent structures (FDS)*: dynamical structures created and conserved because of their functioning. To illustrate this concept he used the example of the self assembly of proteins in a metabolic pathway. He showed a model of a chain of three enzymes, the product of one enzyme beeing the substrate of next one. He used a partial differential equations system to model the behaviour of the pathway. He showed that the standard behaviour is obtained when the affinity constants between the enzymes are near zero. Then he demonstrate that when the affinity constants between the enzymes are big enough, assemblies appear and in specific conditions, the system can exhibit some regulatory behaviours such as sigmoidicity. He showed another kind of FDS which inhibits the process when the last product is not released: the FDS sequests the enzymes which then are not available to catalyse reactions elsewhere.

Patrick Amar presented a simulation programme, *Hsim*, built to study the dynamics, and in particular the assembly and disassembly of large numbers of molecules in a virtual cell. He described the programme as a stochastic automaton coupled to a modelling language. The language allows the definition of molecular types, the description of interaction rules between pairs of neighboring molecules of given types, and the description of an initial state of the system. The interaction rules include rules to specify the assembly and the dissociation of molecules. Then he demonstrate the simulator with the same kind of example previously used by Michel Thellier: a chain of five enzymes in a metabolic pathway. He showed how, starting with a set of separate enzymes and initial substrate diffusing in the virtual cell, self assemblies of enzymes appears, transform the initial substrate to the final product, and then dissociate when no longer needed (i.e. when all the initial substrate has been transformed). He showed then another kind of simulation, the growth of actin filaments in the virtual cell. Finally he showed that the simulation exhibits an emergent behaviour of the system: the filaments tends to align along the axe of the cell.

On "Self-organization and language evolution" [1]

Gilles Bernot and Janine Guespin

[1] L. Steels, Self-organization and Language Evolution.

Luc Steels (Sony Computer Science Laboratory, Paris & Université libre de Bruxelles) addressed the question of self-organization and language evolution. The main hypothesis is that emergent processes are at the root of language evolution. This is in contrast with Chomski's theory of the " innate structures ". Linguistic arguments in favour of the emergent nature of language evolution are rooted in the study of this evolution itself, through language comparison and assessment of actual speech situations, where communication relies on a common (implicit) knowledge of part of the situation. The relationship between words and meaning is therefore very indirect, including several layers of categorization. The implicit plays a major role, each language differing from the others as to what is implicit (hence explicit). Starting from linguistic processes that give a hint of emergence, Steels uses multi-agent systems and learning robots endowed with a formal communication protocol to study the emergence of linguistic behaviors similar to real ones. In a community of 2 to 400 agents, what instruction must be given to each agent to obtain an evolution of the communication and a consistency of the language between all the agents? This has been tested for the emergence of sounds (one sound can spread into the community) or for a real communication where a common ground of knowledge appears during the experiment. One of the open questions is : what is general enough in these models that might also apply to genetics?

On "Synchronization and Oscillations in neuronal networks" [1]

Georgia Barlovatz

[1] V. Hakim, Synchronization and oscillations in neuronal networks.

In his talk, Vincent Hakim examined the synchronisation properties of neurons networks. This work investigates how the instantaneous firing rate of a neuron can be modulated by a noisy input. The results show that the firing-rate modulation is shaped by the subthreshold resonance. For weak noise, the firing-rate modulation has a minimum near the preferred subthreshold frequency. For higher noise, such as that prevailing in vivo, the firing-rate modulation peaks near the preferred subthreshold frequency.

Using numerical simulations of conductance-based neurons and analytical calculations of one-variable nonlinear integrate-and-fire neurons, the dependence of this synchronisation on the modulated noise has been analysed.

These results were discussed in connection with intrinsic neuron property, according which the characteristics of fast sodium channel could determine the speed with which neurons respond to noisy inputs.

On "Modelling of Biochemical networks of interactions" [1]

Jean-Louis Giavitto

[1] G. Plotkin, Modelling of Biochemical networks of interactions.

The Wednesday's presentations begin with a talk from Gordon Plotkin on "biochemical CCS", a formalism derived from process algebras for the modelling of biochemical networks of interactions (metabolic pathway, genetic regulatory network, signaling, etc).

The fundamental idea is to see the biological processes like calculations and thus, as that appeared fertile in data processing, to study their naming, their composition, their modularity... under an algebraic point of view.

In this approach, the elementary processes correspond to transformations (chemical or enzymatic reactions...), various binding processes (formation and dissociation of complexes) and translocations (diffusion, transport...). These elementary processes can be formalized in several manners, for example by a transition in a Petri net or by a differential equation. This last approach is traditional, but Gordon Plotkin shows on several examples how one can model these elementary processes by a Petri net (in this case, one cannot express the kinetics of the chemical reactions).

The approach of process algebras focus especially on the construction of the networks. Several operations make possible to give an account of it. Let us quote for example the parallel composition (p|q) of two networks p and c or the abstraction c[p] (c is a process parametrized by a name p) which acts as a template that can be instanciated into several context using different names. This lead to the study the properties of these operations of construction. For example the parallel composition is associative and commutative, and the renaming can be distributed onto the parallel composition:

$$c[p|q] = c[p]|c[q].$$

These properties make possible to reformulate any network in a normal form facilitating its study (but this normal form can have an exponential size in the number of operations used to build the network). A network being built, one can precisely associate to him a semantics in term of a Petri net or a system of differential equations.

The approach outlined here can spread in several directions: first of all, several semantics can easily be associated to a network (for example, it would be interesting to develop a stochastic interpretation of the processes); it would then be necessary to take into account the state of the chemical entities (active, inactive, phosphoryled, etc.) and in the same spirit, to consider the various functional fields of proteins and to handle the complexes. A work with a great practical utility, is the development of tools to translate a SBML description into such a network. In the long term, it is necessary to extend the formalism in order to more finely take into account the space structures (localizations), to conceive more practical concept of abstraction and modularity, etc.

Questions:

1) Some ambiguities are possible in the translation of an enzymatic reaction into a Petri net?

   That seems due to the ambiguity of the initial chemical notation. To overcome it, it is enough to ask an expert what is the good interpretation and that determines without ambiguity the corresponding Petri net.

2) How to express constraints like the laws of conservation of mass?

The formalism presented here is purely syntactic: it does not force particular constraints on the built processes. These constraints can come from properties on the basic processes. The situation is the same one with for example the differential equations: the differential formalism does not force by itself any conservation law; it is a property particular of some family of equations.

3) Which is the interest of a notation which one in fine will translate for example into differential equations?

This approach which distinguishes the syntactic construction of the network from its semantics, appeared profitable in data processing. Initially, the networks to be described can be immense and a convenient notation makes it possible to describe them in a concise and precise way. That can also constitute an interchange format. Then, certain properties depend only on construction and can be studied in the notation without referring to semantics.

4) When a network is built, some sub-networks are used (not necessarily disjoins). These sub-networks can correspond to functions or modules having a biological meaning. However, it is the biologist who fixes them. Is it possible of the infer the sub-network starting from the complete network?

It is a track of interesting research, but nothing exists yet in this direction.

"Synthesis of the Short talks and posters sessions"

Patrick Amar & Marie Dutreix

The Monday afternoon, two short talks were given by P. Mazière and S. Randall Thomas about computer science and mathematical modelling. Pierre Mazière described a new integrative language, called BioΨ, developed to describe biological functions with respect to five parameters: schedule, specification, localisation, biochemical state (conformation), and kinetics. Four scales of observation were chosen, from the chemical structure of constituents to the constitution of multimolecular complex. The association of the relevant parameters gives « modules » that then can be used for the description at the next level. The language being imprecise, integrated and non-specific, it allows modelling biological systems. The application of the method was illustrated by the description of the Insulin receptor system.

The following speaker, Randall Thomas, used also medical system to illustrate the application and the usefulness of mathematical modelling for physiologists. Actually, the complexity of the interactions among coupled flows at the level of kidney organisation and the very inaccessibility of the inner structure, prevent in vivo intervention and requires theoretical analysis to quantitatively formulate working hypothesis and predict new experiments. He described tools accessible on the web that provide a hierarchical collection of models « Big Kidney » at different scales and a quantitative database « QKDB ». The database is an open source built for and with the participation of the kidney community. These tools are being developed using generic approach, with a view toward easy adaptation to other fields.

On Tuesday afternoon, three short talks were given by Cecilia Garmendia-Torres, Paul François and Jacques-Deric Rouault.

Cecilia Garmendia-Torres showed how Msn2, a transcriptional activator which is involved in stress response in *Saccharomyces cerevisiae*, migrates periodically between the cytoplasm and the nucleus. The type of stress she studied to induce this oscillatory behaviour is the light emitted by the microscope itself. She presented a model in which nuclear Msn2, after a delay, triggers the process which leads to its exit from the nucleus. She then showed how some domains of Msn2 are involved in the periodic migration of the protein. She has determined that the part of Msn2 which is able to oscillate contains a NES and a NLS, then she showed that the NLS domain is suffcent to trigger the oscillations.

Paul François is interested in modular genetic regulatory networks. He describes an evolutionary procedure *in silico* that creates small gene networks performing basic tasks such as toggle switches (e.g. the lactose operon) or oscillators. The procedure he described use genetic algorithms which select the oscillating networks.

Jacques-Deric Rouault showed how DS^2 models (dynamical systems evolving in a dynamical structure) can be used to predict patterns appearing on Drosophila. He demonstrated his approch with a model using a few rectangular cells and a cell multiplication procedure on a 2D space. This model shows how patterns, such that the abdominal tergites of the Leucophenga genus can be produced.

Part 7

LIST OF ATTENDEES

LIST OF ATTENDEES
(in alphabetical order)

AMAR	Patrick	Patrick.Amar@lri.fr
BALLET	Pascal	pascal.ballet@univ-brest.fr
BANZHAF	Wolfgang	banzhaf@cs.mun.ca
BARLOVATZ-MEIMON	Georgia	barlovatz@creteil.inserm.fr
BASSANO	Vincent	vbassano@lami.univ-evry.fr
BERNOT	Gilles	bernot@lami.univ-evry.fr
BEURTON-AIMAR	Marie	Marie.Aimar@phys-mito.u-bordeaux2.fr
BLAVIER	André	andre.blavier@wanadoo.fr
BOELLE	Pierre-Yves	pierre-yves.boelle@sat.ap-hop-paris.fr
BONABEAU	Eric	eric@icosystem.com
BRIERE	Christian	briere@scsv.ups-tlse.fr
BRÜLS	Thomas	bruls@genomining.com
BRUNEL	Dominique	brunel@versailles.inra.fr
CABIN	Armelle	armelle.cabin@univ-rouen.fr
CAPPELLO	Giovanni	Giovanni.Cappello@curie.fr
CARITEY	Nicolas	nicolas.caritey@rp.lip6.fr
CARON LORMIER	Georffrey	gecaronl@univ-rennes1.fr
CHRISTOPHE	Catherine	cch@jouy.inra.fr
COMET	Jean-Paul	comet@lami.univ-evry.fr
D'ALCHE-BUC	Florence	dalche@epigenomique.org
DANOS	Vincent	Vincent.Danos@pps.jussieu.fr
DAUVILLIER	Jérôme	Jerome.Dauvillier@serono.com
DELAPLACE	Franck	delapla@lami.univ-evry.fr
DELOSME	Jean-Marc	delosme@lami.univ-evry.fr
DEMARTY	Maurice	maurice.demarty@univ-rouen.fr
DESMEULLES	Gireg	desmeulles@enib.fr
DUTREIX	Marie	marie.dutreix@curie.u-psud.fr
ENGELEN	Stéfan	sengelen@lami.univ-evry.fr
EUVRARD	Guillaume	guillaume.euvrard@wanadoo.fr
FAGES	François	francois.fages@inria.fr
FLEURY	Ludovic	lfleury@genoscope.cns.fr
FOURMENTIN	Jean	fgfondat@club-internet.fr
FOURMENTIN	Eric	fgfondat@club-internet.fr
FRANCOIS	Paul	francois@lps.ens.fr
FROIDEVAUX	Christine	chris@lri.fr
GARMENDIA-TORRES	Cécilia	cecilia.garmendia@igmors.u-psud.fr
GAUBERT	Laurent	laurent.gaubert@club-internet.fr
GENIEYS	Stéphane	genieys@maply.univ-lyon1.fr
GHERBI	Rachid	gherbi@limsi.fr
GIAVITTO	Jean-Louis	giavitto@lami.univ-evry.fr
GIDROL	Xavier	GIDROL@cea.fr
GOODWIN	Brian	BCGood1401@aol.com
GRANGE	Thierry	grange@ccr.jussieu.fr
GRONDIN	Yohann	jg69@le.ac.uk
GUESPIN-MICHEL	Janine	Janine.Guespin@univ-rouen.fr
HAKIM	Vincent	hakim@tournesol.lps.ens.fr
HUTZLER	Guillaume	hutzler@lami.univ-evry.fr
IARTSEVA	Anastasia	iartseva@mail.ru

INCITTI	Roberto	incitti@ihes.fr
JACQUET	Michel	michel.jacquet@igmors.u-psud.fr
KEPES	François	Francois.Kepes@genopole.cnrs.fr
KLAUDEL	Hanna	klaudel@lami.univ-evry.fr
KUTTLER	Céline	celine.kuttler@iemn.univ-lille1.fr
LE FEVRE	François	francois.lefevre@genoscope.cns.fr
LECLERCQ	Sébastien	leclercq@genopole.cnrs.fr
LELOUP	Jean-Christophe	jleloup@ulb.ac.be
LESNE	Annick	lesne@lptl.jussieu.fr
MALO	Michel	michel.malo@creteil.inserm.fr
MANCENY	Matthieu	mmanceny@lami.univ-evry.fr
MAZIERE	Pierre	pierre.maziere@ibph.pharma.univ-montp1.fr
MEINHARDT	Hans	hans.meinhardt@tuebingen.mpg.de
MESTIVIER	Denis	mestiv@ebgm.jussieu.fr
MICHAUT	Angélique	angelique.michaut@wanadoo.fr
MICHEL	Olivier	michel@lami.univ-evry.fr
MOLINA	Franck	franck.molina@ibph.pharma.univ-montp1.fr
NATOWICZ	René	r.natowicz@esiee.fr
NORRIS	Vic	victor.norris@univ-rouen.fr
PALDI	Andras	paldi@genethon.fr
PARISEY	Nicolas	nicolas.parisey@wanadoo.fr
PELTEKIAN	Elise	peltek@genethon.fr
PERES	Sabine	sabine.peres@etud.u-bordeaux2.fr
PETROVA	Alexandra	pealvy@yahoo.com
PEYRIERAS	Nadine	Nadine.Peyrieras@biologie.ens.fr
PINAUD	Benjamin	bpinaud@genoscope.cns.fr
PLOTKIN	Gordon	gdp@inf.ed.ac.uk
QUACH	Minh	quach@epigenomique.org
QUERREC	Gabriel	gabriel.querrec@enib.fr
REYMOND	Nancie	nreymond@jouy.inra.fr
RICHARD	Adrien	arichard@lami.univ-evry.fr
RIPOLL	Camille	camille.ripoll@univ-rouen.fr
ROUAULT	Jacques-Deric	jacques.rouault@ibaic.u-psud.fr
SAIDANI	Samir	saidani@info.unicaen.fr
SCHACHTER	Vincent	vs@genoscope.cns.fr
SMIDTAS	Serge	sergi@sergi5.com
SONIGO	Pierre	sonigo@cochin.inserm.fr
SOULE	Christophe	soule@ihes.fr
SPICHER	Antoine	aspicher@lami.univ-evry.fr
SRIRAM	Krishna	sriram@epigenomique.org
STEELS	Luc	steels@arti.vub.ac.be
TABONY	James	TABONY@dsvsud.cea.fr
TAHI	Fariza	tahi@lami.univ-evry.fr
TEUSCHER	Christof	christof@teuscher.ch
THELLIER	Michel	Michel.Thellier@univ-rouen.fr
THIERRY	Alain	thierrya@univ-montp2.fr
THOMAS	Randy	srthomas@necker.fr
TRACQUI	Philippe	Philippe.Tracqui@imag.fr
VANNIER	Jean-Pierre	jean-pierre.vannier@chu-rouen.fr
VAN REGENMORTEL	Marc	Vanregen@esbs.u-strasbg.fr
VESPIGNANI	Alessandro	Alessandro.Vespignani@th.u-psud.fr
ZEMIRLINE	Abdallah	Abdallah.Zemirline@univ-brest.fr