List of attendees

Patrick	AMAR	pa@lri.fr
Martyn	AMOS	M.R.Amos@exeter.ac.uk
Hugues	ASCHARD	hugoaschard@hotmail.com
Pascal	BALLET	pascal.ballet@univ-brest.fr
Georgia	BARLOVATZ	gbm@lami.univ-evry.fr
Vincent	BASSANO	vbassano@lami.univ-evry.fr
Gilles	BERNOT	bernot@epigenomique.org
Grégory	BEURIER	beurier@lirmm.fr
marie	BEURTON-AIMAR	aimar@labri.fr
Lélia	BLIN	lelia.blin@lami.univ-evry.fr
Christian	BRIERE	briere@scsv.ups-tlse.fr
Benoît	CALVEZ	bcalvez@lami.univ-evry.fr
Geoffrey	CARON-LORMIER	gecaronl@univ-rennes1.fr
André	CHARRIER	andre.charrier@ensam.inrafr
Chafika	CHETTAOUI	chetta@lami.univ-evry.fr
Olivier	COLLIN	olivier.collin@imag.fr
Jean-Paul	COMET	comet@lami.univ-evry.fr
Gilles	CURIEN	gcurien@cea.fr
Florence	D'ALCHÉ-BUC	dalche@lami.univ-evry.fr
olivier	DELANEAU	olivier.delaneau@wanadoo.fr
Franck	DELAPLACE	delapla@lami.univ-evry.fr
Gireg	DESMEULLES	desmeulles@enib.fr
Peter	DITTRICH	dittrich@minet.uni-jena.de
Raouf	DRIDI	dridi@lifl.fr
Jacques	DUMAIS	jdumais@oeb.harvard.edu
stefan	ENGELEN	sengelen@lami.univ-evry.fr
Abraham	ESCOBAR	aescobar@lusignan.inra.fr
David	FELL	dfell@brookes.ac.uk
Nicolas	FEREY	ferey@limsi.fr
Pascal	FERRARO	ferraro@labri.fr
jean-louis	GIAVITTO	giavitto@lami.univ-evry.fr
Christophe	GODIN	godin@cirad.fr
Brian	GOODWIN	BCGood1401@aol.com
Andrew	GRIFFITHS	griff@mrc-lmb.cam.ac.uk
Yohann	GRONDIN	jg69@le.ac.uk
Pierre-Emmanuel	GROS	gros@limsi.fr
Janine	GUESPIN	janine.guespin@univ-rouen.fr
François	HOULLIER	francois.houllier@cirad.fr
Guillaume	HUTZLER	hutzler@lami.univ-evry.fr
Anastassia	IARTSEVA	aiartsev@lami.univ-evry.fr
Liliana	IBANESCU	ibanescu@loria.fr
François	KÉPÈS	Francois.Kepes@genopole.cnrs.fr
Sébastien	KERDELO	kerdelo@enib.fr
Bernard	KORZENIEWSKI	benio@mol.uj.edu.pl

Charles	LALES	charleslales@hotmail.com
Marc	LAMARINE	Marc.Lamarine@serono.com
Benjamin	LEBLANC	benjamin.leblanc@cpbs.univ-montp1.fr
Mikaël	LUCAS	mlucas@ens-lyon.fr
Claire	LURIN	lurin@evry.inra.fr
Michel	MALO	mmalo@lami.univ-evry.fr
Matthieu	MANCENY	mmanceny@lami.univ-evry.fr
Pierre	MARTINEAU	pmartino@valdorel.fnclcc.fr
Jean-Pierre	MAZAT	jpm@u-bordeaux2.fr
Pierre	MAZIERE	pmaziere@ebi.ac.uk
Olivier	MICHEL	michel@lami.univ-evry.fr
Franck	MOLINA	franck.molina@cpbs.univ-montp1.fr
Thomas	MONCION	tmoncion@lami.univ-evry.fr
Violaine	MOREAU	violaine.moreau@cpbs.univ-montp1.fr
Abderrazzak	NEJEOUI	a.nejeoui@ucam.ac.ma
Vic	NORRIS	victor.norris@univ-rouen.fr
Aida	OUANGRAOUA	aida.ouangraoua@labri.fr
Nicolas	PARISEY	parisey@labri.fr
Bernard	PAU	bernard.pau@ibph.pharma.univ-montp1.fr
Sabine	PÉRÈS	sabine.peres@etud.u-bordeaux2.fr
Michel	PETITOT	petitot@lifl.fr
karine	PIOT	karine.piot@cpbs.univ-montp1.fr
Przemyslaw	PRUSINKIEWICZ	pwp@cpsc.ucalgary.ca
Gabriel	QUERREC	gquerrec@enib.fr
Nancie	REYMOND	nreymond@umpa.ens-lyon.fr
Adrien	RICHARD	arichard@lami.univ-evry.fr
Camille	RIPOLL	camille.ripoll@univ-rouen.fr
Pierre	ROUX	pierre.roux@crbm.cnrs.fr
Nicolas	SALVETAT	nicolas.salvetat@cpbs.univ-montp1.fr
Saidani	SAMIR	saidani@info.unicaen.fr
Ian	SMALL	small@evry.inra.fr
Antoine	SPICHER	aspicher@lami.univ-evry.fr
Jörg	STELLING	joerg.stelling@inf.ethz.ch
Fariza	TAHI	tahi@lami.univ-evry.fr
Alain	THIERRY	thierrya@univ-montp2.fr
Randall	THOMAS	srthomas@lami.univ-evry.fr
Nizar	TOULEIMAT	nizar@epigenomique.org
Philippe	TRACQUI	Philippe.Tracqui@imag.fr
Sylvie	TRONCALE	troncale@lami.univ-evry.fr
Sandrine	VIAL	svial@lami.univ-evry.fr
Andreas	WAGNER	aw@ihes.fr
Abdallah	ZEMIRLINE	zemirlin@univ-brest.fr

CONTENTS

Part 1

ARTICLES

Cellular Computing

Martyn Amos*

*Department of Computer Science, University of Exeter, UK

Abstract

Complex natural processes may often be expressed in terms of networks of computational components, such as Boolean logic gates or artificial neurons. The interaction of biological molecules and the flow of information controlling the development and behaviour of organisms is particularly amenable to this approach, and these models are well-established in the biological community. However, only relatively recently have papers appeared proposing the use of such systems to perform useful, *human-defined* tasks. Rather than merely using the network analogy as a convenient technique for clarifying our understanding of complex systems, it may now be possible to harness the power of such systems for the purposes of computation. In this paper we review several such proposals, focusing on the molecular implementation of fundamental computational elements.

1 Introduction

When, in the 17th century, the French mathematician and philosopher René Descartes declared to Queen Christina of Sweden that animals could be considered a class of machines, she challenged him to demonstrate how a clock could reproduce. Three centuries later, with the publication of *"The General and Logical Theory of Automata"* [15] John von Neumann showed how a machine could indeed construct a copy of itself. von Neumann believed that the behaviour of *natural* organisms, although orders of magnitude more complex, was similar to that of the most intricate machines of the day. He believed that life was based on logic.

The concept of molecular complexes forming computational components was first proposed by Richard Feynman in his famous talk *"There's Plenty of Room at the Bottom"* [10]. The idea was further developed by Bennett [4] and Conrad and Liberman [6], and since then there has been an explosion of interest in performing computations at a molecular level. In 1994, Adleman showed how a massively-parallel random search may be implemented using standard operations on strands of DNA [1].

Since Adleman's original experiment, several other attempts to implement computations using DNA have been reported [3]. However, these experiments, although different in many ways are characterized by the fact that they are all implemented *in vitro* - information is encoded as strands of DNA and these are then manipulated in solution in order to perform a computation. It may be argued that this approach is sub-optimal in that it fails to utilize the true potential of the DNA molecule in its natural environment (i.e., in the cell). The advantages of working *in vivo* as opposed to *in vitro* are numerous; rather than using DNA as a passive information carrier, we may take advantage of the fact that it can also be *meaningful* in a biological context. By reprogramming part of the cellular machinery to our advantage, the DNA program can affect its own execution by the production of proteins, for example. In [8], Eng describes how, in principle, cellular processes may be harnessed in order to perform computations. We now describe in more detail and expand on this work.

2 Physical Implementations

In 1999, Weiss *et al.* [16] described a technique for mapping digital logic circuits onto genetic regulatory networks such that the resulting chemical activity within the cell corresponds to the

computations specified by the digital circuit (this work is described in a later Chapter). There was a burst of activity in 2000, when two papers appeared in the same issue of *Nature*, both being seminal contributions to the field. In [7], Elowitz and Leibler described the construction of an oscillator network that periodically caused a culture of *E. coli* to glow by expressing a fluorescent protein. Crucially, the period of oscillation was slower than the cell division cycle, indicating that the state of the oscillator is transmitted from generation to generation. In [11], Gardner *et al.* implemented a genetic toggle switch in *E. coli*. The switch is flipped from one stable state to another by either chemical or heat induction.

These “single cell" experiments demonstrated the feasibility of implementing artificial logical operations using genetic modification. In [12], Savageau addresses the issue of finding general design principles among microbial genetic circuits, citing several examples. This theme is further developed in several Chapters of [2].

3 Bacterial Self-Assembly

When millions of bacteria act collectively as a group (as is the case in any bacterial colony), a large variety of different patterns arise as a result of how the individual cells respond to the others in their neighbourhood and to the conditions in their environments (which will be different for different individual bacteria, even if they are relatively close to each other) [5, 9, 13, 14]. The physical appearance of these patterns lies (largely) due to cells clumping together in different ways at different positions. These patterns are the emergent result of local interactions and environmental conditions, and can be usefully viewed as being *programmed* by both the particular way the bacteria in question interact, and by the particular way these bacteria respond to environmental signals. Also, these patterns are widely regarded as functional; for example, in stress conditions (such as a toxic chemical in the environment), the pattern formed will be ideal for protecting a maximal number of individuals (e.g. by minimising the number of cells without immediate neighbours) while promoting the colony’s “search" for a less harmful environment (by, e.g., extending “arms" of cells which explore the environment). This spontaneous development of functional patterns, purely as a result of environmental conditions and local interactions between individuals, is a form of computation which is extremely common and useful in nature, but which is currently very poorly understood and highly unexplored and unexploited in computer science.

3.1 Open questions

Previous work has focussed solely on the simulation of bacterial colonies, and coarse levels of similarity between actual and simulated pattern generation have already been achieved by previous researchers. Our goals are much more ambitious. We wish not only to be able to *predict* the final pattern, given the initial conditions, but to *control* the form of a final pattern by discovering the initial conditions that give rise to it. The level of control required for microtechnological applications can be gained only by a much better understanding of the effects of the factors contributing to pattern determination. These factors include environmental conditions (both physical and chemical), the rules and processes underlying bacterial interaction and the initial “program" manifest by the starting conditions. The fundamental questions that we seek to address are: **(1)** Is it possible to simulate the growth and movement of bacterial populations such that the complex patterns they generate *in vitro* can be *specified in advance* in computationally feasible times? **(2)** If achieved, how may the associated abstract computational architectures, as well as the directed self-assembly of bacteria *in vitro* be applied in various domains? **(3)** What level of complexity at the individual (bacterial) level is required to generate the observed complexity at the collective level? Additionally, how much of the observed complexity at the collective level is a reflection of environmental rather than individual complexity?

4 Conclusions

In this paper we argued that the execution of molecular algorithms using DNA is perhaps best performed *in vivo*, in order to take advantage of the numerous cellular processes within the cell (and between cells). Rather than attempting to rigidly implement computational models in an artificial manner we may, in future, prefer to modify our models to complement evolved molecular processes

References

[1] Leonard M. Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266:1021–1024, 1994.

[2] Martyn Amos, editor. *Cellular Computing*. Series in Systems Biology. Oxford University Press, USA, 2004.

[3] Martyn Amos. *Theoretical and Experimental DNA Computation*. Springer, 2005. In press.

[4] C.H. Bennett. The thermodynamics of computation – a review. *International Journal of Theoretical Physics*, 21:905–940, 1982.

[5] Elena O. Budrene and Howard C. Berg. Complex patterns formed by motile cells of *Escherichia coli*. *Nature*, 349:630–633, 1991.

[6] Michael Conrad and E.A. Liberman. Molecular computing as a link between biological and physical theory. *Journal of Theoretical Biology*, 98:239–252, 1982.

[7] M. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, January 2000.

[8] Tony L. Eng. On solving 3CNF-satisfiability with an *in vivo* algorithm. In *Preliminary Proceedings of the Fourth International Meeting on DNA Based Computers, University of Pennsylvania, June 15-19, 1998*, pages 163–171, 1998.

[9] Budrene E.O. and Berg H.C. Dynamics of formation of symmetrical patterns by chemotactic bacteria. *Nature*, 376:49–53, 1995.

[10] Richard P. Feynman. There's plenty of room at the bottom. In D. Gilbert, editor, *Miniaturization*, pages 282–296. Reinhold, 1961.

[11] T. Gardner, R. Cantor, and J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403:339–342, January 2000.

[12] Michael A. Savageau. Design principles for elementary gene circuits: Elements, methods and examples. *Chaos*, 11(1):142–159, 2001.

[13] J.A. Shapiro. Thinking about bacterial populations as multicellular organisms. *Ann. Rev. Microbiol.*, 52:81–104, 1998.

[14] J.A. Shapiro and M. Dworkin, editors. *Bacteria as Multicellular Organisms*. Oxford University Press, 1997.

[15] John von Neumann. The general and logical theory of automata. In *Cerebral Mechanisms in Behavior*, pages 1–41. Wiley, New York, 1941.

[16] R. Weiss, G. Homsy, and T.F. Knight Jr. Toward in-vivo digital circuits. In L.F. Landweber and E. Winfree, editors, *DIMACS Workshop on Evolution as Computation*. Springer, January 1999.

Introduction to Chemical Organization Theory

P Dittrich[1] & P Speroni di Fenizio[1,2]

[1]Bio Systems Analysis Group, Jena Centre of Bioinformatics and Department of Mathematics and Computer Science, Friedrich Schiller University Jena, D-07743 Jena, Germany
[2]ProtoLife, Parco Vega Via della Liberta' 12, Porta dell'Innovazione, 30175, Marghera, Venezia, Italia

Abstract

Complex dynamical networks consisting of many components that interact and produce each other are difficult to understand, especially, when new components may appear over time. Here, we outline a theory to deal with such systems. The theory consists of two parts. The first part introduces the concept of a chemical organization as a closed and self-maintaining set of components. This concept allows to map a complex (reaction) network to its set of organizations, which provides a new view on the system's "organizational structure". The second part connects dynamics with the set of organizations, providing a link to classical dynamical systems theory, e.g., by mapping a movement of the system in state space to a movement in the set of organizations. Finally, examples illustrating how the theory can be applied practically are presented.

1 Introduction

The knowledge concerning the molecular machinery of life is rapidly increasing; and with this increasing knowledge the size of dynamical models of bio-chemical reaction systems is increasing as well. For example, a recently published stochastic model of the central sugar metabolism of E. Coli consists of about 100 molecular species and 200 reaction rules [3]. However, these models are still relatively small compared to reaction systems that we obtain when molecules and reaction rules are defined implicitly, as for example in models of DNA computing or in models of combinatorial molecules or molecular complexes, such as, the nuclear pore complex or complexes formed by scaffold proteins. In these cases, combinatorial explosion can lead to thousands, millions, if not an infinite number of molecular species possible in the model.

A second aspect that should be noted is that usually not all possible species are present in a chemical reaction vessel. When we take the example of a DNA computing model, at a certain point in time, only a small subset of all possible DNA strands will be present in a reaction tube. So, when time proceeds, new molecular species may appear, and vice versa, existing molecular species may disappear completely. Following Fontana and Buss [2], we call systems where new objects may appear constructive. The theory described in this paper aims at such constructive dynamical systems.

1.1 Starting Point

The starting point when applying the theory is a reaction network given by a set of molecular species $\mathcal{M}$ and a set of reaction rules $\mathcal{R}$. So, the theory assumes that the topology of the reaction network is known. The aim of the theory is to (1) reveal the hierarchical structure of the network, (2) predict potential dynamical behavior, (3) describe the dynamics with respect to the network's structure.

The structure of the network will be described by the set of organizations, which form a hierarchical and overlapping structure. Inspired by Fontana and Buss [2] we define an organization as a set of molecular species that possesses two important properties, namely closure and self-maintenance[1]. So, this approach abstracts the state of a reaction systems by a set of molecules that are present in that state. In other words, instead of taking, e.g., a concentration vector to represent the state of a reaction system, we take the set of molecular species that have concentrations above zero (or above a small positive threshold Θ).

In the following, we denote sets and multisets of molecules by upper cases characters, such as $A, B \subseteq \mathcal{M}$. Note that by the very nature of a set, a molecule from $\mathcal{M}$ can only appear once in a set A. The set of all sets of molecules is denoted by $\mathcal{P}(\mathcal{M})$ (power set). We use multisets in order to represent the left hand side and the right hand side of reaction rules. A multiset is like a set, but the same molecule can appear several times. $\mathcal{P}_M(C)$ denotes the set of all multisets with elements from C, e.g., $\mathcal{P}_M(\{a,b\}) = \{\{\}, \{a\}, \{b\}, \{a,b\}, \{a,a\}, \{a,b\}, \{b,b\}, \{a,a,a\}, \dots\}$. The frequency of occurrence of an element a in a multiset A is denoted by $\#(a \in A)$.

Definition 1 (algebraic chemistry) Given a set $\mathcal{M}$ of molecular species and a set of reaction rules given by the relation $\mathcal{R} : \mathcal{P}_M(\mathcal{M}) \times \mathcal{P}_M(\mathcal{M})$. We call the pair $\langle \mathcal{M}, \mathcal{R} \rangle$ an *algebraic chemistry*.

For simplicity, we adopt a notion from chemistry to write reaction rules. Instead of writing $(\{s_1, s_2, \dots, s_n\}, \{s'_1, s'_2, \dots, s'_{n'}\}) \in \mathcal{R}$ we write: $s_1 + s_2 + \cdots + s_n \rightarrow s'_1 + s'_2 + \cdots + s'_{n'}$. And instead of writing $a + a \rightarrow b$ we can also equivalently write $2a \rightarrow b$. Note that "+" is not an operator here, but is used to separate the elements on both sides. Given the left hand side species $A = \{s_1, s_2, \dots, s_n\}$ and the right hand side species $B = \{s'_1, s'_2, \dots, s'_{n'}\}$, we write $(A \rightarrow B) \in \mathcal{R}$ instead of $(A, B) \in \mathcal{R}$. $A \rightarrow B$ represents a chemical reaction equation where A is the multiset of species on the left hand side (also called *reactants*) and B the multiset of species on the right hand side (also called *products*).

Example 1 (algebraic chemistry) In the example that we will use throughout the paper, the set of molecular species consists of five species: $\mathcal{M} = \{a, b, c, d, e\}$. There are 11 reaction rules: $\mathcal{R} = \{$ $a + b \rightarrow 2b$, $a + c \rightarrow 2c$, $b \rightarrow d$, $c \rightarrow d$, $b + c \rightarrow e$, $\emptyset \rightarrow a$, $a \rightarrow \emptyset$, $b \rightarrow \emptyset$, $c \rightarrow \emptyset$, $d \rightarrow \emptyset$, $e \rightarrow \emptyset\}$.

An algebraic chemistry is basically a reaction network, which represents the structure of a reaction system. Alternatively we might represent the reaction network by a directed bipartite graph, which consists of two node types: species and rules. An algebraic chemistry contains information about the stoichiometry, but no further details concerning the dynamics. Note that Def. 1 allows to define reaction rules or reaction networks that are not balanced, such as, $\{a \rightarrow a+b\}$ or $\{2a \rightarrow b, b \rightarrow a\}$, respectively. In chemistry, we usually demand that after a chemical reaction mass is conserved,

[1] Note that our definition of self-maintenance and organization is compatible with the concepts of Fontana and Buss. However, our form can be applied to arbitrary reaction networks.

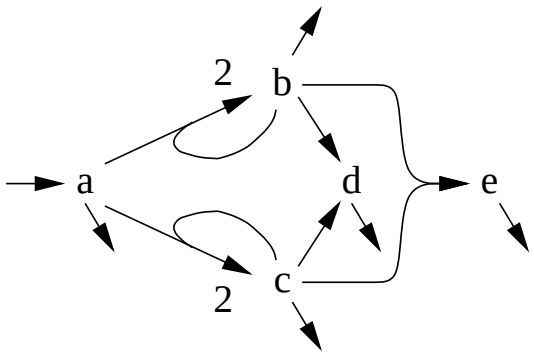


Figure 1: Reaction network that consists of 5 species and 11 rules. Lower case characters denote molecular species. Numbers written on the edges denote stoichiometric information. For example, the "2" at the lower left edge means that two molecules of species c are produced when one molecule of species a and one molecule of species c react. E.g., $a + c \rightarrow 2c$. Note that in order to produce d we need b *or* c, while in order to produce e we need b *and* c.

ie. the mass on the left hand side of a reaction rule is equal to the mass on its right hand side. Chemical organization theory was made to also handle systems that are not balanced and where mass is not necessarily conserved. A simple example of a model that is useful and at the same time not mass-conserving is given by the reaction rule $a \rightarrow 2a$, i.e. exponential growth.

1.2 Stoichiometric Matrix and Differential Equations

A common approach to describe chemical reaction systems is by a stoichiometric matrix, which can be used to derive an ordinary differential equation (ODE) model for the dynamics of the system based on mass action kinetics. Given a reaction system with m molecules and n reactions the *stoichiometric matrix* $\mathbf{M}$ has the form:

$$\mathbf{M} = \begin{pmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,n} \\ m_{2,2} & m_{2,2} & \dots & m_{2,n} \\ \dots & \dots & \dots & \dots \\ m_{m,1} & m_{m,2} & \dots & m_{m,n} \end{pmatrix} \tag{1}$$

where each row corresponds to a molecular species and each column to a reaction rule. An ODE describing the dynamics of the concentration vector $\mathbf{x} \in \mathbb{R}^m$ can be defined by

$$\dot{\mathbf{x}} = \mathbf{M}\mathbf{v}(\mathbf{x}) \tag{2}$$

where $\mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x}), \dots, v_n(\mathbf{x})) \in \mathbb{R}^n$ is a flux vector describing the rate of each reaction depending on the current concentrations $\mathbf{x}$. The following example illustrates how $\mathbf{v}(\mathbf{x})$ can be inferred from the reaction rules according to mass action kinetics.

Example 2 (stoichiometric matrix and mass-action kinetics) Given $m = 5$ molecular species $\mathcal{M} = \{a, b, c, d, e\}$ and $n = 11$ reactions as shown in Example 1, the stoichiometric matrix is

$$\mathbf{M} = \begin{pmatrix} -1 & -1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}. \tag{3}$$

Let $\mathbf{x} = (x_a, x_b, x_c, x_d, x_e)$ be the state of the reaction system at time t, where x_i denotes the concentration of species $i \in \mathcal{M}$, then we can define a flux vector $\mathbf{v}$ by

$$\mathbf{v}(\mathbf{x}) = \begin{pmatrix} k_1 x_a x_b \\ k_2 x_a x_c \\ k_3 x_b \\ k_4 x_c \\ k_5 x_b x_c \\ k_6 \\ k_7 x_a \\ k_8 x_b \\ k_9 x_c \\ k_{10} x_d \\ k_{11} x_e \end{pmatrix} \tag{4}$$

where k_j is a rate constant for reaction rule j. The ODE system reads:

$$\begin{aligned} \dot{x}_a &= -k_1 x_a x_b + -k_2 x_a x_c + k_6 - k_7 x_a, & (5)\\ \dot{x}_b &= +k_1 x_a x_b - k_3 x_b - k_5 x_b x_c - k_8 x_b, & (6)\\ \dot{x}_c &= +k_1 x_a x_c - k_4 x_c - k_5 x_b x_c - k_9 x_c, & (7)\\ \dot{x}_d &= k_3 x_b + k_4 x_c - k_{10} x_d, & (8)\\ \dot{x}_e &= k_5 x_b x_c - k_{11} x_e. & (9) \end{aligned}$$

Note that $\dot{x}_i$ denotes the current production rate of molecular species i. In the following we will denote the production rate of molecule i by $f_i := \dot{x}_i$.

2 Static Chemical Organization Theory

We will now define the central concept of the theory.

Definition 2 (organization) Given an algebraic chemistry $\langle \mathcal{M}, \mathcal{R} \rangle$ with $m = |\mathcal{M}|$ molecules and $n = |\mathcal{R}|$ reactions, and let $\mathbf{M} = (m_{i,j})$ be the $(m \times n)$ stoichiometric matrix implied by the reaction rules $\mathcal{R}$, where $m_{i,j}$ denotes the number of molecules of species i produced in reaction j. A set of molecules $O \subseteq \mathcal{M}$ is called an *organization*, if there exists a flux vector $\mathbf{v} \in \mathbb{R}^n$ such that the four following condition apply: (1) for all reactions $(A \to B)$ with $A \in \mathcal{P}_M(O)$ the flux $v_{(A \to B)} > 0$; (2) for all reactions $(A \to B)$ with $A \notin \mathcal{P}_M(O)$, $v_{(A \to B)} = 0$; (3) for all molecular species $i \in O$ is the production rate $f_i \geq 0$ (self-maintaining); and (4) for all molecular species $i \notin O$ is the production rate $f_i = 0$ (closure). The production rates are given by $(f_1, \ldots, f_m)^T = \mathbf{M}\mathbf{v}$.

A set C where conditions (1)-(3) apply but not necessarily condition (4) is called *self-maintaining*, since it is possible to maintain all its components from a stoichiometric point of view. If conditions (4) is true for a flux vector that obey rule (1) and rule (2), then the set C is called *closed*. The condition of closure can be defined more easily, without referring to possible flux vectors:

Definition 3 (closed set) Given an algebraic chemistry $\langle \mathcal{M}, \mathcal{R} \rangle$, a *closed set* $C \in \mathcal{M}$ is a set of molecular species such that for all $A \subseteq C$ the following holds: if there is a reaction $(A \to B) \in \mathcal{R}$, then $B \subseteq C$.

Given a set S, $S \subseteq \mathcal{M}$, it is always possible to generate its *closure* $G_{CL}(S)$ [4]. To generate the closure of a set we expand it by interacting the molecules of the set and adding to the set any newly generated molecule. When no new molecule is generated, the set is closed.

Definition 4 (generate closed set)) Given a set of molecules $S \subseteq \mathcal{M}$, we define $G_{CL}(S)$ as the smallest closed set C containing S. We say that S *generates the closed set* $C = G_{CL}(S)$ and we call C the *closure* of S.

2.1 Consistent Reaction Systems

We can now define a generate operator for self-maintaining sets in the same way as for closed sets by saying that the self-maintaining set generated by a set S is the biggest self-maintaining set C contained in S. "Unfortunately" this set C is not unique for arbitrary reaction systems, ie. for arbitrary algebraic chemistries. We call chemistries where a self-maintaining set can be uniquely generated for any set by the procedure described above *consistent*. So, in a consistent algebraic chemistry we can always generate uniquely for any given set $S \subseteq \mathcal{M}$ a self-maintaining set C by taking the biggest self-maintaining set C that contains S.

Definition 5 (consistent) An algebraic chemistry is called *consistent*, if the closure and self-maintaining set generated by a set can uniquely be defined, ie. given any set $S \subseteq \mathcal{M}$, the smallest closed set that contains S and the largest self-maintaining set contained in S are unique, respectively.

If our reaction system is consistent, a couple of interesting properties hold, which are explained in the following. First of all we define "generate self-maintaining set" formally:

Definition 6 (generate self-maintaining set) Given a consistent algebraic chemistry $\langle \mathcal{M}, \mathcal{R} \rangle$ and a set of species $S \subseteq \mathcal{M}$, we define $G_{SM}(S)$ as the biggest self-maintaining set C contained in S. We say that S *generates the self-maintaining set* $C = G_{SM}(S)$.

Since in consistent reaction systems the closure and self-maintaining set can be generated uniquely, we can also uniquely define the organization generated by a set S in the following way:

Definition 7 (generate organization) Given a set of molecules $S \subseteq \mathcal{M}$, the organization $O = G(S)$ generated by S is defined as $G(S) \equiv G_{SM}(G_{CL}(S))$.

If O is an organization $G(O) = O$. So, the organizations are the fixed points of the "generate organization operator" G. The generate operator implies a union and an intersection operator on the set of organizations $\mathcal{O}$ of an algebraic chemistry.

Definition 8 (Union $\sqcup$ and intersection $\sqcap$ of organizations) Given two organizations U and V, the organization generated by their union ($U \sqcup V$) and intersection ($U \sqcap V$) are defined as

$$U \sqcup V \equiv G(U \cup V), \tag{10}$$
$$U \sqcap V \equiv G(U \cap V). \tag{11}$$

Lemma 1 Given an algebraic chemistry $\langle \mathcal{M}, \mathcal{R} \rangle$ and all its organizations $\mathcal{O} = \{O \subseteq \mathcal{M} | O \text{ is an organization}\}$, then $\langle \mathcal{O}, \sqcup, \sqcap \rangle$ is an (algebraic) lattice.

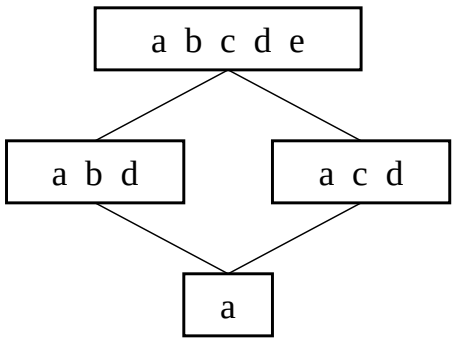


Figure 2: Hasse-diagram of the lattice of organizations of the network shown in Fig. 1.

A lattice is an algebraic structure that can be visualized by a Hasse-diagram (Fig. 2). Here the vertical position of an organization is determined by the number of molecules contained in it. The largest organization, which always exists in a finite lattice, can be found at the top of the Hasse-diagram. At the bottom, we can see the smallest organization, which is the organization generated by the empty set.

3 Dynamics

The static part of the theory (Sec. 2) deals with molecules $\mathcal{M}$ and reaction rules $\mathcal{R}$ but not with time. For that reason we have defined the concept of an *algebraic chemistry*, which is sufficient as input for the static part. In order to add dynamics to the theory, we have to formalize the dynamics of a system. In a very general approach, the *dynamics* is given by a *state space* X and a formal definition (mathematical or algorithmical) that describes all possible movements in X, so given an initial state $\mathbf{x}_0 \in X$, the formal definition describes how the state changes over time. Now, for simplicity, we assume a deterministic dynamical process, which can be formalized by a phase flow $(X, (T_t)_{t\in\mathbb{R}})$ where $(T_t)_{t\in\mathbb{R}}$ is a one-parametric group of transformations from X. $T_t(\mathbf{x}_0)$ denotes the state at time t of a system that has been in state $\mathbf{x}_0$ at $t = 0$.

For connecting the static theory with dynamics we introduce a mapping ϕ called *abstraction*, from X to $\mathcal{M}$, which maps a state of the system to the set of molecules that are present in the system being in that state. The exact mapping can be defined precisely later, depending on the state space, on the dynamics, and on the actual application.

The concept of *instance* is the opposite of the concept of abstraction. While $\phi(\mathbf{x})$ denotes the molecules represented by the state $\mathbf{x}$, an instance $\mathbf{x}$ of a set A is a state where exactly the molecules from A are present according to the function ϕ.

Definition 9 (instance of A) We say that a state $\mathbf{x} \in X$ is an instance of $A \subseteq \mathcal{M}$, iff $\phi(\mathbf{x}) = A$.

In particular, we can define an instance of an organization O (if $\phi(x) = O$) and an instance of a generator of O (if $G(\phi(x)) = O$). Loosely speaking we can say that $\mathbf{x}$ *generates organization* O. Note that a state $\mathbf{x}$ of a consistent reaction system is *always* an instance of a generator of one and only one organization O. This leads to the important observation that a lattice of organizations partitions the state space X, where a partition X_O implied by organization O is defined as the set of all instance of all generators of O: $X_O = \{\mathbf{x} \in X | G(\phi(\mathbf{x})) = O\}$. Note that as the system state

evolves over time, the organization $G(\phi(\mathbf{x}(t)))$ generated by $\mathbf{x}(t)$ might change. A movement in state space can thus be mapped to a movement in the set of sets of species, and finally to a movement in the set of organizations, which provides a new way of visualizing a trajectory in a high-dimensional state space (see Ref. [5, 1]).

3.1 Fixed Points are Instances of Organizations

In this section we present a theorem stating that every fixed point must be an instance of an organization. Or in other words, in a continuous dynamical reactions system given by an ODE, we cannot obtain a stationary state with a combination of molecular species that are not an organization. For the theorem, we have to specify formally the abstraction function:

Definition 10 (abstraction) Given a dynamical system $\dot{\mathbf{x}} = f(\mathbf{x})$ and let $\mathbf{x}$ be a state in X, then the abstraction $\phi(\mathbf{x})$ is defined by

$$\phi(\mathbf{x}) = \{i | x_i > \Theta, i \in \mathcal{M}\}, \quad \phi : X \to \mathcal{P}(\mathcal{M}), \quad \Theta \geq 0 \tag{12}$$

where x_i is the concentration of molecular species i in state $\mathbf{x}$, and Θ is a threshold chosen such that it is smaller than any positive coordinate of any fixed point of $\dot{\mathbf{x}} = f(\mathbf{x}), x_i \geq 0$.

Theorem 1 **Hypothesis:** Let us consider a general reaction system whose reaction network is given by the algebraic chemistry $\langle \mathcal{M}, \mathcal{R} \rangle$ and whose dynamics is given by a differential equation $\dot{\mathbf{x}} = \mathbf{M}\mathbf{v}(\mathbf{x}) = f(\mathbf{x})$ as defined before. Let $\mathbf{x}' \in X$ be a fixed point, that is, $f(\mathbf{x}') = \mathbf{0}$, and let us consider a mapping ϕ as given by Def. 10, which assigns a set of molecules to each state $\mathbf{x}$. **Thesis:** $\phi(\mathbf{x}')$ is an organization. (Proof see Ref. [1].)

Acknowledgment: We acknowledge support by Federal Ministry of Education and Research (BMBF) Grant 0312704A to Friedrich Schiller University Jena.

References

[1] Peter Dittrich and Pietro Speroni di Fenizio. Chemical organization theory: Towards a theory of constructive dynamical systems. *(submitted), preprint arXiv:q-bio.MN/0501016*, x(x):1–7, 2005.

[2] W. Fontana and L. W. Buss 'The arrival of the fittest': Toward a theory of biological organization. *Bull. Math. Biol.*, 56:1–64, 1994.

[3] J. Puchalka and A.M. Kierzek. Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. *Biophys. J.*, 86(3):1357–1372, 2004.

[4] P. Speroni di Fenizio, P. Dittrich, J. Ziegler, and W. Banzhaf. Towards a theory of organizations. In *German Workshop on Artificial Life (GWAL 2000), in print*, Bayreuth, 5.-7. April, 2000, 2000.

[5] Pietro Speroni Di Fenizio and Peter Dittrich. Artificial chemistry's global dynamics. movement in the lattice of organisation. *The Journal of Three Dimensional Images*, 16(4):160–163, 2002.

The Role of Chemistry and Mechanics in the Morphogenesis of Walled Cells

Jacques Dumais*

*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, USA

Abstract

Cell morphogenesis is the product of chemical and mechanical processes. How chemistry and mechanics interact to create cell shape is one of the fundamental questions in biology. Substantial progress has been made in our understanding of morphogenesis in walled cells, that is, cells that are surrounded by a stiff extracellular matrix. Walled cells are found in bacteria, fungi, algae, and plants. They encompass a large fraction of the morphological diversity present in living organisms; yet, the observed cellular shapes can be understood from a few basic mechanical principles first derived from studies of soap bubbles and rubber balloons. In this lecture, I will consider some examples of cell morphogenesis and show how the internal pressure of the cell and the molecular architecture of the cell wall contribute to cell shape. I will also present a general paradigm for how chemistry and mechanics contribute to the morphogenesis of walled cells.

1 The question

Several decades of intense molecular investigations have highlighted the ingenuity of Nature in using chemistry to fulfil specific functions. The capture of solar energy through photosynthesis, the storage of information in DNA, and the generation of forces by molecular motors are but a few examples of Nature's great achievements in chemistry. There are, however, realms where chemistry falls short of providing complete answers. One of them is cell morphogenesis where both chemistry and mechanics play important roles. D'Arcy W. Thompson [1] is widely credited for popularizing the idea that many aspects of cell morphogenesis are governed by geometry and physical principles such as surface tension and pressure. Although some of Thompson's ideas have not survived further scrutiny, the premises on which his work was based remain valid today. We must therefore ask to what extent mechanics contributes to cell morphogenesis? I want to address this question in the context of walled cells.

2 Cell walls and their role in Nature

The cell wall is a stiff extracellular matrix found in archea, eubacteria, fungi, algae, and plants. The structure and composition of cell walls in prokaryotes and eukaryotes vary greatly although in fungi, algae and plants, the cell wall takes the form of a fiber re-inforced composite. Typically, fibrils of chitin in fungi or cellulose in plants provide stiffness to the wall while an amorphous matrix of pectins helps maintain the cohesion between the different wall components.

Cell walls are widely distributed because they offer an effective solution to a fundamental problem faced by living cells – that of maintaining osmotic equilibrium with their environment. The concentrations of solutes (ions, proteins, sugars) inside cells typically exceed that of the surrounding environment. Therefore, water tends to flow into cells by osmosis. If left unchecked, the influx of water would soon burst the cell. A stiff cell wall can prevent bursting by resisting the expansion

of the cell surface. The net influx of water stops when the osmotic force driving water inside the cell is balanced by the opposing turgor pressure of the cell. The osmotic balance is governed by the equation $\Pi_{out} = \Pi_{in} + P$, where Π_{out} and Π_{in} are the osmotic potentials of the environment and the cytoplasm, and P is the turgor pressure of the cell. Cells can regulate their turgor pressure by modifying the osmotic potential of the cytoplasm. A typical turgor pressure for plant cells is 0.5 MPa (˜5 atm).

The evolution of cell walls created a new problem, that of cell expansion. To acquire a specific shape, walled cells must extend the stiff "straitjacket" that surrounds them. Cell expansion is achieved by modulating the mechanical properties of the cell wall to promote local yielding to turgor pressure [2-6]. Accordingly, the relative rate of wall expansion is determined by two mechanical factors: the turgor-induced tensile stresses present in the cell wall and the local mechanical properties of the wall. Variations in the morphology of walled cell must ultimately be traced back to variations in one of these two controlling factors.

3 Mode of morphogenesis in walled cells

Walled cells have adopted a limited number of approaches for growth. The two most widespread modes of morphogenesis in single cells are known as "diffuse growth" and "tip growth". Morphogenesis by diffuse growth was studied extensively in the alga *Nitella*. The cylindrical internodal cell of *Nitella* elongates greatly along its axis while circumferential growth is much reduced [7]. A similar phenomenon is observed in the internode of *Acetabularia* (Fig. 1B). On the other hand, a simple calculation reveals that the tensional stresses created by turgor pressure are twice as high in the circumferential direction compared to the axial direction. In other words, turgor pressure favors a fattening of the cell not its elongation. The discrepancy between the observed deformation of the cell and the forces exerted by turgor pressure can be explained if the mechanical properties in the plane of the cell wall are varying with direction. It has been argued that circumferentially-oriented cellulose microfibrils provide the mechanical anisotropy required to explain the expansion of diffusely-growing cells [8,9]. The cellulose reinforcement prevents extension in the circumferential direction despite the greater tension in that direction. The deposition of cellulose microfibrils is believed to be guided by circumferentially-oriented cortical microtubules [10]. If the preferential orientation of cellulose microfibrils is perturbed, the expansion anisotropy is lost and the cell soon adopts a spherical shape [8].

The second major mode of cell morphogenesis, tip growth, is found in a wide range of cells including some prokaryotes (e.g. Streptomycetes) and many eukaryotes such as fungal hyphae, certain unicellular algae, pollen tubes, and root hairs. The shape of tip-growing cells is characterized by a long cylinder capped with a prolate dome as illustrated in Fig. 1D. Wall expansion is limited to the tip of the cell and forms a meridional gradient with a maximum at or near the pole and decreasing to zero near the equator. It has also been observed that over most of the dome surface, wall expansion in the circumferential direction exceeds expansion in the meridional direction [11,12]. Contrary to what is observed in diffusely-growing cells, the anisotropy of wall expansion in tip-growing cells does not seem to depend on structural anisotropy in the cell wall. This conclusion can be drawn from reports that wall structure on the dome of various tip-growing cells is random [13-16]. On the other hand, a quantitative analysis of wall expansion during tip growth has established that the tensional stresses created by turgor pressure contribute to a large fraction of the observed wall expansion [12]. An analysis of expansion in rubber balloons has revealed striking similarities with expansion in tip-growing cells suggesting that common mechanical principles may be at work in these disparate structures [11].

The shape of cells in tissues depends not only on cell expansion but also on the positioning of new walls when the cell divides. Thompson [1] has explained how complex cellular patterns, such

as the pattern shown in Fig. 1A, can be built from three simple rules: i) cells divide so that the two daughter cells have equal volumes, ii) new cell walls contact older cell walls at right angle, iii) new cell walls are as short as possible. These rules are reminiscent of the laws that govern organization of soap bubbles in a froth and have lead Thompson to argue for similar laws in biology. All of these observations emphasize the importance of mechanics in determining fundamental aspects of cell morphogenesis.

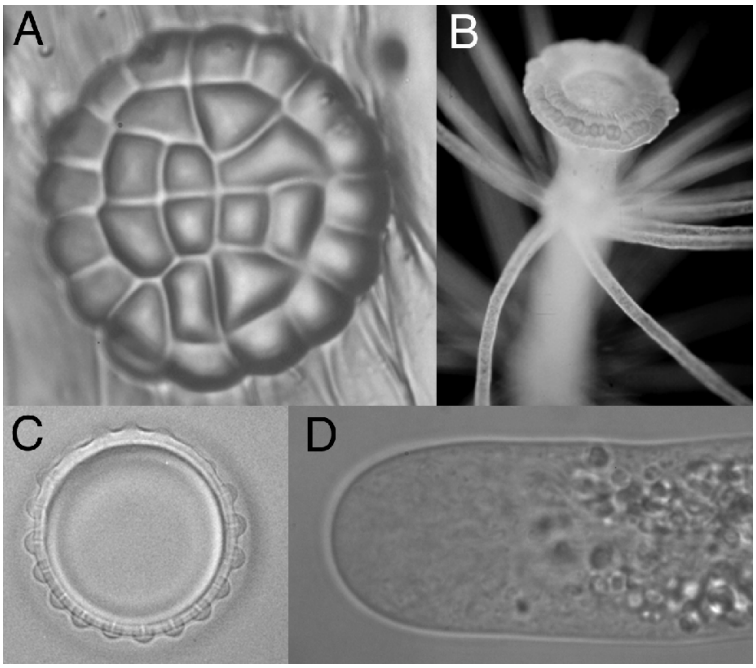


Figure 1: Some examples of shapes that can be achieved in walled cells. (A) Two-dimensional cellular pattern in a digestive gland of the Venus flytrap. (B) A vegetative whorl and a young reproductive cap in *Acetabularia acetabulum*. (C) Breaking of radial symmetry during the formation of the vegetative whorl in *Acetabularia acetabulum*. (D) Tip growth in a Lily pollen tube.

4 On the control of growth and form in walled cells

Although mechanics can explain many aspects of cell morphogenesis, particularly in walled cells, a satisfactory explanation for growth and form in the context of modern biology must relate features of the genotype to features of the phenotype, here cell shape. The objective of this closing section is to present a paradigm for the molecular and mechanical control of growth and form.

Most theories of cell morphogenesis postulate an involvement of the cytoskeleton. In walled cells, the cytoskeleton can serve to localize and orient wall structure because if its ability to transduce biochemical information into spatial information [17]. I propose a general causal chain for the genetic control of morphogenesis where genes and form are connected through four major transduction steps beyond the well understood steps of transcription, translation, and protein folding (Fig. 2). First, cytoskeletal self-assembly and dynamic instability are used by the cell to transduce molecular signals into a spatial structure. These cytoskeletal structures provide polarity and anisotropy to the cell. Second, the labile cytoskeletal structures serve as templates for cell wall assembly. Third, wall architecture dictates the mechanical behavior of the cell wall. This connection is encapsulated in the structure-property relationship of the cell wall. Finally, the turgor stresses and mechanical properties of the cell wall are integrated over time to specify growth and form.

The causal chain outlined in Fig. 2 indicates that the effect of genes on form is masked by several complex transduction steps. Attempts to explain organic form in molecular terms must take into account these transduction steps where new properties can emerge.

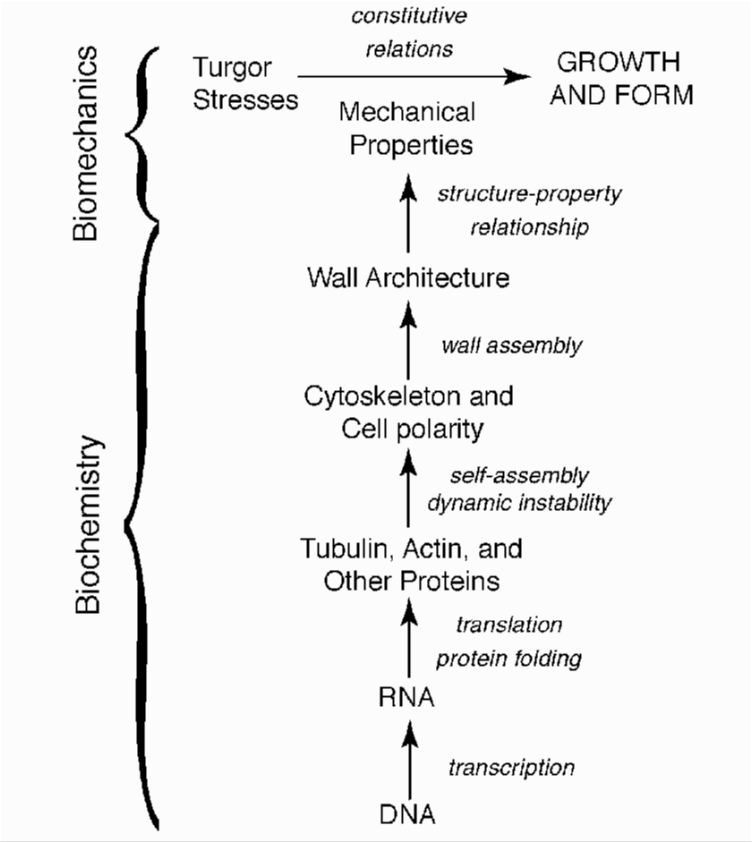


Figure 2: Mechanical and chemical controls of cell morphogenesis. The causal chain that connects DNA to form includes six “transduction” steps. Note that feedbacks exist between the different levels but have not been included in this figure for simplicity.

References

[1] D’A.W. Thompson, D’A.W. (1942) On Growth and Form. Cambridge University Press (Dover republication, 1992)

[2] A.N.J. Heyn. Bot. Rev., 6: p515, 1940.

[3] P.B. Green. J. Cell Biol, 27: p343, 1965.

[4] J.A. Lockhart. J. Theor. Biol., 8: p264, 1965.

[5] P.M. Ray. In D.J. Cosgrove and D.P. Knievel, eds., Physiology of Cell Expansion During Plant Growth. American Society of Plant Physiologists. p1, 1987.

[6] D.J. Cosgrove. Plant Physiol., 102: p1, 1993.

[7] P.B. Green. Brookhaven Symp. Biol., 16: p203, 1963.

[8] P.B. Green. Science, 138: p1404, 1962.

[9] M.C. Probine and R.D. Preston. J. Exp. Bot. 13: p111, 1962.

[10] C. Lloyd. Int. Rev. Cytol., 86: p1, 1984.

[11] Z. Hejnowicz, B. Heinemann, and A. Sievers. Z. Pflanzenphys., 81: p409, 1977.

[12] J. Dumais, S.R. Long, and S.L. Shaw. Plant Physiol., 136: p3266, 2004.

[13] D.S. Belford and R.D. Preston. J. Exp. Bot., 12: p157, 1961.

[14] A.L. Houwink and P.A. Roelofsen. Acta Bot. Neerl., 3: p385, 1954.

[15] E.H. Newcomb and H.T. Bonnett. J. Cell Biol., 27: p575, 1965.

[16] J.C. O'Kelley and P.H. Carr. Amer. J. Bot., 41: p261, 1954.

[17] M. Kirschner and T. Mitchison. Cell 45: p329, 1986.

Structural analysis of metabolic networks using elementary modes

David A. Fell*

* School of Biological and Molecular Sciences, Oxford Brookes University, Headington, Oxford, OX3 0BP, U.K.

We have understood for several decades how the genome encodes the metabolic phenotype of an organism by means of genes specifying the enzymes that catalyze the reactions of the metabolic network. Surely we should be able to exploit that understanding to make correct predictions about metabolic properties and responses to perturbations? There are several potential application areas, including metabolic engineering and the design of effective drug therapies. Furthermore, if we can achieve this with metabolism, related cell processes should yield to a similar approach, such as signal transduction, the cell cycle, and apoptosis. Here we will consider approaches and issues in analysing the functioning of metabolic networks.

1 Formal representation of metabolic pathways

The metabolic network is composed of an interconnected set of enzymic reactions. The rates of these reactions are non–linear functions of the concentrations of the metabolites, with the familiar Michaelis–Menten equation for a single–substrate, irreversible enzyme being just the starting point:

$$v = \frac{SV}{S + K_{\mathrm{m}}}$$

In practice, even for a single substrate enzyme, a more realistic equation is:

$$v_{net} = \frac{(V_{\mathrm{f}}/K_{\mathrm{m,S}})\,(S - P/K_{\mathrm{eq}})}{1 + S/K_{\mathrm{m,S}} + P/K_{\mathrm{m,P}}}$$

Predicting the behaviour of the whole network would therefore seem to require the knowledge of many complex enzyme rate functions, and even more enzyme rate parameters. Mathematically, there are techniques for solving such problems by numerical simulation, but that still leaves the problem of acquiring all the enzyme data needed. However, there are properties of the network that can be deduced without needing all this information. To understand this, it is necessary to consider how metabolic networks can be represented mathematically.

Consider the following specimen pathway:

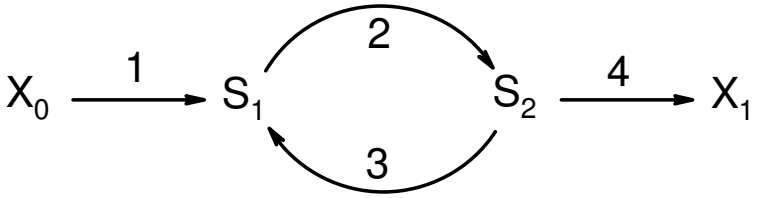
$$X_0 \xrightarrow{xase} Y \xrightarrow{ydh} Z \xrightarrow{zase} X_1$$

X_0 is termed the *source*, and X_1 is the *sink*. They are also termed *external metabolites*. Y and Z are the variable, or *internal* metabolites that reach constant levels at steady state, when their rates of formation equal their rates of utilization.

In such a metabolic pathway there is a flow of matter from the source to the sink. At *s*teady state, the concentrations of the intermediates remain constant because their rates of formation exactly equal their rates of degradation. The flow through the pathway also remains constant. Even if there are very slow changes in the concentrations of metabolites, or the pathway flux, because of slow changes in the source or sink, the pathway may be regarded as being in *quasi–steady state* provided the time scale of the changes is very much longer than the time taken by the pathway

to approach steady state. Just as metabolic experiments are simpler to reproduce and interpret if they are designed to give steady state or quasi–steady state conditions, so it is simpler to analyse network properties at steady state.

To illustrate this, consider a slightly more complex, though still simple pathway:



Apart from this diagrammatic form, two other types of representation contain the same information:

$$\begin{array}{c} \\ S_1 \\ S_2 \end{array} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix}$$

r1: X_0 -> S_1 $\sim$
r2: S_1 -> S_2 $\sim$
r3: S_2 -> S_1 $\sim$
r4: S_2 -> X_1 $\sim$

The one on the left is the *stoichiometry matrix* where the numbers in the matrix represent the moles of the substrate specified by the row that take part in the reaction specified by the column (with negative numbers indicating substrates of the reaction). The form on the right is a symbolic list of reactions. All three forms can be interconverted, but the stoichiometry matrix is the basis for further mathematical analysis.

If the rate of reaction 1 is given by v_1, and so on, inspection of the diagram shows that the net rates of change of S_1 and S_2 will be given by:

$$\frac{dS_1}{dt} = v_1 - v_2 + v_3$$

$$\frac{dS_2}{dt} = v_2 - v_3 - v_4$$

At steady state $\frac{dS_1}{dt} = 0$ and $\frac{dS_2}{dt} = 0$. In addition to the above two equations, this implies $v_1 = v_4$. How can we generalize this, so that the relationships between rates 1 and 4 would be evident even in complex networks?

The rate at which the substrate concentrations are changing can be written more generally and compactly as $\mathbf{N}.\mathbf{v}$, where $\mathbf{N}$ is the stoichiometry matrix, and $\mathbf{v}$ is a vector of the enzyme kinetic functions. So for our substrate cycle pathway:

$$\begin{bmatrix} \frac{dS_1}{dt} \\ \frac{dS_2}{dt} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

where each v_i is the rate function for enzyme i, depending on the metabolites, V_{m}, K_{m} etc.

It turns out that this has separated the structural constraints from the detailed kinetics of the network, because the equation above is a linear equation in the rates, with coefficients given by $\mathbf{N}$, which represents the structure of the network, even though the rates themselves are complicated non–linear functions of the metabolite concentrations.

2 Methods of structural analysis

Any metabolic pathway at steady state satisfies the relationship $\mathbf{N}.\mathbf{v} = \mathbf{0}$, where $\mathbf{N}$ is the stoichiometry matrix, exemplified by the substrate cycle pathway:

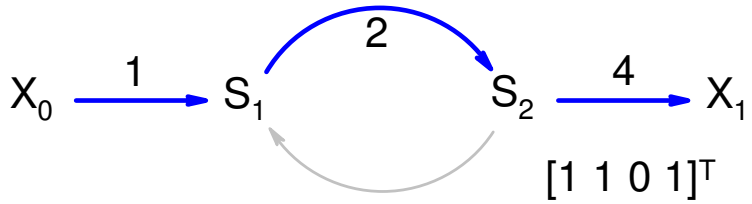
$$\begin{array}{c} S_1 \\ S_2 \end{array} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

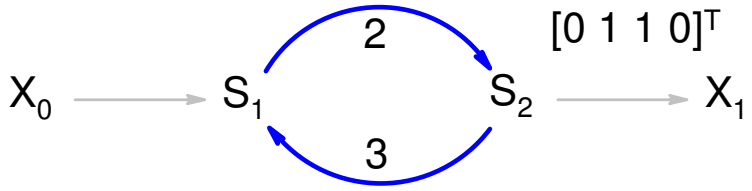
Any observed set of velocities at steady state will be a linear combination of a set of vectors $\mathbf{K}$ referred to as a basis for the *null space* of the stoichiometry matrix. It does not matter for our purposes how a basis is calculated, though computer algebra programs generally provide a function for doing this. In this case, a suitable basis could be:

$$\mathbf{K} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

The vectors are in essence, prototype solutions of the problem, since there are no unique solutions in this case (or for metabolic networks in general) . The number of null space vectors tells us the number of independent fluxes that can exist in the pathway — in this case, a linear flux and a cyclic flux.

The null space vectors can be considered as pathways through the network:





Any feasible set of velocities at steady state is a linear combination of these null space vectors, e.g.:

$$\mathbf{K} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a \\ a+b \\ b \\ a \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

Without the kinetic rate functions,we cannot predict actual values of a and b for a cell, but it seems that we can decompose the network into components that correspond to different ways in which the reactions can operate.

This may seem limited, but it means that we can decide questions about metabolism that depend only on the routes through a network, and not on how fast they work. For example, if we can prepare a list of the reactions coded by the genome of an organism, we should be able to decide:

- what nutrients it can utilize and what products it can produce;
- whether there is a route from a particular nutrient to a product;
- which route to a product has the highest yield;
- the consequences of deleting an enzyme, and
- whether the genome annotations generate a connected and self-consistent metabolism.

It turns out that these questions are answerable, though null space calculation is not the only or best way to answer them. To get round some of the problems with using null space vectors to answer these questions, we use the concept of elementary modes, originally developed by Stefan Schuster.

3 Elementary modes analysis

An elementary mode is a minimal set of enzymes that can operate at steady state with all irreversible reactions working in the thermodynamically favoured direction, and enzymes weighted by the *relative* flux they carry. 'Steady state' implies that there is only net production or consumption of external metabolites. Production and consumption of all internal metabolites is balanced. 'Minimal' means that deleting any enzyme in the set would prevent a steady state. By definition, an elementary mode is not decomposable into component elementary modes. It turns out, given these conditions, that the set of elementary modes of a reaction network is unique.

For example, if we start with the set of carbohydrate–metabolizing enzymes shown in Fig. 1, we can determine the six elementary modes of the system.

The elementary modes turn out to include glycolysis and several modes of operation of the pentose phosphate pathway, even though no biochemical knowledge is involved in the mathematical procedure that calculates the modes. In this case, we already knew most of the answers beforehand; can we find something that was not previously known.

Predicting new pathways

In Schuster et al(1999), we analysed the elementary modes of the TCA cycle and glyoxylate cycle enzymes. Amongst the modes we found was one that used a hybrid of the two cycles to catabolize glycolytic phosphoenolpyruvate completely to CO_2, according to the equation:

ADP + FAD + 4 NAD + PEP -¿ ATP + FADH2 + 4 NADH + 3 CO_2

Thsi catabolic form of the the glyoxylate cycle seemed unusual, since it is widely regarded as a route for forming cell biomass from two–carbon compounds. In 2003, Fischer and Sauer observed this pathway in *E coli* cells growing at low glucose levels. Flux measurements showed it combined with a background level of the TCA cycle.

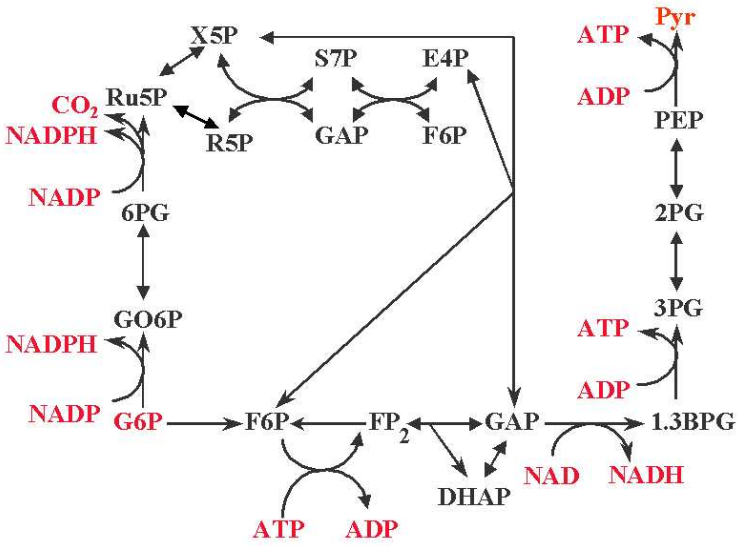


Figure 1: Reactions of glycolysis and pentose phosphate metabolism used for elementary modes analysis.

Enzyme knockouts: G6PDH

Glucose-6-phosphate dehydrogenase deficiency (favism) is a common enzymopathy in humans. Targeted G6PDH knockout in mouse cells produces clones that can grow (i.e. make ribose for nucleic acid synthesis and some NADPH), but that are sensitive to oxidative stress (increased NADPH demand). The elementary modes obtained for the pentose phosphate pathway shown above do include a mode for making ribose by sugar rearrangements without use of glucose-6-phosphate dehydrogenase. In other cases, it is possible to find enzymes whose absence cannot be bypassed in this way, and therefore inhibition by a drug or mutation of their genes would be expected to give rise to a metabolic deficiency.

Engineering pathways

Carlson, Srienc and I (2002) used elementary modes analysis to investigate how effectively *S*accharomyces cerevisiae could make the biodegradable plastic precursor polyhdroxybutyrate (PHB) after expression of the bacterial genes for this pathway in yeast. Analysis of the yeast metabolic network suggested that acetate and ethanol could be used as effective precursors after the diauxic shift, and 67% of the input carbon could be incorporated in PHB. This corresponds with experimental observations that the formation is highest after the diauxic shift of glucose–grown yeast.

We also investigated whether a cytosolic ATP citrate synthase (EC 2.3.3.8) would assist in PHB synthesis, which was occurring in the cytosol, rather than in the mitochondrion where the substrate acetyl coAwould be more readily available. We found that elementary modes analysis predicted that adding this enzyme to the yeast network would increase the carbon conversion efficiency from acetate to 83%.

4 Enzyme subset analysis

An even simpler structural feature than an elementary mode is an enzyme subset. This is defined as a group of enzymes that carry flux in a fixed proportion in any steady state where they are active, so that if any of them are present in an elementary mode, they all are, or else none at all.

For a simple linear pathway with no branches, all the enzymes are in a single enzyme subset, but the subsets are not so easy to see in a complicated network. A use of enzyme subsets is in aiding genome annotation of newly–sequenced organisms by suggesting missing enzymes from 'broken' subsets. For example, elementary mode of the carbohydrate–metabolizing enzymes that had been assigned to open reading frames in the *T. pallidum* genome only generated a single, rather strange elementary mode. This mode failed to use several of the enzymes that had been identified in the annotation, including transketolase and ribose-5-phosphate epimerase. However, in general microbial metabolism, these two enzymes form a subset with transaldolase, which had not been found in the annotation. It therefore seems likely that the genome does contain a transaldolase somewhere, since the other two enzymes would have no function if it did not.

5 Summary

This survey of structural modelling of metabolic networks has not covered all the tools that can be used, but my intention was to illustrate that there are aspects of metabolism that can be modelled with information that is limited to the presence or absence of enzymes in the system. In conclusion, structural analysis can be used for:

- Determining valid routes through metabolic networks;
- Analyzing metabolic capabilities;
- Finding susceptibilities to drugs and mutations;
- Aiding metabolic engineering, and
- Identifying metabolic subsystems to aid genome annotation.

6 Further reading

- S Schuster, T Dandekar & D A Fell, Trends in Biotechnology, 17, 53-60 (1999).
- S Schuster, D A Fell & T Dandekar, Nature Biotechnol. 18, 326-332 (2000)
- Papin, J. A. et al, Metabolic pathways in the post-genome era, Trends Biochem.Sci. 28, 250-258(2003).

Meaning in Evolution

Brian Goodwin*

*Schumacher College, Dartington Devon TQ9 6EA, UK bcgood1401@aol.com

Abstract

Nature and culture have been sharply separated in modern thought. A major reason for this is the belief that language and meaning apply to humans and culture, but not to the evolution of species generally. However, recent studies on the structure of proteomes, on genetic and metabolic networks, are leading to a new perspective on the nature of the processes involved in reading and expressing the information in the genome. These are beginning to be recognised as having the network properties of a language, so that a reading of the genetic text by an organism is a process that makes meaning of the text through the self-construction of the organism. The members of a species are then participants in a culture with a language. They make meaning of their inherited texts by generating a form (a distinctive morphology and behaviour pattern) that is dependent on both genetic text and external context. This understanding of development and evolution arises from experimental observation and mathematical modelling that are described, and leads to an extended conceptual context for understanding living processes.

1 Information and the Genome.

The molecular biology revolution of the 20^{th} century was driven by the realisation that the hereditary polymers of organisms, their DNA and RNA, carry information that is used to make the molecular components of cells whose finely-discriminating interactions can also be understood as forms of information. Instead of chemical machines, organisms became information machines, like computers, only more complex since they self-replicate. The genome projects were a logical extension of this perspective, based on the premise that since we know how to read the genetic code we can make sense of all aspects of organisms by reading the information in their genomes. This led to a period of dramatic technological and analytical development in biology as genetic data-bases were revealed and the details of molecular information-processing unfolded.

However, natural processes have an uncanny way of escaping from well-designed research nets intended to catch and reveal their essential natures. We need only recall the astonishment and bewilderment of the physics community in the early years of the 20^th century at the revelations of quantum mechanics as the constituents of atoms disappeared into a cloud of energetic entanglements that remain mysteriously holistic despite mathematical descriptions of their behaviour. Living processes can likewise be expected to lead us on to some quite unexpected revelations.

Something rather similar to the quantum mechanical revelation is now happening to the genes, which seemed to have been caught in the genome information net only to escape into deeper subtlety. One of the early shocks of the genome project was that the perceived morphological and behavioural complexity of higher organisms is not reflected in the number of genes in their DNA, measured in terms of coding sequences for proteins. Humans, mice, Drosophila and higher plants don't differ nearly as much as had been believed: the expectation of something like 100,000 genes for humans has been scaled down to about 33,000, while Arabidpsis may have as many as 27,000. In fact, the protein-coding sequences account for somewhat less than 2% of the human genome,

while 97-98% of its transcriptional readout is estimated to be non-coding RNA. Inevitably the focus now shifts from number of protein-coding sequences to how they are translated and how their products interact during development so as to make organisms of distinct morphology and behaviour.

The view that is emerging is that organisms have much more sophisticated ways of reading and editing their DNA than was previously believed, so that what was believed to be a single coding sequence is in fact many. For example, in the hair cells of the inner ear of the chick there is a gene that can be translated into 576 different proteins, each one altering the tuning of cells to sound frequencies (Black1998). The single sequence of bases in the DNA that contains the information for making these proteins is read in many different ways by alternative splicing, the different versions of the protein being produced in hair cells in different positions of the inner ear, in systematic spatial order. In the fruit fly, it has been estimated that the number of different messages that could arise from a single 'gene' sequence is 38,016! Furthermore, the non-coding regions of the DNA, originally described as 'junk', turn out to be major players in the selection, modification, and editing of the DNA during development and throughout the life of the organism.

Reflecting on such questions in her thoughtful book, 'The Century of the Gene', Evelyn Fox Keller (2000) observed: "What is most impressive to me is not so much the ways in which the genome project has fulfilled our expectations but the ways in which it has transformed them... Contrary to all expectations, instead of lending support to the familiar notions of genetic determinism that have acquired so powerful a grip on the popular imagination, these successes pose critical challenges to such notions. Today, the prominence of genes in both the general media and the scientific press suggests that in this new science of genomics, twentieth century genetics has achieved its apotheosis. Yet, the very successes that have so stirred our imagination have also radically undermined their core driving concept, the concept of the gene. As the human genome project nears the realisation of its goals, biologists have begun to recognise that those goals represent not an end but the beginning of a new era in biology".

The new era is shifting its attention from the genome, the genetic text of a species, to the context that makes sense of this text, the sophisticated processes that use information networks to construct an organism of a specific type with a morphology and a behaviour that allow it to survive and reproduce in a particular habitat. This is of course the evo-devo project, the union of developmental biology with evolutionary principles to understand the whole organism as a life-cycle. The organised agency that is the living organism makes sense or meaning of its genetic text by making itself in a manner that reveals context sensitivity within the developing organism and is also responsive to its external context, its habitat. What is the nature of this agency, and what concepts are going to be useful in transcending the information and the molecular machine metaphors, whose limitations are now only too apparent? I suggest that we now extend and transform these metaphors by introducing the concepts of coherence and meaning into the evo-devo dialogue.

2 Biology and Meaning

The gap between information and action always involves an agent that makes sense of the information within a context. What kind of agent is a living organism? Organisms survive and flourish in particular habitats by taking discriminating actions appropriate to circumstance, actions that matter to them for their well-being and the perpetuation of their kind. Discriminating and mattering are the components of meaning as expressed in language (cf Caws, 1988). Meaning is a high-level concept not normally used in biology, though I think its time has arrived. However, to inform detailed research it needs to be explored and articulated, just as the information and computer concepts had to be developed in connection with molecular processes.

To say that organisms take discriminating actions that matter to them for their survival is common sense. How does this translate into molecular activities within cells? One significant aspect of post-genomic biology is the recognition that the sophisticated patterns of molecular interactions whereby genetic information is read and instructions are executed is through complex networks. Genomes and proteomes are now being modelled as self-referential networks, a description that beckons in a new direction. As stated by Jeong et al (2001) in connection with their study of the structure of the protein interaction network in yeast: "Proteins are traditionally identified on the basis of their individual actions as catalysts, signalling molecules, or building blocks in cells and microorganisms. But our post-genomic view is expanding the protein's role into an element in a network of protein-protein interactions as well, in which is has a contextual or cellular function within functional macromolecules". These networks have generic properties that identify them as self-organised systems defined by power laws. If we ask what is the probability that a particular protein in the network interacts with k other proteins, the answer is $P(k) = ak^{-\alpha}$ where a and α are constants (Jeong et al, 2001). It has been known for many years that such relationships are also characteristic of written texts. If we rank-order the frequencies of words in any text and calculate the frequency is of the r^{th} most frequent word, then we find that it is $f(r) = br^{-\beta}$ where again b and β are constants (Zipf's law). Organisms appear to be using forms of language in their discriminating functional activities, which have meaning in relation to their goals. Here meaning depends on the embodied or tacit knowledge and information used by organisms, forms of knowledge now recognised by cognitive scientists as primary also in human culture (Varela et al, 1997; see also Maturana and Varela, 1987, 1998)

Power laws are turning up in many different contexts, all of which appear to reflect self-organising properties in complex systems. In a study of the properties of heart-beat intervals determined from electrocardiograms, it has been discovered that the lengths of these intervals in healthy individuals obeys a power law relationship: the frequency of intervals of duration d is $f(d) = cd^{-\gamma}$. In individuals with abnormalities of heart function, this relationship does not hold. Here is a case where health or well-being correlates with a distinctive dynamic pattern that describes long-range order: the dynamic structure of the interbeat interval distribution means that a series of shorter durations (faster heart rate) will be followed by longer duration (slower beats), in such a way that the overall pattern maintains a balanced or harmonious quality that is characterised by self-similarity of the dynamic pattern (Peng et al, 1995; Kalon et al, 1997; Solé and Goodwin, 2000). This results in a robust dynamic structure. The fractal distribution of heart-beat intervals describes a chaotic element in heart rate dynamics that is interpreted to result in responsive sensitivity of the heart to whatever demands may be made on it within its unpredictable context, the whole body with its continuously changing adaptive response to circumstances. The ECG of an individual with disturbed heart function, such as congestive heart failure or a heart arrhythmia, or even a respiratory disorder such as apnoea, is dynamically more ordered than that of a healthy individual (Poon and Merrill, 1997; Ivanov et al, 1996). The principle that emerges here is that too much order, or order of the wrong kind that is not responsive to circumstance, is a sign of danger. This also conforms to common sense.

Fractal structures are also widely encountered in biological form, as in the branching patterns of trees, the venation patterns of leaves, and the structure of respiratory or circulatory systems. Here the principle can be connected with efficient flow of fluids, compressible or incompressible, throughout the living system, as demonstrated by West et al (1997). Another domain of fractal order in connectivity patterns is in metabolic networks, often taken to be the very essence of biological process. Morowitz (1992) proposed that the pattern of interactions of metabolites in organisms reflects the natural interactive order of chemicals, an hypothesis confirmed by Fell and Wagner (2000), Wagner and Fell (2001) in a study of bacterial metabolism. These networks have evolved their complexity through gradual addition of components to core reaction pathways. It

appears that they have self-organising properties that are naturally fractal and are not a result of natural selection. The agency that is expressed in organisms through their sophisticated contextual networks of molecular interactions is robust and evolutionarily adaptable, but not designed simply for survival. There seems to be a deeper principle at work that has survival and adaptability as a consequence, not a primary, imposed design feature. Natural selection looks to be the fine tuner, not the initial generator of the organisational principles of living organisms, which reflect principles of order that extend beyond the living state to its physical and chemical substratum..

3 Wholeness as Coherence.

Power laws are characteristic of any process in nature in which order is emerging. The most familiar of these are phase transitions such as the condensation of a gas to a liquid, of a liquid to a solid, or the appearance of magnetism in a ferromagnet as it is cooled below its transition temperature.

Steam undergoes a transition to liquid water through a process in which the water molecules begin to stick together in clusters of different size. At the critical temperature of 100^oC these clusters have a size distribution that obey a power law relationship: the size of a cluster (its correlation length) behaves according to the relationship $\xi = At^{-\nu}$, where $t = (T - Tc)/Tc$, Tc being the critical temperature while A and ν are constants. This means that there are many small clusters, a very few very large ones, and clusters of all sizes in between that define a straight line when $log\xi$ is plotted against log t on a graph. As the critical temperature is approached from above, clusters grow very rapidly and some extend over the full size of the container. It is as if the system is exploring all possible states available to it, including becoming a new whole that is the liquid form.

The situation regarding the appearance of a magnetic field in a ferromagnet seems to have nothing to do with the condensation of a gas. However, it turns out that these two very different states of matter behave in not just similar but identical ways. As the critical temperature of 770^o is approached from above, the molecular magnets of which the metal consists begin to align with one another, forming small magnetic fields whose length grows just as do the clusters of molecules in the gas. At the critical temperature these fields have a size distribution that behaves in exactly the same way as the condensing gas, following a power law of size distribution. However, the ferromagnet has an additional degree of freedom compared with a gas, which can only condense to a liquid. The magnetic field can have its north-south polarity oriented either one way, or the opposite way, along the metal. Which orientation develops depends upon chance, or upon external magnetic influences if there are any. But whichever polarity emerges, it does so according to the same universal principles as any phase transition, in any physical system.

These surprising discoveries of universality in critical point phenomena were the result of intensive research during the 60s, 70s and 80s, when it gradually dawned on physicists that something highly significant is at work when new order comes into being. In studying the transitions from gas to liquid in substances as different as oxygen, neon and carbon monoxide, it emerged that the physical details about the different molecules involved were essentially irrelevant to their behaviour at the transition point: a system undergoing transition from one state of order to another looks the same at all dimensions. This is the meaning of power laws and self-similarity: no matter what length you choose as your unit of measurement, the system looks the same because the clusters of order are all self-similar. What is operating here is geometry, the emergence of order, which transcends physical detail. Even in the realm of relativistic quantum mechanics, where Einstein's revelations about space-time and quantum insights into the world of elementary creation come together in the form of guage theories, based on symmetry and geometry in quantized space-time,

phase transitions obeying power laws describe the processes that give rise to elementary particles, from photons to Higgs bosons, currently a focus of research. And recent experiments at the Brookhaven National Laboratory, where physicists are attempting to recreate the conditions of the Big Bang, there is good evidence that a quark-gluon plasma has emerged through a phase transition of atoms back to primordial matter (Brumfiel: Nature 430. 498-499, 2004).

4 Coherence and Meaning.

Quantum mechanics continues to provide other scientific disciplines with insight into subtle principles of dynamic order, and quantum coherence is one that has already been suggested as a literal aspect of biological processes that involve long-range coherence and order (Penrose, 1989; Ho,1998). Despite the absence of clear, repeatable experimental evidence for such behaviour at the level of, say, molecular dynamics in brain activity or the condition of health in organisms, it is useful to examine how such a concept could be used to explore the properties of the molecular networks that are emerging as the locus of integrated activity in organisms.

Phase transitions in quantum systems and their description takes us to another level of understanding of what happens when order comes into being. A familiar example of quantum coherence is provided by a laser, in which the photons are coherently aligned with one another regarding their phase (i.e., they are oscillating as electromagnetic waves in alignment with one another). As Roger Penrose (1989) describes it, such a state of coherence occurs " when large numbers of particles can collectively cooperate in a single quantum state..."

Quantum mechanics has a distinctive and precise way of describing this state of coherent order. We need to consider the relationships between the pattern of events at one point of space, say X_1, and a neighbouring point, X_2, Order is described over a spatial domain in terms of the behaviour in time of observable events, such as the arrival of photons recorded by instruments located at different points. The whole space-time pattern is referred to as a field, described precisely by mathematical functions of position and time in the field. When a field is coherent, there are strict relationships between events recorded at different positions in the field by what are called correlation functions. Such a function, say $G(X_1, X_2)$, gives a measure of the frequency with which events at the two points coincide. Consider first the meaning of a self-correlation function such as $G(X_1, X_1)$. Clearly events such as the arrival of photons at a point always coincide with themselves: they are self-correlated. So we would expect that $G(X_1, X_1)$ would always be greater than $G(X_1, X_2)$, unless there is a relationship between the events at the two points so that they are correlated. There is a mathematical relationship, the Schwartz inequality, that describes this precisely:

$$|G(X_1, X_1)||G(X_2, X_2)| \geq |G(X_1, X_2)|^2$$

Self-correlations of events at two points of space will always be greater than, or equal to, the square of the correlations between these events at two different points of space. In general, the inequality holds. However, in the case of coherence, equality prevails. This is a very special situation, because the events are independent of each other: that is, if you look at the pattern of arrivals of photons at two points of space, X_1 and X_2, they will appear to be uncorrelated. However, there is a subtle relationship between them such that over an extended period the events are observed to be correlated. This is a kind of quantum entanglement such that the events appear to be independent but are in fact an expression of a single quantum state, not a mixed state. The square of the correlation function for the two different points factorises precisely into the product of the two self-correlation functions, which is a sufficient condition for coherence.

Quantum mechanics reveals some strange, counterintuitive properties about possible forms of behaviour in the elementary events that make up our world. What we are seeing at the quantum level is actually order coming into being and passing away all the time, with creation and annihilation operators defining the conditions under which new particles can appear and disappear in well-defined patterns of relationship. Quantum coherence tells us about conditions of order that are subtle and intriguing. The description above is for coherence between pairs of points in a filed, but coherence can be defined for three or more points simultaneously. Then the product above involves correlation functions for many points in the spatial field, $G(X_1, X_2, X_3, \ldots X_n)$ on the right hand side, and self-correlation functions of all these points taken separately on the left. Here is a situation in which events occurring at any point of a field appear to be occurring independently of one another, but in the coherent state they are all correlated as defined by equality of the products of the self-correlations and the nth power of the cross-correlations:

$$|G(X_1, X_1)||G(X_2, X_2)||G(X_3, X_3)| \ldots |G(X_n, X_n)| = |G(X_1, X_2, X_3, \ldots X_n)|^n$$

What is being described here is a condition of coherent wholeness in a system of many components that emerges when conditions are such that the coherent state becomes the attractor of the whole. The state of coherent wholeness is an expression of meaning. All phase transitions to order can be described in these terms, whether in so-called inanimate or in animate systems. In physics these transitions are singular processes that occur under well-defined conditions. In organisms, where order is continually emerging and being dissolved in the flux of molecular interactions that maintain the organisms in a far-from-equilibrium process of dynamic order, the condition of self-similarity in networks of relationships appears to be a generic signature of the living state. This holds also for relationships between organisms, as in emergent order in social insect colonies (Solé, Miramontes and Goodwin,1993; Solé and Miramontes, 1995)), for the taxonomic relationships between taxa , or indeed for the pattern of species extinctions during evolution (Solé and Goodwin, 2000). What general conclusions can we draw from these observations that could direct research into principles of the living condition that take us beyond the information metaphor to another guiding principle based on coherence and meaning?

5 Nature and Culture

I mentioned Zipf's law for the probablility distribution of words in written texts, rank-ordered according to their frequency, which exhibits a fractal, self-similar power-law pattern. The same law holds for the frequency of cities with n inhabitants, which behaves according to the relation

$$f(n) \propto n^{-\alpha}$$

A similar relationship holds for the frequency of cities with area a, with the consequence that cities grow with a fractal, self-similar pattern. Language and cities are usually associated with culture, not nature, but the relationships being revealed between the pattern of living networks and cultural networks suggests that these are deeply related in their generative dynamics. Furthermore, Ferrer y Cancho and Solé (2001, 2003) have shown that human languages can be seen as the result of a phase transition arising from a tension between the demands of a speaker and a receiver of verbal communication in human culture, resulting in Zipf's law for word frequencies. The self-similar, self-referential properties of genetic, protein, and metabolic networks can now be seen as proto-languages whereby organisms make sense of their inheritance and their environmental contexts by generating forms (organisms of specific morphology and behaviour) that express embodied meaning in coherent wholes. The dynamics and patterns of phase transitions and coherent quantum processes may be similarly interpreted as expressions of a similar type of process. In consequence,

the distinction made by humans between nature and culture is somewhat arbitrary, as both may be understood to be the result of similar generative dynamic processes that underlie the emergence of order throughout the physical, biological and cultural domains.

This unifying insight is perhaps the most significant result of the complexity revolution that has swept through physics, biology, and the humanities in the past twenty years. What may be emerging here is the formulation of a new principle of creativity in natural/cultural processes that manifests as the generation of coherent wholes. We are familiar with this in culture since all our creative activities including science, literature, art and craft take the shape of stories and forms that seek coherence and wholeness as their signature. Nature can now be seen to do likewise. The result is that the nature/culture boundary, long cherished by humans as that which distinguishes humanity from mere animals or mechanical nature, erodes. The self-referential networks that characterise the organisation of living organisms, both internally and in their interactions with each other and with the environment, can be seen to take on the characteristics of language and culture, so that evolution is a process whose intrinsic tendency is to generate coherent wholes with meaning.

In vol. 4 of the series "Towards a Theoretical Biology" (1972), Waddington remarks in an Epilogue that a better title for the series would have been a Theory of General Biology, broadening the principles of the exploration. And his conclusion was that " The situations which arise when there is mutual interaction between the complexity-out-of-simplicity (self-assembly), and simplicity-out-of-complexity (self-organisation) processes are, I think, to be discussed most profoundly at the present time with the help of the analogy of language". This suggestion was prompted by observations about programmes, symbolism and languages at the Serbelloni meetings, particularly by Howard Pattee (1972). The issue of meaning in biology was also raised by Goodwin (1972) in the same volume. Already more than 30 years ago, before the full expression of the molecular biology and genome revolutions, and the complexity movement revealing the emergent properties of self-organising wholes, here was an anticipation of the fusion of nature and culture that is now proceeding through biological semiotics and hermeneutics (Sebeok and Umiker-Sebeok, 1992; Hoffmeyer, 1998; Markos, 2002), as well as many other similar developments. Physical, biological and cultural evolution are undergoing unification into a coherent whole in terms of new general principles whereby all natural processes generate coherent wholes of distinctive quality. In place of variational or extremum principles for natural process such as least action or energy, or maximum fitness, new formulations of emergent wholeness within physics, biology, and the humanities are under exploration. These principles remain to be clearly articulated but will help to make the necessary transition from a mechanical world-view to one with creativity and meaning, as well as intelligibility.

References

[1] Black, D.L. (1998). Splicing in the inner ear**?**: a familiar tune, but what are the instruments**?**? Neuron 20**?**: 165-168.

[2] Brumfiel,G. (2004) What's in a name? Nature 430. 498-499. Caws, P. (1988). Structuralism; The Art of the Intelligible. Humanities Press International, Inc

[3] Fell, D., and Wagner, A. (2000). The small world of metabolism. Nature Biotechnology 18, 1121-1122

[4] Ferrer i Cancho, R., and Solé, R.V. (2001). The small world of human language. Proc Roy. Soc. Lond. B 268, 2261-2266.

[5] Ferrer i Cancho, R., and Solé, R.V. (2003). east effort and the origins of scaling in human language. PNAS 100, 788-791

[6] Goodwin, B.C. (1972). Biology and meaning. In 'Towards a Theoretical Biology' 4, 259-275. Ed. C.H.Waddington. Edinburgh: Edinburgh University Press.

[7] Ho, M-W. (1998) The Rainbow and the Worm**?**; The Physics of Organisms. Singapore**?**: World Scientific Hoffmeyer, J. (1998). The unfolding semiosphere. In: Evolutionary Systems. Biological and Epistemological Perspectives on Selection and Self-Organisation. G. van de Vijver et al, eds (Dordrecht, The Netherlands: Kluwer), 281-293.

[8] Ivanov, P. Ch., Rosenblum, M.G., Peng, C-K, Mietus, J., Havlin, S., Stanley, H.E., and Goldberger, A. L. (1996). Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis. Nature 383, 323-327.

[9] Jeong, H., Mason, S.P., Barabasi, A-L., Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature 411, 41-42

[10] Kalon, K.L., Moody, G.B., Peng, C-K, Mietus, J.E., Larson, M.G., Levy, D., and Goldberger, A.L. (1997). Circulation 96, 842-848.

[11] Keller, E.F. (2000). The Century of the Gene. Harvard**?**: Harvard University Press.

[12] Markos, A. (2002). Readers of the Book of Life; Contextualising Developmental Evolutionary Biology. New York: Oxfrord University Press

[13] Maturana, H., and Varela, F. (1987, 1998). The Tree of Knowledge. Boston: Shambala.

[14] Morowitz, H. (1992) The Origin of Cellular Life. New Haven: Yale University Press Pattee, H.H. (1972). Towards a Theoretical Biology 4, 248-258. Ed. C.H.Waddington. Edinburgh: University of Edinburgh Press.

[15] Peng, C-K., Havlin, S., Stanley, H.E., and Goldberger, A.L. (1995). Quantitative analysis of heart rate variability. In J.Belair, L. Glass, U, an der Heiden, and J Milton (eds). Dynamical Disease : Mathematical Ananlysis of Human Illness. Los Angeles : Springer Verlag 82-87.

[16] Penrose, R.(1989). The Emperor's New Mind. Oxford: Oxford University Press.

[17] Poon, C-S., and Merrill, C.K. (1997). Decrease of cardiac chaos in congestive heart failure. Nature 389, 492-495.

[18] Sebeok, T.A., and Umiker-Sebeok, J., eds. (1992). Biosemiotics. The Semiotic Web 1991 (Berlin: Mouton de Gruyter)

[19] Solé, R.V, and Goodwin, B.C. (2000). Signs of Life; How Complexity Pervades Biology. New York: Basic Books.

[20] Solé, R.V., Miramontes, O., and Goodwin, B.C. (1993). Oscillations and chaos in ant societies. J. theoret. Biol. 161, 343-357.

[21] Solé, R.V. and Miramontes, O. (1995) Information at the edge of chaos in fluid neural netsworks. Physica D 80, 171-180.

[22] Varela, F.J., Thompson, E., and Rosch, E. (1997) The Embodied Mind; Cognitive Science and Human Experience. Cambridge, Mass.: The MIT Press.

[23] Wagner, A., and Fell, D. (2001) The small world inside large metabolic networks. Proc. Roy. Soc. London B 280, 1803-1810

[24] West, G., Brown, J.H., and Enquist, B.J. (1997) A general model for the origin of allometric scaling laws in biology. Science 276, 122-126

Regulation of oxidative phosphorylation. Comparing computational models of cardiac and skeletal muscle.

Bernard Korzeniewski*

*Jagiellonian University, Krakow, Poland

Abstract

The classical mechanism of the regulation of the rate of ATP supply by oxidative phosphorylation in response to varying energy demand in the cell postulated by Chance and Williams consists in the negative feedback with [ADP] (and [P_i]) as the relevant regulatory signal. The discovery of the activation of three 'key' tricarboxylic acid (TCA) cycle dehydrogenases (pyruvate dehydrogenase, isocitrate dehydrogenase, 2-oxoglutarate dehydrogenase) in vitro by calcium ions prompted several authors to postulate that also the NADH-supplying block is directly activated during transition from low work to high work. Finally, computer-aided studies using the dynamic model of oxidative phosphorylation developed by myself and co-workers suggest that all oxidative phosphorylation enzymes must be directly activated by some cytosolic factor/mechanism in parallel with the activation of ATP usage and NADH supply by calcium ions, in order to explain the changes in fluxes (oxygen uptake, ATP turnover) and metabolite concentrations ([ADP], [Pi], [PCr], [NADH], Δp) during transition from low work to high work (the so-called each-step-activation mechanism or parallel-activation mechanism).

Quantitative analysis of the available experimental data suggest that the contribution of particular mechanisms to the regulation of ATP supply by oxidative phosphorylation may be different in different muscles and various experimental conditions. The negative-feedback mechanism seems to predominate in perfused glycolytic skeletal muscle stimulated electrically. In oxidative skeletal muscle in vivo stimulated neurally the each-step-activation mechanism is likely to be the main mechanism, while the negative-feedback mechanism is only responsible for fine-tuning regulation. In intact heart in vivo the direct parallel activation of ATP usage and all steps of ATP production seems to be perfectly balanced, judging from the perfect stability of [ADP], [P_i], [PCr] and [NADH] during transition from low work to high work.

Nevertheless, the physical nature of the cytosolic factor activating particular oxidative phosphorylation complexes remains unknown, although it seems to have much to do with calcium ions. Therefore, the discovery of this factor constitutes a great challenge for the future experimental studies.

Computer model of oxidative phosphorylation

Oxidative phosphorylation in mitochondria is the main source of energy in the form of ATP in most muscles under most conditions. Therefore the quantitative description of the functioning and regulation of this system is crucial to our understanding of the bioenergetic aspect of muscle work.

The basic scheme of the enzymatic reactions involved in the oxidative ATP production has been known since Mitchell proposed his chemiosmotic theory [1]; according to this theory the key intermediate in ATP synthesis is the so-called protonmotive force, that is the electrochemical potential associated with the proton gradient across the inner mitochondrial membrane. This thermodynamic potential constitutes the link between respiratory chain complexes, which couple electron flow from NADH to oxygen with proton pumping outside mitochondria (building up the protonmotive force) and ATP synthase, which couples proton return to mitochondrial matrix (dissipating the protonmotive force) with the synthesis of ATP from ADP and inorganic phosphate (P_i). The transport of ATP, ADP and P_i across the inner mitochondrial membrane is carried out by ATP/ADP transporter and P_i transporter..

However, to understand fully the dynamic behaviour of oxidative phosphorylation in vivo we must know not only the static metabolic scheme of this process, but also the quantitative kinetic behaviour of the entire system, its control and regulation. The human brain has but limited capacity to integrate the large set of experimental data concerning the kinetic properties of particular elements of the system (enzymes, processes and so on). Therefore if a deep explicit insight into the kinetic behaviour of some complex metabolic system is to be achieved, reliable dynamic computer models are necessary.

I developed with my co-workers a dynamic model of oxidative phosphorylation in isolated muscle mitochondria and in intact skeletal muscle [3,4]. The following enzymes, processes, and metabolic blocks are explicitly taken into account within the model: substrate dehydrogenation (hydrogen supply to the respiratory chain including TCA cycle, glycolysis, glycogenolysis, glucose transport, fatty acid β-oxidation, fatty acid transport, and so on), complex I, complex III, complex IV (cytochrome c oxidase), proton leak, ATP synthase, ATP/ADP carrier, phosphate carrier, adenylate kinase, creatine kinase and ATP usage (actomyosin-ATPase, Ca^{2+}-ATPase). The dependencies of metabolite fluxes through particular elements of the system on intermediate metabolite concentrations are described by proper kinetic equations. The time variations of the metabolite concentrations are expressed in the form of a set of ordinary differential equations. The set of differential equations is integrated numerically. In each iteration step, new values of rates, concentrations and other parameters are calculated on the basis of the corresponding values from the previous step. The Gear procedure was used for numerical integration, and the simulation programs were written in the FORTRAN programming language. A complete description of the model is located on the web site www.mol.uj.edu.pl/staff/benio .

The discussed computer model has been extensively tested for a very broad set of different properties of the oxidative phosphorylation system [3-8]. Only a computer model that is very well tested by comparison with a possibly broad set of different parameter values and system properties encountered in experimental studies may be used for reliable theoretical studies on the modelled metabolic system and for predicting the kinetic behaviour of this system.

Theoretical studies

During the transition from rest to work in skeletal muscle, the increased ATP demand must be matched by an elevated ATP supply in order to prevent a rapid complete exhaustion of ATP, which would lead to termination of exercise and possibly to muscle cell death. Three main mechanisms of adjusting the rate of ATP supply by oxidative phosphorylation to the current energy demand have been proposed in the literature.

According to the first mechanism, which can be called the output-activation mechanism, only ATP usage (output of the system) is directly activated by calcium ions during elevated muscle work, and oxidative phosphorylation is activated only indirectly through negative feedback involving an increase in [ADP] (and $[P_i]$). It was originally postulated by Chance and Williams [9] that the dependence of the respiration rate (VO_2) on ADP concentration is hyperbolic (first-order at low ADP concentrations). Jeneson and co-workers [10] modified this proposal and postulated that the mechanistic VO_2/[ADP] dependence in skeletal muscle is much steeper, at least second order.

The discovery of the activation in vitro of the "key" (having the greatest control over the flux of metabolites) tricarboxylate acid (TCA) cycle dehydrogenases (pyruvate dehydrogenase, isocitrate dehydrogenase and 2-oxoglutarate dehydrogenase) by Ca^{2+} ions prompted several authors to postulate that substrate dehydrogenation (input of the system) is activated in parallel with ATP usage (input/output-activation mechanism) [11,12]. According to this proposal, however, oxidative phosphorylation is still activated only indirectly via an increase in [ADP] and/or an increase in the NADH/NAD^+ ratio.

Theoretical studies by means of the discussed computer model of oxidative phosphorylation in skeletal muscle led to the conclusion that only a direct activation by some cytosolic factor(s) (e.g., calcium ions) of all oxidative phosphorylation enzymes is able to account for large relative changes in VO_2 (respiration rate) and ATP turnover accompanied by only a very moderate relative increase in ADP concentration during rest-to-work transition in skeletal muscle (see Table 1). This mechanism has been named the each-step-activation mechanism or parallel-activation mechanism [5-8].

Table 1. Changes in [ADP] and oxygen consumption during transition from resting or a low-work state to an intensive-work state in different organisms and muscle types. In all instances the relative increase in VO_2 is much greater than the relative increase in [ADP] (the phenomenological VO_2/[ADP] is much steeper than the first order), which is in contradiction with simple Michaelis-Menten (hyperbolic) kinetics postulated by Chance and Williams for isolated mitochondria [9].

Organism/muscle	Increase in VO_2 (times)	Increase in [ADP] (times)	Literature
Dog heart	5	1 (no increase)	[15]
Rat gastrocnemius	10	2 - 2.5	[16]
Dog gastrocnemius	18	3 - 6	[16]
Human forearm flexor	>10	4	[10]
Greyhound biceps femoris	200	4 - 5	[16]
Thoroughbred leg muscle	60	~ 2	[16]
Human calf muscle	15	5	[16]
Trained rat hind-limb muscle	30	2	[17]
Insect flight muscle	600	2	[16]

The output-activation mechanism in its original form [9] assumes a hyperbolic (first-order at low [ADP]) dependence of VO_2 on [ADP], and therefore it obviously can not explain the experimental data presented in Table 1. This intuitive prediction was fully confirmed by computer modelling [13]. The very steep phenomenological VO_2/[ADP] relationship could be at least potentially accounted for by the modified output-activation mechanism [10], assuming at least second-order mechanistic VO_2/[ADP] dependence, and by the input/output activation mechanism and the each-step-activation mechanism. Therefore more experimental data and quantitative analysis were needed to distinguish between these mechanisms.

An important experimental finding is that the phenomenological VO_2/[ADP] relationship is much steeper in intact skeletal muscle than in isolated muscle mitochondria (where this relationship is hyperbolic [9]) and that the maximum oxygen consumption (normalized for the amount of mitochondrial proteins) in intact skeletal muscle is 2 to 4 times greater than in isolated mitochondria, skinned fibres, muscle homogenate or intact muscle during recovery [4]. These kinetic properties of the system can not be reconciled with both the original [9] and the modified (high-sensitivity) [10] version of the output-activation mechanism, since this mechanism assumes the same VO_2/[ADP] relationship in isolated mitochondria and intact muscle.

The $NADH/NAD^+$ ratio and the protonmotive force Δp seem to be relatively constant (or even to increase) in skeletal muscle during rest-to-work transition (compare [5,7]). This remains in contradiction not only with the output-activation mechanism, which implies a significant decrease in $NADH/NAD^+$ and Δp during rest-to-work transition, but also with the input/output-activation mechanism. In the latter situation, the quantitative analysis performed by use of the computer model of oxidative phosphorylation shows [7,13] that small relative changes in [ADP] and [NADH] can not account (through metabolite-mediated activation) for the large increase in the oxygen consumption encountered in experimental studies, and therefore the oxidative phosphorylation system must be directly activated by some factor/mechanism during rest-to-work transition.

Computer simulations also demonstrate that it is not enough to directly activate one (e.g., ATP synthase) or a few oxidative phosphorylation complexes. If a large increase in VO_2 and ATP turnover is to be caused and dramatic changes in [ADP] and [NADH] are to be avoided, it is absolutely necessary that all oxidative phosphorylation complexes be activated [13]. These simulations constitute the primary basis for the each-step-activation proposal.

Subsequently, it was demonstrated that several other properties of the system can be explained by the each-step-activation mechanism, but not by the output-activation mechanism or the input/output-activation mechanism. These properties comprise:

(1) Training-induced increase in VO_2 recalculated per mg of mitochondrial protein at a given [ADP] and an increase in the ATP/ADP homeostasis (smaller increase in [ADP] for a given increase in VO_2 during rest-to-work transition) in skeletal muscle [6,14].

(2) Asymmetry of the half-transition time $t_{1/2}$ for [PCr] between the on-transient (rest-to-work transition) and off-transient (work-to-rest transition) in skeletal muscle [6].

(3) PCr overshoot during muscle recovery [6].

(4) Differences in the kinetic properties of oxidative phosphorylation in different muscles and various experimental conditions [6].

Finally, it has been postulated that in skeletal muscle both the each-step-activation mechanism and negative feedback via [ADP] participate in the regulation of oxidative phosphorylation, while in intact heart in situ each-step activation is the only relevant mechanism [unpublished data].

References

[1] Mitchell P: Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. Nature 191: 144-148, 1961

[2] Bohnensack R: Control of energy transformation of mitochondria. Analysis by a quantitative model. Biochim Biophys Acta 634: 203-218, 1981

[3] Korzeniewski B and Mazat J-P: Theoretical studies on the control of oxidative phosphorylation in muscle mitochondria: application to mitochondrial deficiencies. Biochem J 319: 143-148, 1996

[4] Korzeniewski B and Zoladz JA: A model of oxidative phosphorylation in mammalian skeletal muscle. Biophys Chem 92: 17-34, 2001

[5] Korzeniewski B: Theoretical studies on the regulation of oxidative phosphorylation in intact tissues. Biochim Biophys Acta 1504: 31-45, 2001

[6] Korzeniewski B: Regulation of oxidative phosphorylation in different muscles and various experimental conditions. Biochem J 375: 799-804, 2003

[7] Korzeniewski B: Regulation of ATP supply in mammalian skeletal muscle during resting state → intensive work transition. Biophys Chem 83: 19-34, 2000

[8] Korzeniewski B and Zoladz JA: Factors determining the oxygen consumption rate (VO_2) on-kinetics in skeletal muscle. Biochem J 397: 703-710, 2004

[9] Chance B and Williams GR: The respiratory chain and oxidative phosphorylation. Adv Enzymol 17: 65-134, 1956

[10] Jeneson JA, Wiseman RW, Westerhoff HV and Kushmerick MJ: The signal transduction function of oxidative phosphorylation is at least second order in ADP. J Biol Chem 271: 27995-27998, 1996

[11] McCormack JG, Halestrap, AP and Denton RM: Role of calcium ions in regulation of mammalian intramitochondrial metabolism. Physiol Rev 70: 391-425, 1990

[12] Hansford RG: Control of mitochondrial substrate oxidation. Curr Top Bioenerg 10: 217-277, 1980

[13] Korzeniewski B: Regulation of ATP supply during muscle contraction: theoretical studies. Biochem. J. 330: 1189-1195, 1998

[14] Korzeniewski B and Zoladz JA: Training-induced adaptation of oxidative phosphorylation in skeletal muscle. Biochem J 374: 37-40, 2003

[15] Katz LA, Swain JA, Portman MA and Balaban RS: Relation between phosphate metabolites and oxygen consumption of heart in vivo. Am J Physiol 256: H265-H274, 1989

[16] Hochachka, PW: Muscles as Metabolic Machines. CRC Press, Boca Raton, 1994

[17] Dudley GA, Tullson PC and Terjung RL: Influence of mitochondrial content on the sensitivity of respiratory control. J Biol Chem 262: 9109-9114, 1987

Lindenmayer vs. D'Arcy Thompson: A comparison of two approaches for modeling plant development and structure

P Prusinkiewicz, P Federl, R Karwowski, B Lane, A. Runions & C. Smith

Department of Computer Science, University of Calgary, Calgary, AB T2N 1N4, Canada

Abstract

A commonly used approach for modeling plants is to view them as assemblies of semi-autonomous discrete modules. As exemplified by Lindenmayer systems, models and simulations are used to explain the behavior of larger systems in terms of their constitutive components.

A contrasting approach has its origins in the theory of transformations, proposed by D'Arcy Thompson and extended to developmental processes by Richards and Riley. The growth of an organism, organ, or tissue is described in continuous terms, as a deformation of space. This technique was originally used to explain changes in proportions during growth. Here we discuss how the theory of transformations can be extended to patterning and differentiation.

1 Introduction: Bottom-up vs. top-down approaches to the modeling of development

Developmental plant models are commonly constructed using the bottom-up approach. Philosophically, this amounts to a reductionist explanation of a system in terms of its components. Thus, for example, the development of a tissue is viewed as a result of the division and elongation of individual cells, the development of a plant is described in terms of the behavior of architectural plant modules (e.g., apices, internodes, leaves, and flowers), and properties of an ecosystem are derived from the behavior of individual plants. The bottom-up approach is the cornerstone of the individual-based methods in ecology (*c.f.* [3]) and the view of a plant as a population of semi-autonomous units, [Harper and White], among others. Its applications to modeling are best exemplified by the notion of L-systems [7,24], and have been described in detail in the literature (*e.g.*, [14,16]).

A contrasting approach has its origins in the theory of transformations, which was proposed by D'Arcy Thompson [23] to capture the dependencies between the form of related species, and extended by Richards and Riley [18] (see also [17]) to capture the relations between developmental stages of the same organism. According to this approach, the growth of a tissue, organ, or organism is described in continuous terms, as a deformation of space (represented by a coordinate system), in which these structures are embedded. The transformation approach characterizes development in a top-down, holistic manner, in the sense that the changes in the form of the system components are derived from the growth of the system as a whole. It was originally applied to describe changes of proportions between components of an organism during growth (allometry) (Figure 1).

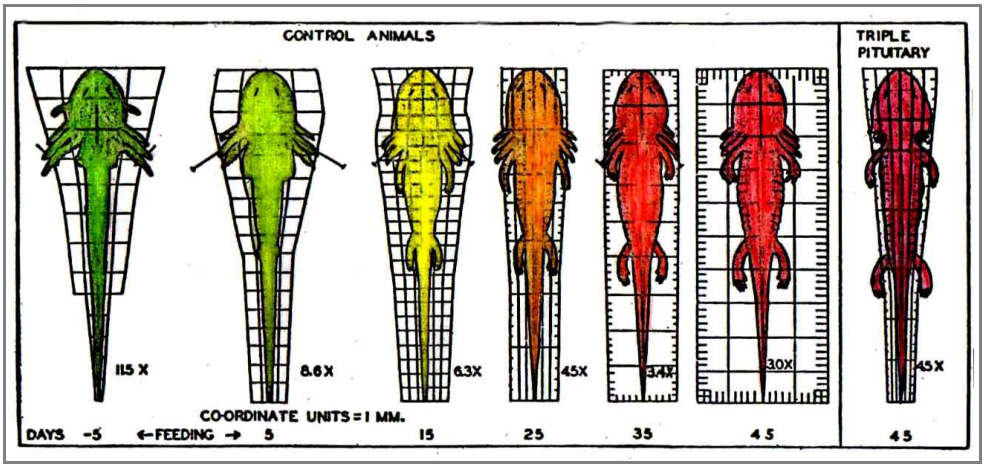


Figure 1. Dorsal views of growing *Amblystoma punctatum* (salamander) larvae related to the transformations of the coordinate system. Adapted from [17].

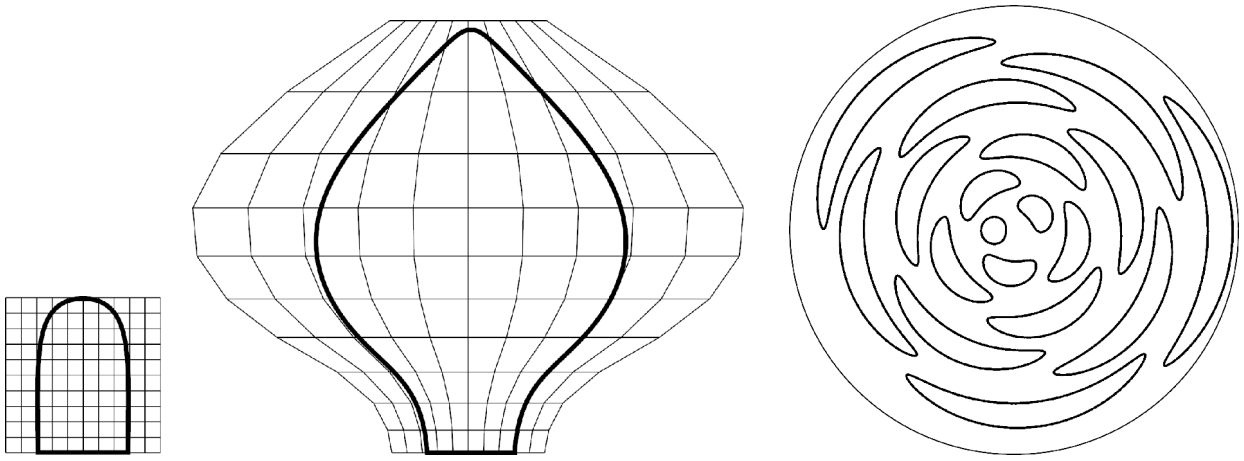


Figure 2. Deformations of organs due to non-linear growth. Left and center: two stages of a simulation of leaf growth. Right: A snapshot of a simulation of a phyllotactic pattern formation on a growing circular apex. Radial growth is assumed to occur near the center of the receptacle. Circular primordia are formed near the center of the receptacle and more away with the apical tissue; this causes their gradual deformation from circular to crescent shape. Such a deformation is often observed in transversal sections of real apices.

In its original form, the theory of transformations characterizes the changes of size and shape of organs that are already present in the structure (Figures 1 and 2). During development, however, we also deal with the emergence of new organs or structures, absent in earlier developmental stages. For example, the hind legs of the salamander larva shown in Figure 1 emerge relatively late, and are created in a process of morphogenesis, as opposed to an enlargement of preformed components. Similarly, in the phyllotactic pattern formation illustrated in Figure 2 (right), not only the existing primordia change shape from circular to crescent-shaped, but also new primordia are created near the center of the apex. This raises the question of whether the transformation approach can be extended to morphogenesis, which may include dynamic modifications of the type, number, and topological configuration of system components.

2 Introducing morphogenesis into the theory of transformations: the artists-on-canvas metaphor

At a conceptual level, the notion of expanding space and morphogenesis have been united in Coen's metaphor of an organism as a canvas operated upon by a population of artists [1]. The artists, representing local morphogenetic processes, modify the structure of the organism according to their own history (lineage information) and interaction with the neighbors (signaling). Growth of an organism corresponds to the expansion of canvas and entails an increase in the number of artists; consequently, the number of artists per unit area of the canvas remains approximately constant. The artists-on-canvas metaphor does not specify whether the increase in the number of artists is the cause or an effect of the increase of the canvas size; the first interpretation is consistent with the bottom-up view of model construction, while the second interpretation is consistent with the top-down view. The second interpretation, however, suggests how the transformation approach can be extended to account for morphogenesis.

The basic principle is illustrated in Figure 3, which shows the development of a file of cells embedded in a non-uniformly growing tissue. The notion of "embedding" implies here that the cells do not move with respect to each other and the tissue as a whole. The only changes in cell configuration are due to cell division (and, possibly, death), as is the case in the development of plants. In this example, as the file becomes longer, the individual cells reach a threshold length, which leads to a pattern of cell division.

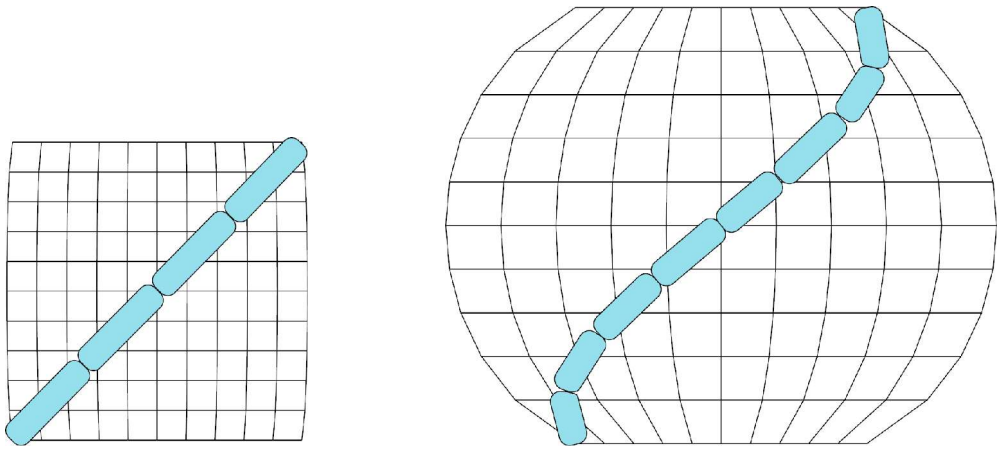


Figure 3. Simulation of the development of a cell file embedded in a non-uniformly growing tissue. Cells divide upon reaching the threshold length.

A slightly more involved example of morphogenesis in a growing surface is shown in Figure 4 (bottom). A branching structure is created by subdividing each segment that has reached a threshold length into three sub-segments with the associated lateral branches. The growth of the segments is a result of the expansion of the surface that includes them.

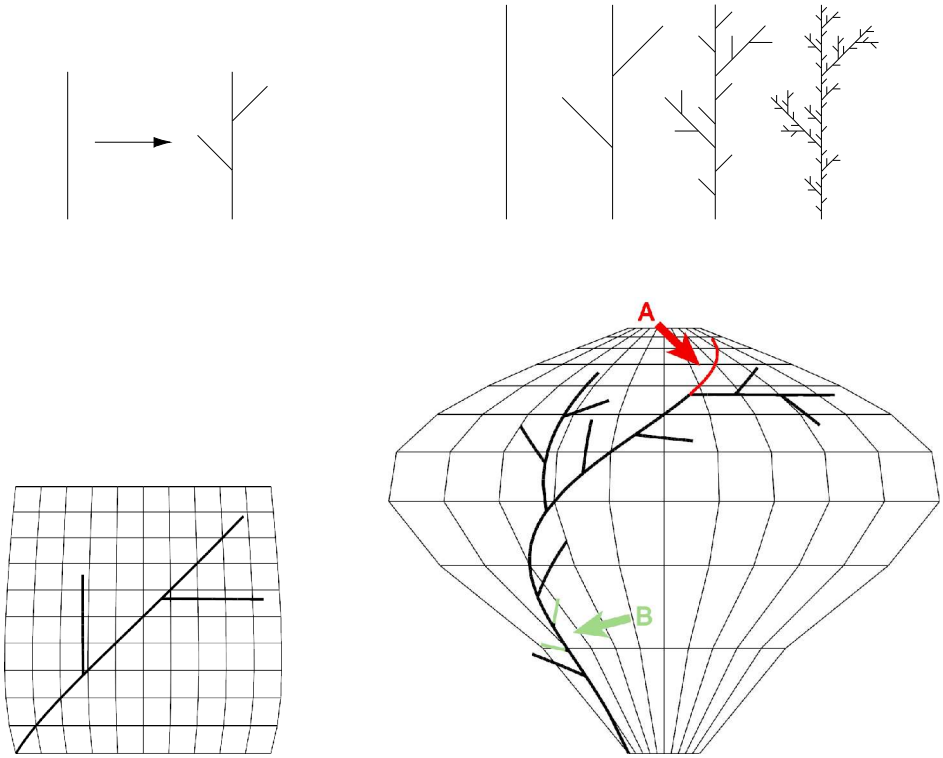


Figure 4. A comparison of two approaches to the simulation of branching structure development. Top left: the developmental rule replaces a single segment by a branching structure of five segments. Top right: The initial stages of the developmental sequence generated by an L-system using this rule (*c.f.* [14]). Following the definition of L-systems, production application is driven by the progress of time, and all segments are subdivided synchronously in each derivation step. Bottom: Two stages of a simulation of the development of a related structure in an expanding surface. The production is applied when a segment reaches the threshold length. Due to the non-uniform expansion of the surface, the segments are subdivided asynchronously. For example, the top segment (A) has not divided at all between the two stages shown, whereas a part of the bottom segment underwent two-fold division (B).

3 Examples

Spacing of stomata and trichomes in a leaf is an example of a process in which distances are measured on the growing surface, rather than just along lines [10,21]. An approximately constant spacing between elements can be maintained if new elements are inserted between the existing ones as the surface grows (Figure 5).

An extension of this process has recently been proposed as a mechanism of vein pattern formation [20]. In this case, the expansion of leaf blade is supposed to create room for the sources of a plant hormone auxin, which promotes the growth of veins towards these sources. Once formed in an acropetal direction (from the existing veins towards the sources of auxin), the veins elongate symplastically within the expanding leaf blades, in a manner similar to that shown in Figure 3.

A further, well studied example of morphogenesis on expanding surfaces is due to Nakielski and Hejnowicz, who modeled patterns of cell division driven by the expansion of tissues [12,13]. A simulation based on their work is shown in Figure 6. A yet another example is the process of bark pattern formation driven be the expansion of cambium (Figure 7).

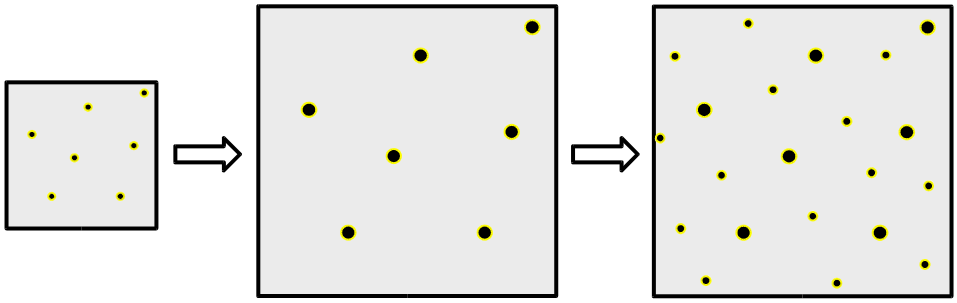


Figure 5. Illustration of a mechanism for maintaining an approximately constant spacing between organs embedded in a growing surface. As the distance between organs increases, new organs emerge between the existing ones.

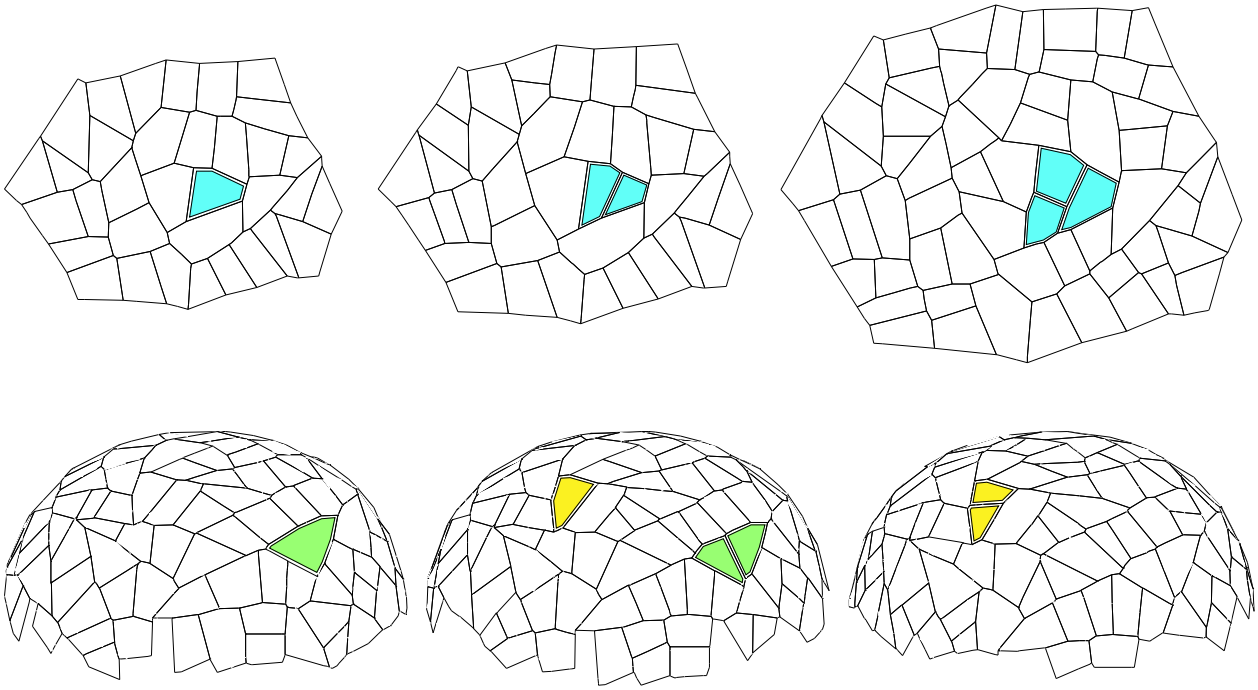


Figure 6. Simulation of cell division driven by the expansion of a planar (top) and spherical (bottom) tissue layer according to the method of Nakielski and Hejnowicz [12,13]. Cells divide upon reaching a threshold size. Shading indicates cell lineages.

4 Further aspects of morphogenesis in expanding space

4.1 Growth vs. change of space metric

In all the above examples, morphogenesis is driven by the expansion of space, as measured with a given unit of distance. A dual point of view is to assume that the it is the unit of distance (or, more generally, the metric properties of space), rather than the shape of surface, that changes. In the context of morphogenesis, the duality of these points of view has been observed by Coen [1], and is well illustrated by the model of spiral phyllotaxis proposed by Ridley [19]. According to this model, distribution of primordia on the surface of an apex or receptacle is controlled by a density function: the number of primordia per unit area. The total number of primordia can thus

Figure 7. Simulation of bark formation. The inner layer represents the interface between cambium and bark, the outer layer represents bark. Expansion of cambium causes cracks in the initially smooth bark surface (left). Depending on the mechanical properties of bark, different bark patterns are obtained (center and right).

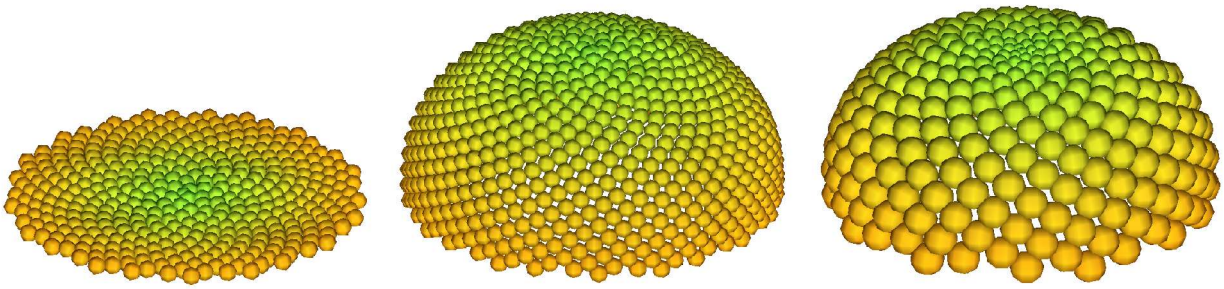


Figure 8. Simulation of pattern formation according to the model of Ridley [19]. The number of primordia depends both on the size of the receptacle and its metric properties (the size and density of placement of primordia).

be increased either by increasing the size of the receptacle, or by decreasing the unit of distance used when measuring the area (Figure 8). The latter point of view has proven particularly useful in interactive exploration of patterns [15].

4.2 Time progress vs. space expansion as factors driving morphogenesis

An interesting aspect of the top-down approach is its reliance on the expansion of space, rather then progress of time, as the factor that drives morphogenesis. This is illustrated in Figure 4, which contrasts the two approaches using the formation of a fractal branching structure as an example. The question of whether the factor that drives morphogenesis is the progress of time or the expansion of space is biologically significant. For example, the key difference between the models of phyllotaxis proposed by Hofmeister [8] and Snow and Snow [22] (see also [4]) is that in the first model new primordia are assumed to emerge at fixed time intervals, whereas in the second model they appear where there is enough space for them.

4.3 The embedding of expanding surfaces in 3D space

In the examples shown, the expanding surfaces were assumed to be flat or spherical. However, in contrast to one-dimensional objects, where growth rates and curvature are independent of each other, the curvature of surfaces is directly affected by growth. Growth that is locally faster in an area than in its surroundings leads to an "excess of surface", which increases Gaussian curvature of the object (c.f. [9]). Analogously, growth that is locally slower leads to a decrease of Gaussian curvature. These phenomena have a profound impact on the shape of surfaces, since positive Gaussian curvature implies cup-like shapes, whereas negative curvature yields undulations. The curvature, in turn, has a profound effect on the simulation of morphogenesis, because fundamental notions such as line and distance between points must be defined and computed taking the shape of the underlying surface into account [2].

5 Conclusions

The bottom-up and top-down approaches can often be applied to describe the same developmental process (Figure 9). The question nevertheless arises of which point of view is preferable in specific situations.

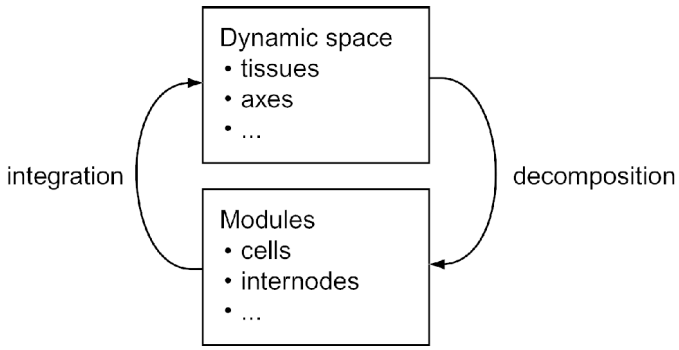


Figure 9. The relationship between the bottom-up and top-down approaches. In the bottom-up approach, exemplified by the work of Lindenmayer, a higher-level structure emerges as a result of the development of individual components (modules). In the top-down approach, derived from the work of D'Arcy Thompson, the behavior of the modules results from the development of the structure as a whole.

It appears that the top-down approach, derived from D'Arcy Thompson's ideas, is particularly convenient in the presence of long-distance coordination of growth, at scales larger than the size of the individual components of the system. The large-scale trends coordinating the growth can be described in continuous terms, such as the growth and curvature tensors. The mechanisms that provide this coordination may be external to the described phenomenon, as exemplified by the model of bark pattern formation driven by the expansion of cambium, or may result from the interaction between the elements of the system itself. In the latter case, the top-down approach is a convenient abstraction, in which the spatial coordination can be considered without delving into the mechanisms that cause it. In contrast, the bottom-up approach offers a more convenient conceptual framework if the mechanism of coordination of the behavior of individual components is investigated, or if the such coordination is lacking at the scale of observed events, rendering continuous notions of growth and curvature tensors less useful.

Although many examples of the simulation of morphogenesis based on the notion of expanding space already exist, development of a generic software environment facilitating such simulations – a top-down counterpart of simulation environments based on L-systems, for example [16] – remains

an interesting area of further research, combining mathematics, physics, biology, and computer science. Interesting topics include:

- mathematical and computational aspects of performing geometric operations in growing, curved spaces (e.g., efficient methods for computing geodesic lines: the counterpart of straight lines in the Euclidean space);
- support for expressing morphogenetic processes in chemical, rather than geometric terms (e.g., simulation of reaction-diffusion in expanding spaces, and the feedback between concentration of substances and the rates of growth);
- effective computational methods for embedding curving surfaces in 3D space (as pointed out by Koenderink [9], these methods are usually related to a description of surfaces in physical terms);
- general methods and user interface for specifying arbitrary growth of space;
- a programming interface (language) tailored to the specification of morphogenetic processes in growing, curved spaces.

References

[1] Coen, E. *The Art of Genes: How Organisms Make Themselves*. Oxford University Press, Oxford, 1999.

[2] Coen, E., Rolland-Lagan, A.G., Matthews, M., Bangham, A. and Prusinkiewicz, P. The genetics of geometry. *PNAS* **101** (2004): 4728-4735.

[3] Deussen, O., Hanrahan P., Lintermann B., Mech, R., Pharr M. and Prusinkiewicz, P. Realistic modeling and rendering of plant ecosystems. Proceedings of SIGGRAPH 1998, ACM SIGGRAPH, New York, 1998, pp. 275-286.

[4] Douady, S. and Couder, Y. Phyllotaxis as a dynamical self organizing process, Parts I-III. *Journal of Theoretical Biology* **178** (1996): 255-312.

[5] Federl, P. Modeling fracture formation on growing surfaces. Ph.D. Thesis, Department of Computer Science, University of Calgary, 2002.

[6] Federl, P. and Prusinkiewicz, P. Finite element model of fracture formation on growing surfaces. *Lecture Notes in Computer Science* **3037** (Proceedings of ICCS 2004), Springer, Berlin, 2004, pp 138-145.

[7] Harper, J.L. *Population Biology of Plants*. Academic Press, London, 1977.

[8] Hofmeister, W. Allgemeine Morphologie der Gewachse. In: *Handbuch der Physiologischen Botanik*, 1, Engelmann, Leipzig, 1868, pp. 405-664.

[9] Koenderink, J. J. *Solid Shape*. MIT Press, Cambridge, MA, 1993.

[10] Korn, R.W. Heterogeneous growth of plant tissues. *Botanical Journal of the Linnean Society* **112** (1993): 351-371.

[11] Lindenmayer, A. Mathematical models for cellular interaction in development, Parts I and II. *Journal of Theoretical Biology* **18** (1968): 280-315.

[12] Nakielski, J. Tensorial model for growth and cell division in the shoot apex. In: A. Carbone, M. Gromov and P. Prusinkiewicz (Eds.), *Pattern Formation in Biology, Vision, and Dynamics*. World Scientific, Singapore, 2000, pp. 252-267.

[13] Nakielski, J. and Hejnowicz, Z. The description of growth of plant organs: A continuous approach based on the growth tensor. In: J. Nation, I. Trofimova, J. D. Rand and W. Sulis (Eds.), *Formal Descriptions of Developing Systems.* NATO Science Series II: Mathematics, Physics and Chemistry, Vol. 121 Kluwer, Dordrecht, 2003, pp. 119-136.

[14] Prusinkiewicz, P. and Lindenmayer, A. *The Algorithmic Beauty of Plants*. Springer-Verlag, New York, 1990. With J.S. Hanan, F.D. Fracchia, D.R. Fowler, M.J.M. de Boer, and L. Mercer.

[15] Prusinkiewicz P., Muendermann, L., Karwowski, R., and Lane, B. The use of positional information in the modeling of plants. Proceedings of SIGGRAPH 2001, ACM SIGGRAPH, New York, 2001, pp. 289-300.

[16] Prusinkiewicz, P. Art and science for life: Designing and growing virtual plants with L-systems. *Acta Horticulturae* **630** (2004): 15-28.

[17] Richards, O.W. and Kavanagh, A.J. The analysis of growing form. In: W. E. Le Gros Clark and P. B. Medawar (Eds.), *Essays on Growth and Form presented to d'Arcy Wentworth Thompson*. Clarendon Press, Oxford, 1945, pp. 188-230.

[18] Richards, O.W. and Riley, G.A. *J. exper. Zool*. **77** (1937): 159.

[19] Ridley, J. N. Ideal phyllotaxis on general surfaces of revolution. *Mathematical Biosciences* **79** (1986): 1-24.

[20] Runions, A., Fuhrer, M., Lane, B., Federl, P., Rolland-Lagan, A., and Prusinkiewicz, P. Modeling and visualization of leaf venation patterns. Manuscript, University of Calgary, 2005.

[21] Sachs, T. *Pattern Formation in Plant Tissues*. Cambridge University Press, Cambridge, 1991.

[22] Snow, M. and Snow, R. A Theory of the regulation of phyllotaxis based on *Lupinus albus*. *Phil. Trans. Roy. Soc. London* **244B** (1962) : 483-513.

[23] Thompson, d'Arcy. *On Growth and Form*. Cambridge University Press, Cambridge, 1952.

[24] White, J. The plant as a metapopulation. Annual *Review of Ecology and Systematics* **10** (1979): 109-145.

Role of the p53 tumour suppressor in migration and invasion

Pierre Roux*

*CRBM, FRE CNRS2593, 1919, route de Mende 34293 Montpellier cedex

Introduction

One of the most critical modifications during tumorigenesis is the conversion from a static primary tumor to an invasive, disseminating metastasis. During this switch, tumor cells, whose proliferating properties have already been largely affected, complete their transformation by increasing their capacity to migrate. This requires remodelling of the actin cytoskeleton to modify cell shape and dynamic interactions with the surrounding extracellular matrix (ECM). Numerous studies have defined the Rho family of small GTPases as key regulators of actin cytoskeleton reorganization and adhesion mechanisms. Three main classes of Rho GTPases, Rac, Cdc42 and Rho, control cell protusions during migration. Rac promotes the formation of large membrane extensions, at the leading edge of the cell called lamelipodia. These are associated with an elongated mode of cell motility. Cdc42 generates the formation of filopodia, presumably to sense the surrounding extracellular environment, providing spatial information to direct movement. Rho kinase (ROCK), a key effector of RhoA and RhoC (known as Rho) provoke actin stress fibers and focal adhesion formation, generating a rounded-bleb-associated mode of motility, similar to amoeboid movement [1]. Aberrant regulation of Rho proteins is associated with metastasis by promoting cell motility [2][3][4][5]. Despite this crucial role, genes encoding RhoGTPases have almost never been found mutated in human cancers, where it seems likely that only their functional activities are deregulated [6][7]. This suggests that genomic alterations of others genes can account for functional modifications of RhoGTPases accompanying actin cytoskeleton remodeling during the metastasic process. Tumor progression is a multistep process wherein several defined events are common to cancer cells [8]. The initiating step is first marked by genetic alterations of genes controlling cell cycle events confering neoplastic properties, leading to progression and ultimately metastasis. We hypothesize that this is a set of genes controlling cell cycle which, once mutated during the tumor initiating phase in cancer, can modify the behaviour of proteins involved in actin cytoskeleton dynamics, such as Rho GTPases, confering on affected cells a migratory and invasive phenotype.

1 p53 and migration

Beyond theses genes appear the tumor suppressor p53, mutations of which occur in more than 50% of human tumors [9] essentially in late tumor stages [10][11][12]. p53 protects cells from malignant transformation by regulating cell cycle arrest or by promoting apoptosis [13][14]. The p53 tumor suppressor protein is among the most frequent targets for inactivation in human cancers [9][14]. Functional inactivation of wild type p53 is generally considered to favor the emergence of abnormal cells with deregulated growth that give rise to malignant transformation. p53 maintains genomic stability by regulating responses to DNA damage and other forms of genotoxic insults. Both protein accumulation and post-translational modifications appear to be essential to support the full tumour suppressor function of p53 [13]. The predominant biochemical activity of p53 relies on the transactivation and transcriptional repression of specific target genes. This appears to be the major mechanism by which p53 implements its antiproliferative activities: cell-cycle arrest in G_1/S and G_2/M phases and induction of apoptotic cell death, mechanisms that underpin its role

of suppressor of oncogenic transformation.However, we have recently shown that the contribution of p53 to the control of tumorigenesis is not restricted to its anti-proliferative activities, but is extended to the modulation of cell migration: p53 negatively modulates Rho GTPases functions and regulates cell polarisation and migration[15][16][17]. Activation of functional wt p53 prevents formation of specific actin-containing structures protruding from the cell surface and characterized as filopodia. Both overexpression of p53 wt and activation of endogenous p53 counteract Cdc42-induced filopodia formation. Our data suggest that transcriptional activity of p53 is required to inhibit Cdc42-induced filopodia formation, since p53 H273 and p53 H175, two mutants deficient in the DNA binding activity were inactive in our assay. Consistent with this observation, cells lacking p53 activity (p53-/-) exhibited constitutive membrane filopodia. This behaviour has direct effects on cell functions: i/ non-adherent cells which overexpressed p53, take more time to fully spread upon reattachment onto the extracellular matrix**?** ii/ during cell migration, the reorientation of the Golgi apparatus in the direction of movement is severely decreased by p53 wt expression, thus preventing cell polarity.

2 p53 and tissue repair

Repair of tissue damage requires the recruitment of fibroblasts to sites of tissue injury, which is mediated in part by the cytokine Tumour Necrosis Factor α (TNFα). Since dynamic rearrangements of actin cytoskeleton control cell locomotion, this implicates that TNFα is a potent coordinator of cellular actin changes. We have investigated the role of TNFα in regulating the cortical actin-containing structures essential for cell locomotion called filopodia. Kinetic analysis of TNFα-treated mouse embryonic fibroblasts (MEFs) revealed a dual effect on filopodia formation: a rapid and transient induction mediated by Cdc42 GTPase that is then counteracted by a subsequent sustained inhibition requiring activation of the mitogen-activated protein kinase p38 but not Cdc42 activity. This inhibition also involves the tumour suppressor p53, since it is activated in response to TNFα following the same time course as the decrease of filopodia formation. This functional activation of p53, measured by transcription induction of its target $p21^{WAF1}(p21)$, is also associated with p38 kinase-dependent phosphorylation of p53 at serine 18. Furthermore TNFα did not inhibit filopodia formation in MEFs treated with the transcription inhibitor actinomycin D, in p53-deficient MEFs, or MEFs expressing p53 mutants H273 or H175, supporting a role for the transcription al activity of p53 in mediating TNFα-dependent filopodia inhibition. These data delineate a novel inhibitory pathway in which TNFα prevents filopodia formation and cell migration through the activation of the MAPK p38 that in turn activates p53. This shows that TNFα on its own initiates antagonistic signals that modulate events linked to cell migration.

3 p53 and cell invasion

However, little is known about the role of p53 in cells moving in a three-dimensional environment (3D) that mimics the in vivo microoenvironement found in tumor cells. Our recent data have demonstrated that the elongated, spindle-shaped fibroblastoid mode of motility can be converted to a rounded blebbing movement by abrogation of the function of the tumour suppressor p53 through a mechanism that requires the RhoA/ROCK signaling in fibroblasts moving in a 3D environment. This amoeboid mode of motility confers higher velocity and invasive properties to cells deficient for p53. This finding demonstrates that p53 controls cell locomotion, suggesting that p53 is an important determinant of the dissemination of metastatic cells.

4 Conclusion

p53 is a tumor suppressor gene that is mutated in more than 50% of human cancers [9]. Such a prevalence define p53 as an appealing target for anti-cancer therapeutic intervention. Diverse therapeutic strategies have been developped in the attempt to restore p53 wild type function to cancerous cells, including pharmacological rescue of mutant p53 function and reactivation of wt p53 [18]. To date, the action of p53 as a tumour suppressor has been attributed to its ability to inhibit cell proliferation through growth arrest, apoptosis and cell senescence [14]. Our results highlight a novel concept of the role of p53 in human cancer ; genetic alterations of p53 in advanced tumours favours motility and invasion. Thus, mutations of p53 would affect not only proliferative properties but also predispose cells to illegitimate dissemination and invasion. This extends the interest of designing anti-cancer agents aiming at restoring p53 wild type functions : such agents may combine antiproliferative and anti invasive activities.

References

[1] E. Sahai, C. J. Marshall, Nat Cell Biol 5, 711 (Aug, 2003).

[2] E. A. Clark, T. R. Golub, E. S. Lander, R. O. Hynes, Nature 406, 532 (Aug 3, 2000).

[3] E. E. Evers, R. A. van der Kammen, J. P. ten Klooster, J. G. Collard, Methods Enzymol 325, 403 (2000).

[4] P. S. Steeg, Nat Rev Cancer 3, 55 (Jan, 2003).

[5] A. B. Jaffe, A. Hall, Adv Cancer Res 84, 57 (2002).

[6] M. Nakamoto, H. Teramoto, S. Matsumoto, T. Igishi, E. Shimizu, Int J Oncol 19, 971 (Nov, 2001).

[7] S. Rihet et al., J Cancer Res Clin Oncol 127, 733 (Dec, 2001).

[8] D. Hanahan, R. A. Weinberg, Cell 100, 57 (Jan 7, 2000).

[9] M. Hollstein et al., Nucleic Acids Res 22, 3551 (1994).

[10] C. J. Kemp, L. A. Donehower, A. Bradley, A. Balmain, Cell 74, 813 (1993).

[11] T. C. Thompson et al., Oncogene 10, 869 (1995).

[12] E. R. Fearon, B. Vogelstein, Cell 61, 759 (1990).

[13] A. J. Giaccia, M. B. Kastan, Genes Dev 12, 2973 (1998).

[14] A. J. Levine, Cell 88, 323 (1997).

[15] G. Gadea et al., J Cell Sci 117, 6355 (Dec 15, 2004).

[16] G. Gadea, L. Lapasset, C. Gauthier-Rouviere, P. Roux, Embo J 21, 2373 (2002).

[17] F. Guo, Y. Gao, L. Wang, Y. Zheng, J Biol Chem 278, 14414 (Apr 18, 2003).

[18] M. V. Blagosklonny, Int J Cancer 98, 161 (Mar 10, 2002).

Robustness and cellular design principles

Jörg Stelling

Institute of Computational Science, Swiss Federal Institute of Technology (ETH) Zurich, HRS H 28, 8092 Zurich, Switzerland; joerg.stelling@inf.ethz.ch

Abstract

Robustness, the ability to maintain performance in the face of perturbations and uncertainty, is a key property of living systems. Its molecular and cellular basis, however, have only recently begun to be understood because intimate links between robustness and cellular complexity require formal methods reaching beyond intuitive reasoning. This talk will consider robustness in cell biology from a systems theory perspective, including an introduction to concepts and definitions of robustness, and to approaches for measuring robustness. Regarding applications, theoretical approaches to complex engineered systems promise the reverse engineering of general design principles because (cellular) biology and engineering employ a common set of basic mechanisms, namely redundancy, feedback control, modularity and hierarchies to ensure robust functions. Examples from the analysis of specific biological systems such as metabolism, circadian rhythms and cell cycle regulation will serve to illustrate the use of these mechanisms in different combinations. In perspective, robustness may be a key to understand cellular complexity and to elucidate general design principles of cellular networks.

1 Robustness and design of cellular networks

1.1 Robustness: Definitions and measures

The notion of robustness has recently received considerable interest in diverse fields for which the existence of complex networks is characteristic, including the internet, social networks, and biology. Not surprisingly, the term ”robustness” has been associated with different, sometimes conflicting interpretations. In general, robustness means the persistence of a system’s characteristic behavior under perturbation or conditions of uncertainty. Robustness is, hence, defined for a specific system, which, however, may have arbitrary structural and behavorial features. Note that robustness encompasses a relative, not an absolute, property of a system. No system can maintain stability for all its functions when encountering any kind of perturbation. Any operational definition of robustness, thus, requires two additional specifications: (i) which function remains unchanged against (ii) which type of disturbances or uncertainties [7, 12].

For relatively simple systems, the ’desired systems characteristics’ often amounts to a dynamical regime. Investigations of oscillators may thus focus on the persistence of a regular periodic solution. This qualitative property does not preclude quantitative changes (in period or amplitude of the oscillations) to occur [1]. While robustness directly connects to functionality, however, often function cannot easily be assigned to a particular subsystem of a cell or organism. In bacterial chemotaxis, for instance, maintaining the ability to adapt to changing nutrient concentrations, whereas adaptation times are allowed to fluctuate, is intuitively understandable [1]. As a counterexample, robust signal processing relies upon sensitive detection, amplification and decoding of input signals, and key inputs and outputs may not be evident in complex networks [4]. The claim of higher-order behavior or entire modules to be robust therefore requires careful justification. Similar considerations apply for the specification of perturbations. Cellular systems face three broad classes of uncertainties: (i) externally induced perturbations owing to variable environments, (ii)

internal perturbations such as mutations affecting kinetic properties of proteins, or the lack of components, and (iii) intrinsic noise due to low copy numbers.

The first two types of perturbations can be mapped on changes in system parameters, which encompass kinetic constants as well as information on interactions, initial concentrations, and time-varying entities such as inputs. Also structural information (existence of links or components) can by these means be encoded. Robustness measures assess the influence of changes in parameters on the observed behavior - a model of a biochemical network serves as a mapping from parameter space to behavior space. Different measures of robustness based on parametric studies have been employed so far. In particular, parameter sensitivities $\mathbf{S}(t)$ quantify the change of a system's state $\mathbf{x}(t)$ in response to a change in parameters $\mathbf{p}$:

$$\mathbf{S}(t) = \frac{\partial \mathbf{x}(t)}{\partial \mathbf{p}} \approx \frac{\Delta \mathbf{x}(t)}{\Delta \mathbf{p}} \,. \tag{1}$$

However, parameter sensitivities usually do not directly reflect robustness, since they have to be connected to a specific functionality, and they are only locally valid (with respect to parameter space). Because they can efficiently be determined computationally, the measure is, however, often used to explore model properties in parameter space. Other approaches that measure the size of parameter space, in which the desired behavior occurs [8] are limited because interactions between many parameters may yield complex synergistic or antagonistic outcomes. The analysis of robustness, thus, has to be clear about functions, disturbances, and the measures for robustness considered.

2 Mechanisms conferring robustness

In complex engineered systems as well as in biology, design for protection against deleterious disturbances mainly employs four ingredients:

Redundancy: The simplest strategy to protect against failure of a specific component is to provide for alternative ways to carry out the function the component performs (e.g. by 'genetic buffering') [5]. Either two or more duplicate genes play identical physiological roles, or groups of different genes constitute alternative pathways for achieving the required functionality. In contrast to redundant systems in engineering, however, completely identical genes that do not diverge in functionality or regulation would not survive evolution. Instead, structurally different entities perform similar functions owing to functional overlap, which lead to the suggestion to use 'degeneracy' as a more appropriate term [3]. However, the question, whether characteristics of individual components or network characteristics contribute most to genetic buffering is highly debated.

Feedback: By using the output of a function to be controlled in order to determine appropriate input signals, feedback enables a system to adjust the output by monitoring it. Negative feedback reduces the difference between actual output and a given set-point, thereby dampening noise and rejecting perturbations. Positive feedback (or autocatalysis) may maintain cellular decisions that are derived from noisy and graded input signals [4]. However, the role of feedback is ambiguous because it invariably introduces fragilities. Such fragilities include the possibility of self-sustaining and cascading failures induced by positive feedback, for instance, in uncontrolled tumor growth. If feedback contributes to robustness compared to an unregulated system, thus, depends on the circuit design, the control objective, and the type of perturbations affecting the system [2].

Modularity: Structuring complex systems into semi-autonomous functional units (modules) can reduce the risk of catastrophic failure by preventing the spread of damage in one module throughout the network [6]. However, two critical issues have to be clarified, namely to prove the existence of modularity in cellular systems, and to establish methods for the unanimous identification of modules. For instance, simple building blocks ('network motifs') can be identified that perform functions such as feedback control [9]. Modularity in biochemical networks entails two important consequences: it lets one refer to concepts in engineering design, and it may simplify modeling and abstraction of network properties [2].

Hierarchies and protocols: Protocols encompass the set of rules aiming at an efficient management of relationships between the parts (i.e. modules). They include, for instance, the organizational structures for embedding modules, and the interfaces between modules that allow for system function [2]. In particular, an efficient means for coordination is to organize a system hierarchically, namely to establish different layers of integration, which helps to reduce the costs of information transmission. For cellular networks, however, a major challenge consists in elucidating the links between different levels of organization such as metabolism and its control at the enzyme and the gene level. In particular, it is unclear, how trade-offs between, for example, robustness and efficiency are handled by means of protocols and hierarchies.

3 Systems analysis of robustness

3.1 Metabolic networks: Structural network analysis

For investigating robustness in metabolism, the relationship between metabolic network structure, functionality, robustness and cellular regulation should be investigated systematically. Such an analysis may, for instance, allow for important insight into the design principles of metabolic regulation and their role in making the system robust toward disturbances. We assumed that, as regulation implies the realization of various behaviors on the background of the metabolic network structure, it should itself be adapted to, and thus affected by, the structure. We focused on the network structure (stoichiometry) by determining and analyzing the non-decomposable pathways able to operate coherently at steady state (elementary flux modes) in the central metabolism of *E. coli* (89 metabolites, 110 reactions) [11].

In particular, we investigated the effects of deletions of single genes coding for metabolic enzymes, since for these disturbances predictions correspond well to the behavior *in vivo*. The effect of random mutations on bacterial growth was assessed by determining number and biomass yield of elementary flux modes after deletion of each single reaction. Our analysis revealed central metabolism of *E. coli* as highly robust, since mutants with significantly (by up to 90%) reduced metabolic flexibility showed a growth yield similar to wild type. Hence, robustness relies, at least in part, on pathway redundancy. This is supported by a multi-level, hierarchical organization of metabolic regulation: according to the substrate regime, transcriptional regulation provides a general set-up for metabolic efficiency and flexibility, which is further fine-tuned at the enzyme (flux) level [11].

3.2 Circadian oscillators: Small gene networks

One can hope to obtain clues on cellular design principles by formal robustness analysis of present-day control in cellular systems, especially when it includes mathematically controlled comparisons. This requires elucidating specific structural characteristics that are responsible for robust performance. Even for genetic circuits of moderate complexity, however, these features are presently unclear. Here, we show that by systematically investigating sensitivities in the parameter

space, one can derive more global properties specific for the design of a genetic circuit as opposed to a specific implementation thereof.

Our analysis focuses on the genetic oscillator responsible for generating circadian rhythms in *Drosophila* as a prototypic dynamic regulatory system. By comparing closely related mathematical models representing alternative feedback architectures, we show that the regulatory structure largely determines the trade-off between robustness and fragility. Sensitivity analysis revealed 'hidden' mechanisms of hierarchical control through general cellular components. This concurs with the theoretical insight that hierarchies might be important for achieving robustness. The complex feedback structures encountered *in vivo*, however, do not seem to enhance robustness *per se*, but to reflect a careful management of robustness that is adjusted to both the physiological function of the clock and its operating conditions. They confer robust precision and adjustability of the clock in the face of frequently occurring perturbations, while simultaneously reducing the risk of catastrophic failure in more rare scenarios [10].

3.3 Cell cycle regulation: Multi-layered networks

The large-scale model-based analysis of robustness in cellular networks currently faces the problem of lacks of adequate mathematical models. We addressed this issues for the complex system of mitotic control in budding yeast. Our analysis involved the three steps of (i) developing a detailed mathematical model based on the known regulatory interactions (represented by $\approx$ 800 chemical reactions) and a limited set of experimental data, (ii) rigorously testing the model's consistency and predictive quality through several independent methods, and (iii) assessing network robustness through simulation studies.

Deletion of many of the regulatory genes hardly affected the ability of the network to control the initiation of mitosis, and the passage through this phase of the cell cycle. For a systematic analysis of the module's robustness, series of random perturbations were applied to the model parameters. Interestingly, no gradual degradation of the ability to control mitotic processes was observed; shifts toward highly imprecise control were rarely encountered. There exist, however, few spots of high sensitivity in the model. The analysis therefore strongly supports the concept of highly optimized tolerance, i.e. of the co-existence of robustness and fragility in cellular control as opposed to uniform insensitivity of the network. Moreover, gene deletions in parallel control circuits did not change this behavior drastically. It is tempting to speculate that *in vivo*, catastrophic breakdown after accumulation of severe damages performs a physiological function (at the population level) by preventing these cells from replicating.

4 Conclusions

Robustness is thought to be a key property of living systems. Current hypotheses on a common set of mechanisms that contribute to robustness in biology and engineering can guide future investigations. The relative importance of redundancy of components vs. pathways, the role of individual feedback circuits, the modular organization of cellular networks, and the integration of cellular functionality across hierarchies are of predominant interest. In particular, the analysis of robustness at multiple levels, using a combination of methods as pursued for the three example systems described above, may ultimately lead to a deeper understanding of complex cellular networks by elucidating general design principles. Robustness of cellular systems, hence, provides us with testable hypotheses derived from top-down studies and with opportunities for a more detailed bottom-up approach. Both approaches should finally converge.

References

[1] N. Barkai and S. Leibler. Robustness in simple biochemical networks. *Nature*, 387:913–17, 1997.

[2] M.E. Csete and J.C. Doyle. Reverse engineering of biological complexity. *Science*, 295:1664–69, 2002.

[3] G.M. Edelman and J.A. Gally. Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. U.S.A.*, 98(24):13763–68, 2001.

[4] M. Freeman. Feedback control of intercellular signalling in development. *Nature*, 408:313–19, 2000.

[5] J.L. Hartman, B. Garvik, and L. Hartwell. Principles of the buffering of genetic variation. *Science*, 291:1001–04, 2001.

[6] L.H. Hartwell, J.J. Hopfield, S. Leibler, and A.W. Murray. From molecular to modular cell biology. *Nature*, 402 (Supp.):C47–C52, 1999.

[7] H. Kitano. Biological robustness. *Nat. Rev. Genetics*, 5:826–36, 2004.

[8] M. Morohashi, A.E. Winn, M.T. Borisuk, H. Bolouri, J. Doyle, and H. Kitano. Robustness as a measure of plausibility in models of biochemical networks. *J. theor. Biol.*, 216:19–30, 2002.

[9] S.S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.*, 31(1):64–68, 2002.

[10] J. Stelling, E.D. Gilles, and F.J. Doyle III. Robustness properties of circadian clock architectures. *Proc. Natl. Acad. Sci. U.S.A.*, 101:13210–15, 2004.

[11] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E.D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420:190–93, 2002.

[12] J. Stelling, U. Sauer, Z. Szallasi, F.J. Doyle III, and J. Doyle. Robustness of cellular functions. *Cell*, 118:675–85, 2004.

Modeling and simulation of the kidney

S. Randall Thomas*

*LaMI, UMR CNRS 8042, Univ. Evry-Val d'Essonne, Genopole, Evry

Abstract

Mathematical modeling has long served as a valuable tool for understanding the many integrated functions of the kidney. We will present an overview of two among the many modeling techniques that have been used at several scales, including channel and transporter kinetics, fluid and solute uptake and secretion by the various segments of the nephron, control of glomerular filtration rate by tubuloglomerular feedback, and several types of models treating the exchanges and recycling of fluid and solutes among the nephrons and blood vessels within the different regions of the renal medulla. We will also describe the Renal Physiome project, an international effort that is under active development to provide quantitative databases and modeling resources for renal physiology.

1 Introduction

Figure 1 shows a slice through a unipapillary mammalian kidney, such as that of a rat or other small rodent.

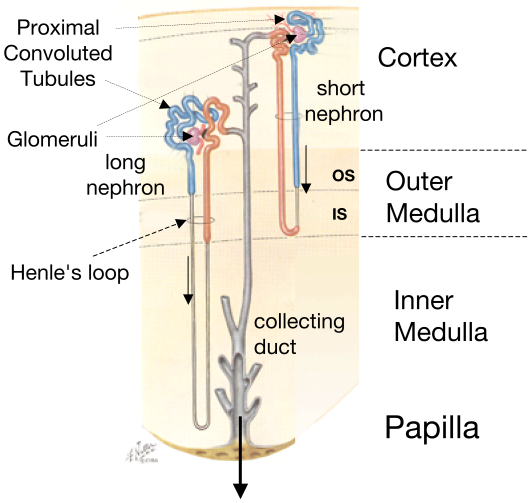


Figure 1. Short and long nephrons (Netter 1973).

Kidneys of large mammals such as humans (figure 2) have several lobes, each of which resembles this single-lobed scheme. While the kidneys of other small animals follow the same general plan as that of the rat, they differ in the size, depth, and thickness of the various regions of the kidney as well as in the number of nephrons, their distribution within the different kidney regions, the relative number of short and long loops of Henle, the length of the papilla & fraction of 'long' loops that reach the papillary tip, and in many other particulars.

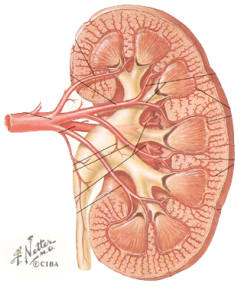


Figure 2. Human kidney (Netter 1973).

These differences generally correspond to differences of habitat (especially with respect to availability of water) and dietary habits, e.g., carnivores like cats and dogs have 100% long loops of Henle, aquatic animals like beavers have only short loops, desert rodents with a vegetarian diet usually have very long papillae and highly developed vascular bundles within the outer medulla, to name only a few of the differences. Nonetheless, despite these interspecies variations, most aspects of the main mechanisms of secretion, absorption, regulation of filtration rate, response to acid-base or hydro-osmotic perturbations are common across all mammalian kidneys. Figure 3 illustrates this variety in a series of latex casts of the vascular system from several small animals. Note the differences in their diets and the animal weights.

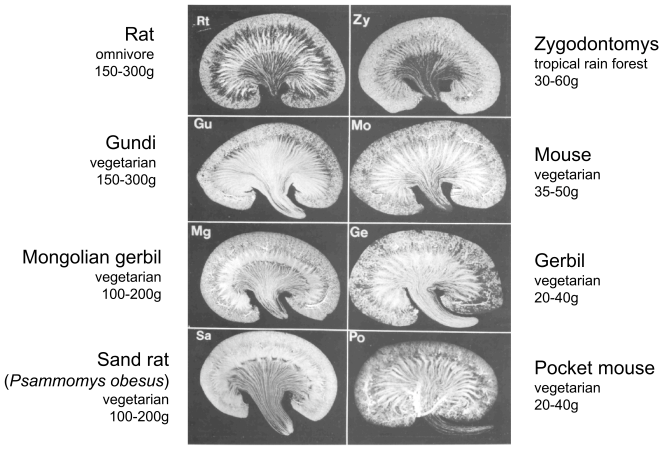


Figure 3. Latex casts of the vascular systems of several unipapillary kidneys (Bankir & de Rouffignac 1985).

Understanding what goes on inside the kidney requires, on the one hand, *in vivo* measurements of flows and concentrations in structures accessible on the surface at the cortical surface and at the papillary tip but not within the kidney itself, and on the other hand, *in vitro* techniques allow characterization of the properties of the components of the kidneyÑ solute and water permeabilities of membranes and epithelia, kinetic descriptions of transporters and channels and their immunohistochemical localization, metabolic studies of the various cell types in the different kidney regions, hormonal and other signaling relationships, both local and long-range, and so on.

However, establishing the functional links between the component characteristics (obtained *in vitro*) and the *in vivo* functioning of the organ, requires the elaboration of hypotheses that are generally too complicated to verify rigorously without recourse to mathematical modeling. hence the rich legacy of hypothesis-based models developed over the years at all levels of kidney organization.

We will present here a brief introduction to two of the modeling techniques used classically at several levels. These are based on differential equation descriptions of the phenomena of interest. This is not the only possible mathematical approach, and it may be that in the foreseeable future more recently developed techniques, such as cellular automata, multi-agent techniques, and others, may find application to problems in renal physiology. In keeping with the tutorial nature of this article, we will not spend time here on the results of the modeling studies based on these techniques; for these, the reader is referred to the original articles.

After this introduction to numerical modeling techniques, we will introduce currently active efforts to facilitate the development of new models through the creation of open databases of quantitative measurements and anatomical data, and to render the legacy of existing models accessible to the larger community via interactive repositories accessible via the worldwide web.

2 Modeling solute and water flows in epithelia

One of the goals of the Renal Physiome project is to provide simulation resources of cells and tissues. To begin to address this need, we are developing a web-based simulation resource that we call the Epithelial Toolbox. The long-term objective is to provide renal investigators a means to construct models of epithelial cell transport for use in their research. The models could be used to interpret data, design experiments, or to examine the adequacy of hypotheses, as expressed in the formalism of a model, to explain observed behaviors. The simulations could also be used for instructional purposes.

Rather than describe fullblown simulations of nephron segments, such as have been developed for questions of reabsorption and secretion along the various segments of the nephron (see review of Weinstein 1994), we will describe a general method for modeling solute and water flows through epithelia, whether they be arranged in flat sheets or in tubular structures.

The core of the epithelial cell model used in the toolbox will be based on an approach to modeling of epithelial cell transport devised twenty years ago (Latta, et al., 1984). Since then, it has mainly been used as a tool in teaching graduate students interested in modeling and/or epithelial transport. More recently, the model has been used in an ongoing research project concerning transport in the thick ascending limb (TAL). The following summarizes the most basic features of the epithelial cell models and how they will eventually be combined into segment models. Then we will outline how the simulation resource will be implemented, including interface design and expansion of the model to permit creation of custom models by users.

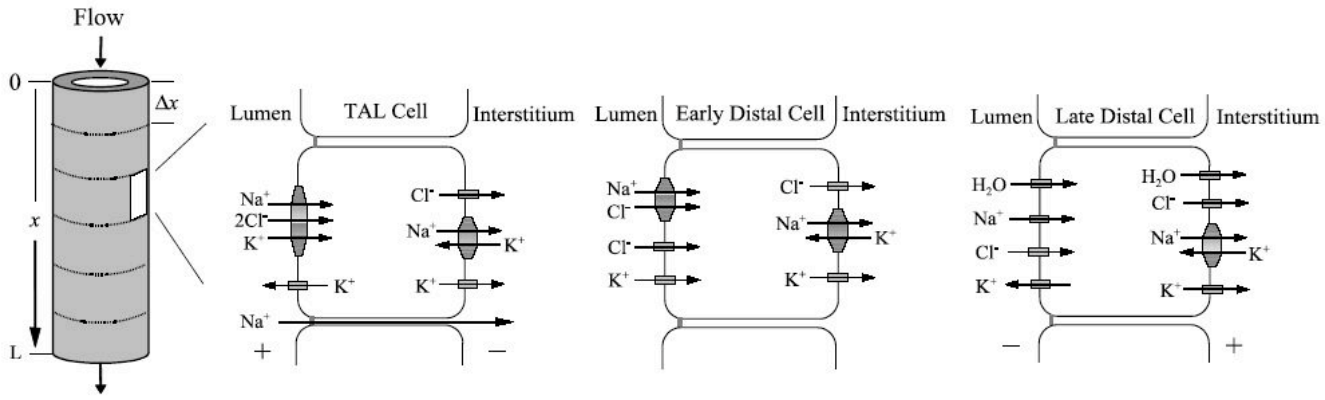


Figure 4. Illustration of elements in proposed epithelial cell and segment simulation resource.

Epithelial cell model.

The epithelial cell model is based on conservation of mass, standard kinetic descriptions of membrane transport, and constraints that ensure maintenance of electroneutrality. The basic model can be used to simulate renal epithelial cells by using appropriate parameters and adding cell-specific membrane transporters, as shown in Figure 4.

Conservation Equations

In the cellular compartment, water conservation requires that

$$\frac{dV}{dt} = A^a J_v^a + A^b J_v^b,$$

where V is cell volume, A is membrane area, the superscripts a and b refer to the apical and basolateral membranes, and J_v is water flux (influxes are positive across both membranes). For each solute:

$$V\frac{dC_i^c}{dt} = A^a J_i^a + A^b J_i^b - C_i^c\frac{dV}{dt},$$

where C_i^c is the concentration in the cellular compartment, and J_i^a and J_i^b are the solute fluxes through the apical and basolateral membranes of the i^{th} solute. Apical solute fluxes are defined as positive cell influxes; basolateral fluxes are positive effluxes.

Passive Flux Equations

Water fluxes are determined by osmotic pressure differences across the apical and basolateral membranes, e.g.,

$$J_v^k = L_p^k RT \sum_i \sigma_i^k (C_i^c - C_i^k),$$

where the range of the superscript k is (a, b, j), corresponding to the apical, basolateral, and junctional membranes, L_p is the applicable membrane hydraulic water permeability, and σ is the relevant reflection coefficient.

For all membranes, the total net flux of the i_{th} solute is the sum of the passive solute flux (described by the Goldman-Hodgkin-Katz constant-field flux equation or limiting case expressions for uncharged solutes or zero transmembrane potentials) and the flux arising from any carrier-mediated, primary or secondary active transport process, J_i^p:

$$J_i^k = P_i^k z_i U^k \frac{C_i^k - C_i^c\, e^{-z_i U^k}}{1 - e^{-z_i U^k}} + J_i^p,$$

where P_i^k are solute permeabilities and

$$U^k = \frac{E^k}{RT}.$$

Carrier-mediated Transport

In the most basic model of an epithelial cell, the only required transporter is Na,K,ATPase located on the basolateral membrane. We employ a kinetic expression for the pump current, I^p, (Lewis and Wills, 1981). A pump stoichiometry of 3/2 (n_{Na}/n_K) is assumed, and the ion fluxes are then:

$$J_{Na}^p = n_{Na} I^p/F \text{ and } J_K^p = n_K I^p/F.$$

To simulate specific epithelial cells, appropriate co-transporter models are placed on the apical and basolateral membranes. In Fortran, the model was programmed as a series of modules so that the addition of an additional carrier merely requires the addition of a subroutine to a module that computes the membrane solute fluxes. Figure 5 illustrates a rather complex epithelial cell model of ion transport in cells of the medullary thick ascending limb (MTAL) (Yang et al. 2001).

Electroneutrality Constraints

Maintenance of electroneutrality of the cell requires that the net electrical current (I) across the apical and basolateral membranes must be identical, implying that $I^a = I^b$, where

$$I^a = FA^a \sum_i z_i J_i^a \text{ and } I^b = FA^b \sum_i z_i J_i^b.$$

z_i is the valence of the i^{th} solute, and F is Faraday's constant.

Maintenance of luminal and basolateral solution electroneutrality under open-circuit conditions requires that the total current passing through the epithelial barrier be zero. This necessitates that the current through the tight junctions, I^j, balance the current through the cell (e.g.,$I^j = -I^a$).

In vivo, the membrane potentials across the apical membrane, E^a, the basolateral membrane, E^b, and the epithelium, E^t, assume values that are at all times consistent with the preservation of electroneutrality. In the model, an iterative procedure (2D-Newton's method) is used to solve for the values of the potentials that satisfy the above electroneutrality constraints. This is embedded in the derivative evaluation subroutine called by the ODE solver. A detailed description of the numerical methods is given by Latta, et al. (1984).

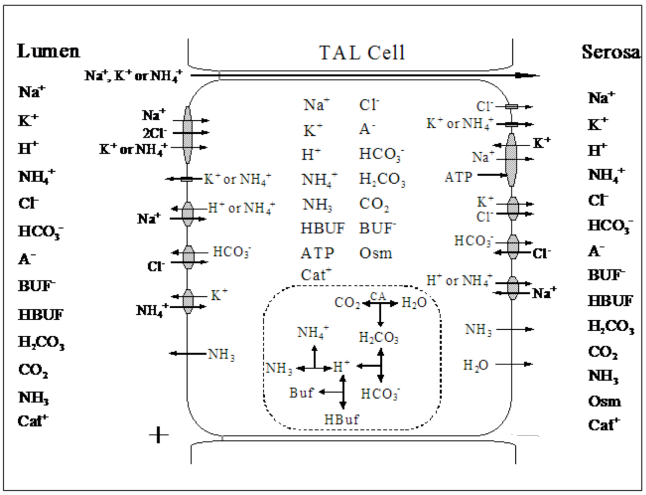


Figure 5. Model of a cell of the thick ascending limb of Henle (TAL).

3 Models of the renal medulla

The renal medulla is characteristic of higher vertebrates, and the inner medulla (i.e., long limbs of Henle with thin epithelium) exists only in mammals. The medulla is formed by the hairpin-turn feature of the nephrons (descending and ascending limbs) and blood vessels (called the *vasa recta* within the medulla), which, by *counter-current exchange* and *counter-current multiplication*, allow the creation of a cortico-papillary gradient of increasing osmolality that serves to concentrate the urine, thus permitting mammals to maintain salt and nitrogen balance without excreting a copious urine that would tend to dehydrate the animal. Thanks to this adaptation, which is developed to

different extents in different species and corresponds to differences of habitat and diet, some mammals can live in very arid environments with no water beyond that available in their food, which may consist only of dry seeds.

Surprisingly, and despite sustained attention for over 50 years (Torossi & Thomas 2001 (a recent translation of Hargitay & Kuhn 1951); Kokko & Rector 1972; Stephenson 1972; Lemley & Kriz 1987), the mechanism by which the osmotic gradient is created within the inner medulla (IM) still has no satisfactory explanation that accounts for measured permeabilities and the lack of active transport within the IM. There have been many modeling studies of this problem, from simple models of the microcirculation (see review of Pallone et al. 2003) and "central core" models that consider the nephrons bathed in a common, ideal vascular space (see review of Stephenson 1992), through multi-nephron models (Lory 1987; Chandhoke et al. 1981, 1985; Thomas 1991), and culminating in 3D models accounting for the lateral separation of structures within, especially, the outer medulla (Wexler et al. 1991; Thomas 1995; Layton & Layton 2003).

Most of these models share the same basic description of tubular flows and transepithelial transport, differing essentially in the topology of the network of tubules and vessels and in the numerical solution techniques, which vary according to the topological complexity (see Layton & Pittman 1994). We will briefly describe a model of medium complexity (Hervy & Thomas 2003) and its rather unusual customized numerical solution technique. An interactive web version of this model is available at http://www.necker.fr/kidneysim.

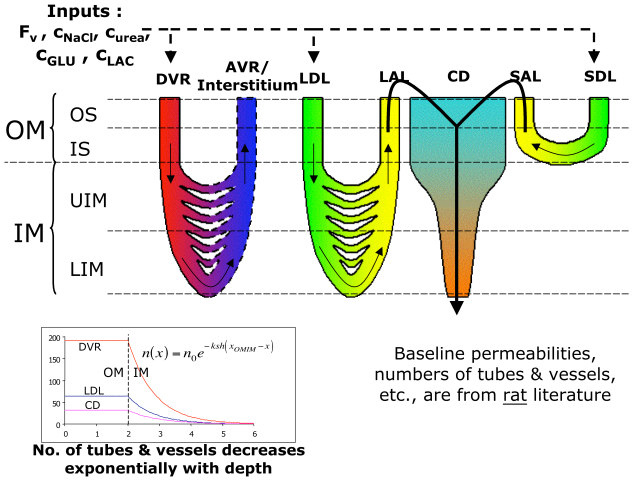


Figure 6. Flat CC model.

The steady state medullary model illustrated in Figure 6 includes vasa recta (descending, DVR, and ascending, AVR), short and long Henle loops (descending, SDL and LDL, and ascending, SAL and LAL), and collecting ducts (CD), and treats flows of volume, NaCl, urea, glucose, lactate, and (only in the CD) KCl. It is thus a system of 35 nonlinear ordinary differential equations (five flow variables along seven tubular structures). The AVR serve to represent the interstitium surrounding all the structures. Rather than explicitly including the equations for transport along the distal tubules, inflow to the OMCD is calculated from flows exiting at the top of the ascending limbs (AHL), based on physiological constraints representing the action of virtual distal tubules (see below under boundary conditions). This model is referred to as "flat" rather than "3D", since all exchange among tubes passes via a common interstitial space instead of being distributed among neighboring structures according to their relative placement within each region.

Model topology. Each type of tube is represented by a single, lumped tubular structure whose circumference at each depth reflects the total number of such tubes at that depth. Within the IM, flows in LDV and in LDL are shunted directly to LAV and LAL, respectively, in proportion to the number of tubes that return at each depth. The axial loop distribution decreases in number exponentially, based on reported anatomy of rat kidney(Han et al. 1992; Knepper et al. 1977). To be explicit, the number of tubes at depth x within the IM, i.e., for $x > xomim$, is given by

$$N(x) = N(0)e^{-ksh(x-xomim)}$$

with $ksh = 1.213$ mm^{-1} for vasa recta and Henle's loops and $ksh = 1.04$ mm^{-1} for IMCD, and $N(0)$ is the number entering the IM. Thus, compared to the number of tubes entering the IM, the fraction of vasa recta and Henle's loops reaching the papillary tip is 1/128 for an inner medullary thickness of 4 mm, and over the same distance, 64 OMCDs converge to a single exiting collecting duct. Also in conformity with the 3D models, two-thirds of the descending vasa recta turn back within the inner stripe of the OM (we call these the short descending vasa recta, SDV) and the remaining third (the long vasa recta, LDV) extend at least part-way into the IM, their number diminishing exponentially as explained just above. The SDV and LDV are distinct structures in the WKM-type 3D models, but in this flat model they are lumped into a single structure, the DVR. For the whole system, the basic scaling factor is $ncd0$ (=64), the number of OMCD entering the outer medulla. Since species other than the rat have different proportions of tubes and vessels extending to the tip (Bankir & de Rouffignac 1985), everything is scaled to the a single exiting CD. By this strategy, the model can represent kidneys containing any number of nephrons simply by varying the medullary length and/or the factor describing the exponential decrease of their number with depth (k_{sh}).

System equations. The equations describing the changes of flows and concentrations with depth in each tube are identical to those used in many other such models. System variables are the axial tubular flows of water and solutes. Concentrations of solutes i in tubes j are calculated from the ratio of solute flow to volume flow, $c_i = F_i^j/F_v^j$. As in Thomas (2000), shunt flows from descending tube j to the corresponding ascending tube at depth x are given by

$$F^j_{shunt}(x) = \frac{F_i^j(x)}{N_j(x)} \frac{d\ N_j(x)}{dx} = k_{sh} F_i^j(x) \tag{1}$$

We then have the following system of differential equations, adopting the usual convention that descending tubule flows are positive and ascending flows are negative:

$$\begin{aligned} \frac{d\ F_v^j(x)}{dx} &= -J_v^j(x) \ {}^{-}_{+}\ k_{sh}\ F_v^j(x) \\ \frac{d\ F_i^j(x)}{dx} &= -J_i^j(x) \ {}^{-}_{+}\ k_{sh}\ F_i^j(x) - J^j_{i,pump}(x), \end{aligned} \tag{2}$$

where in each case the subscript v refers to volume flows and the subscript i refers to NaCl, urea, glucose, or lactate, and j refers to the tubes or vessel. The symbol ${}^{-}_{+}$ means negative for descending tubes and positive for ascending tubes.

In these equations, transmural fluxes of volume and solute i out of tube j are given by:

$$\begin{aligned} J_v^j(x) &= A_j\ Lp_j \sum_{i=1}^{4} RT\sigma_i^j \Big(c_i^{AVR}(x) - c_i^j(x)\Big) \\ J_i^j(x) &= A_j\ P_i^j \Big(c_i^j(x) - c_i^{AVR}(x)\Big) + J_v^j(x)\Big(1 - \sigma - i^j\Big)\Big(\frac{c_i^j(x) + c_i^{AVR}(x)}{2}\Big) \end{aligned}$$

$$J^j_{i,pump}(x) \;=\; A_j \,\frac{Vm^j_i c^j_i(x)}{Km^j_i + c^j_i(x)}. \tag{3}$$

The AVR concentrations here also represent the interstitial concentrations surrounding all the other tubes. Notice also that there is no mention in the above equations of glycolytic conversion of glucose to lactate. This is because the conversion is limited to the "interstitial space", i.e., the AVR. The AVR concentrations and volume flow were calculated from the constraint of global mass balance, as follows. Conservation of mass for the medulla as a whole in the steady state (35) says simply that, at any depth x, the algebraic sum of flows of type i in all tubes j (taking flows to be positive toward the papilla and negative away from the papilla) must equal the exit rate of i from the terminal collecting duct minus the total amount of i synthesized from x to the papillary tip, $x = L$:

$$\sum_j F^j_i(x) = F^{CD}_i(L) - \int_x^L S^j_i(x)\,dx, \tag{4}$$

where flows at level x are taken to be zero in tubes that do not extend all the way to$x$. The S^j_i term is of course zero everywhere for NaCl and urea and applies only in the AVR/interstitium for lactate and for glucose (it is negative for glucose, since glucose is consumed).

The total inner medullary glycolytic glucose consumption, $jglytot$, is specified as a percent of total glucose inflow into DVR, and the rate of glycolysis at a given depth is then scaled to the number of vasa recta at that depth.

Boundary conditions at the bottoms of loops are based on tube connectivity. In general, it is considered that there are no sources or sinks (except glycolysis, for lactate and glucose), that hydrostatic pressure plays a negligible role compared to osmotic pressure forces, and that axial diffusion is negligible relative to convective flow of solutes (these last two assumptions were discussed in Moore & Marsh (1980)).

Inputs and boundary conditions. The inputs to the system are the volume flows and solute concentrations at the entry into the long and short descending limbs and into the vasa recta. F_v^{LDL} and F_v^{SDL} were set at 10 nl/min/nephron, based on an SNGFR (single nephron glomerular filtration rate) of 30 nl/min and TF/P inulin of 3 (i.e., the ratio of tubular fluid to plasma concentration of inulin) at the end of the proximal tubule. F_v into vasa recta was set at 7.5 nl/min/tube, as in (Thomas 1998). For the LDL and SDL, entering concentrations of urea, glucose, and lactate were set at 10mM, 1**?**M, and 1 **?**M, respectively. For the vasa recta, entering concentrations of urea, glucose, and lactate were set at 5mM, 5mM, and 2mM, respectively. NaCl concentrations were calculated from these assuming global entering fluid osmolality of 263 mosM and an osmotic activity coefficient for NaCl of 1.82 (Wang et al. 1994).

Inputs to the OMCD. Rather than include distal tubules explicitly, the entry to the collecting ducts is calculated from flow and concentrations at the top of the short and long ascending limbs, based on constraints deduced from the literature. To calculate the volume flow and four concentrations into the OMCD, we need five constraints. In particular, it was assumed that: (i) fluid entering the OMCD is isosmotic to plasma and is assigned the value $osmcd0$=263 mOsm; (ii) a specified fraction, $ufac = 0.85$, of urea is delivered to OMCD (i.e., the distal tubules reabsorb $(1 - ufac)$ of the urea delivered to early distal tubules); (iii) NaCl concentration entering the OMCD has a fixed value, $cscd0 = 35$ mM; (iv) glucose and lactate flows are conserved along the virtual distal tubules, i.e., their flows into OMCD equal the sum of their flows out of the LAL and SAL. We also assume KCl enters the OMCD at the fixed concentration $ckcd0$ (=20 mM) but that its absolute flow rate, , then remains unchanged along the rest of the CD, and its concentration at depth x is then

$ckcd(x) = fkcd/F_v^{CD}(x)$. This is used along with the other solute concentrations to calculate the osmotic driving force for water flux across the CD wall.

Numerical solution. The system was solved using a method based on that described by Stephenson et al. (1994) and used by us in an earlier model with six cascading nephrons (Thomas 1991). The differential equations are approximated by finite difference equations centered in space. If we consider tube j to be divided into n slices, then the space-centered FDEs between nodes $k-1$ and k are

$$\frac{F_i^j(k-1) - F_i^j(k)}{\Delta x} = -\overline{J}(k - \frac{1}{2}) \tag{5}$$

where i represents flows of volume, NaCl, urea, glucose, or lactate. Thus, the fluxes are evaluated at the middle of the interval ('mid-point method', Tewarson et al. 1985), by assuming that concentrations in the middle of the interval are the arithmetic average of the concentrations at $k-1$ and k.

Solution proceeds as follows. An initial guess is made for the interstitial/AVR concentrations, then taking these as fixed, and given the defined input volume flow and solute flows for LDL and SDL and for the descending vasa recta, the equations for each tube are integrated stepwise (we used a spatial chop of 120 slices (121 nodes)) in the direction of flow using Newton's method on the system of five finite difference equations and five unknowns (F_v and four concentrations, c_i) and using an analytically calculated Jacobian matrix. We found it advantageous to use a much stricter error tolerance ($< 10^{-10}$) on these 'tubular' iterations than was necessary on the 'global' iterations. Using the relative values for tubular flows and concentrations, $F_v^{AVR(k)}$ is calculated to satisfy global mass balance at each mesh node by applying equation (4) to volume flow and rearranging to obtain

$$F_v^{AVR}(x) = F_v^{CD}(L) - \Big(F_v^{DVR}(x) + F_v^{LDL}(x) + F_v^{LAL}(x) + F_v^{SDL}(x) + F_v^{SAL}(x) + F_v^{CD}(x)\Big) \tag{6}$$

Then, using these AVR volume flows, one checks for global mass balance for each solute at each discrete depth. This gives the following "scores", which would ideally equal zero. These are the relative deviations from an ideal solution (see the article for details). If the maximum relative deviation is less than 10^{-6}, we have a solution. If not, then a 'global' Jacobian is constructed numerically by varying each interstitial/AVR concentration in turn (the variation used here was 10^{-4} times the concentration in question) and re-integrating the system. This Jacobian matrix and the error vector based on eq. 6 are then used to solve for a corrections vector s to the interstitial concentrations by LU decomposition (Press et al., 1986). This 'globalÕ Newton-iteration is repeated until global convergence is achieved (i.e., until global mass balance is respected to within our chosen error tolerance).

4 Web resources for renal modeling

We have made initial progress towards the development of web resources for quantitative renal physiology. In particular,

• for the quantitative kidney knowledge database, QKDB, we have developed a working model for the database (entity-relationship model), have implemented this under MySQL, and have built (using PHP) a web GUI with basic functionality. This site is presently under password-protected testing by ourselves and is being seeded with data from our own reference collections and tested for improvements in ergonomy;

• we have developed two proof-of-concept web interfaces for exploration of published renal models and limited customized simulations: one (KSim) uses a Java applet for one of our renal medulla

models (Hervy & Thomas, 2003) and the other (Kidneysim3D) presents a published model of transport along the distal tubule (Chang & Fujita, 1999), implemented via translation into the markup language CellML and superimposed on a zoomable, 3D representation of the kidney (a "virtual kidney").

We will next describe the present state of these initial developments. Following these presentations, we describe existing technologies/ strategies upon which we will build for implementation of the project.

4.1 QKDB present state

We initially considered building QKDB using only XML technologies, which are rapidly occupying an ever greater portion of "dynamic web space". We concluded at the time, however, that the software toolset for this approach was not yet sufficiently stable or widespread given our commitment to an open source, platform-independent implementation. We therefore fell back on the robust and thoroughly multi-platform PHP-MySQL approach. After considering several more complicated schemes, we settled on rather simple but extensible data model described below. If migration to a more XML-centered scheme appears advantageous in the future, the transition will be relatively straightforward, since there are now many database management tools that allow export of MySQL database content in XML format and this can then be transformed using XSLT or similar tools.

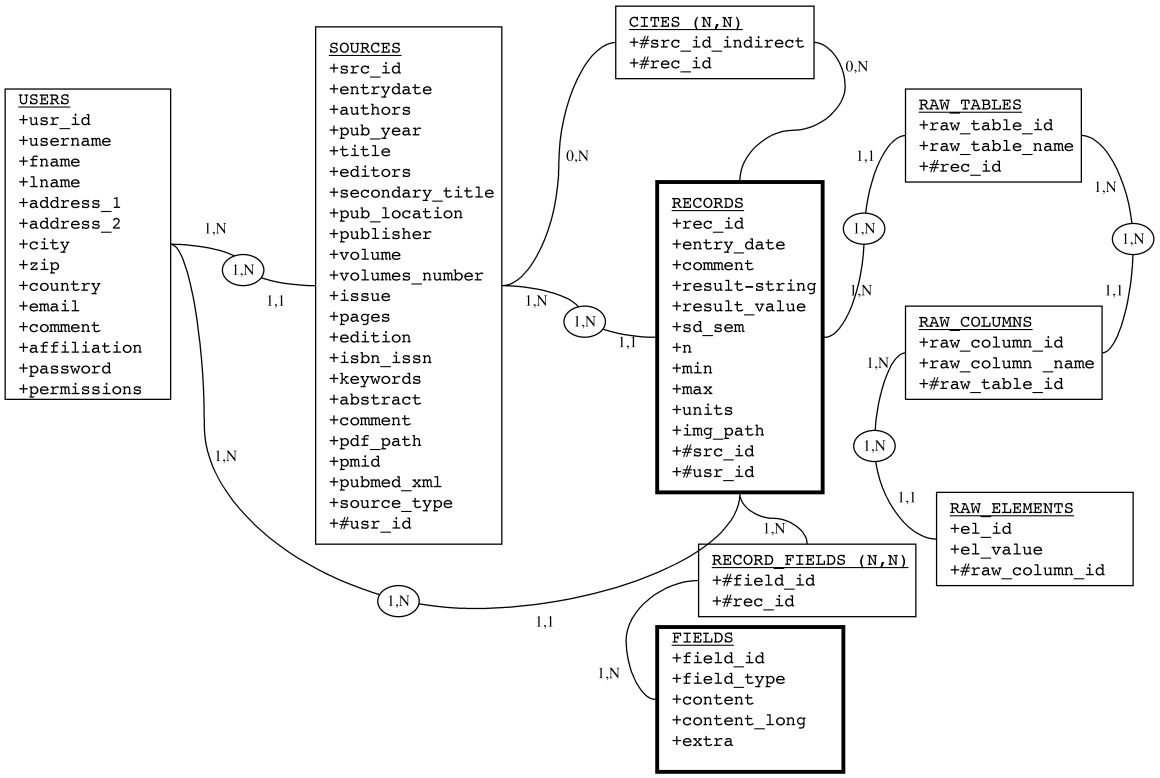


Figure 7. Entity-Relation model for QKDB.

The QKDB data model. Development of an adequate data model was complicated by the varied nature of the data to be included, by the fact that the list of types of data will continue to grow, and by the requirement that this extensibility not necessitate re-programming of the dynamic web

interface for data querying and data entry.

Figure 7 shows the entity-relationship model we have adopted for the phase I development period (Dzodic et al. 2004). Although it is hoped that this model will suffice, it may still evolve as the site matures. We will describe this data model in some detail to illustrate the novel feature that both confers extensibility and ease of conversion for other purposes.

The boxes in the diagram represent tables in the MySQL relational database. The arrows indicate relations between the tables, which may be 1-to-1, 1-to-many, or (via an intermediate indexing table) many-to-many. If each table is thought of as columns and rows, the field names in each box may be thought of as column names, and each row then represents an individual record having a unique identifier. Relational links between records in different tables are established by designating the IDs of related records (indicated by the #-sign) in other tables.

The central item of QKDB is an individual data record, placed in the $RECORDS$ table. The rest of the tables serve merely to qualify and identify the record and indicate who contributed it and when. This table contains fields for a numerical value and its mean, error, min, and max, and the units of measurement, but there is also provision (the $result\_string$ field) for a qualitative result (such as "stimulates", "inhibits"É) or the path and name of an image file (stored as a blob or whose path/url is indicated) in the img_path field. Finally, a comment field contains annotations from the contributor relevant to the individual record.

Since only published data will be included in QKDB, each record is associated with a literature reference, given in the $SOURCES$ table. Provision is also made for indirect citations, such as may appear in review articles, via the $CITES$ indexing table, since this is possibly a many-to-many relationship. The $SOURCES$ table includes fields for standard PubMed style literature references and may be entered by hand or it will be possible to fill it by exporting a citation file supplied by reference managing software such as EndNote or Reference Manager or in the form of an XML file containing the relevant fields. There is also a comment field for annotations from the contributor about the reference as a whole, and a field for the pdf version of the article for cases where its inclusion is feasible and poses no copyright problem.

The $USERS$ table contains fields for identification of the contributor of data items, including his/her password and the level of permission to which (s)he is authorized. Three levels of permission are planned: simple users (can freely browse the contents of the DB), contributors (recognized researchers with the right to contribute to the DB), and administrators (small committee who can modify and delete DB contents as well validate new entries).

The three RAW_xxx tables provide the possibility of entering tables of raw data from articles that include them, though this is rare in publications after about 1975.

Finally, the **FIELDS** table, along with its many-to-many relation with the $RECORDS$ table, through the indexing $RECORD\_FIELDS$ table, is the heart of this data model's flexibility and ready extensibility. The $FIELDS$ table can be extended at will to include not only new items of any particular $field\_type$ but also to include previously unanticipated field types. Since this is the only part of the data model that contains kidney-specific characteristics, the conversion of the whole scheme for a completely different field of research would involve nothing more than changing the entries in the $FIELDS$ table, because the data entry input and query output forms on the web interface are constructed dynamically from these tables.

GUI and quality control for QKDB. The screenshots in Figure 8 give an indication of the GUI of the present prototype version of QKDB. The roll-down lists used to specify a query (or to characterize a new entry on the contibution pages) are built dynamically from the contents of the $FIELDS$ table as the webpage is loaded.

Quality control of the curated data is assured at several levels. Besides the customary types of syntax checking on entry form contents (coded into the PHP program), and the use of pull-down menus built dynamically from the contents of the database (thus avoiding duplication and multiple spellings of field names), the data itself is entered only by authenticated researchers, under password control. Finally, a scientific board of experts oversees and verifies the database contents, including validation of all new entries before they become accessible to queries.

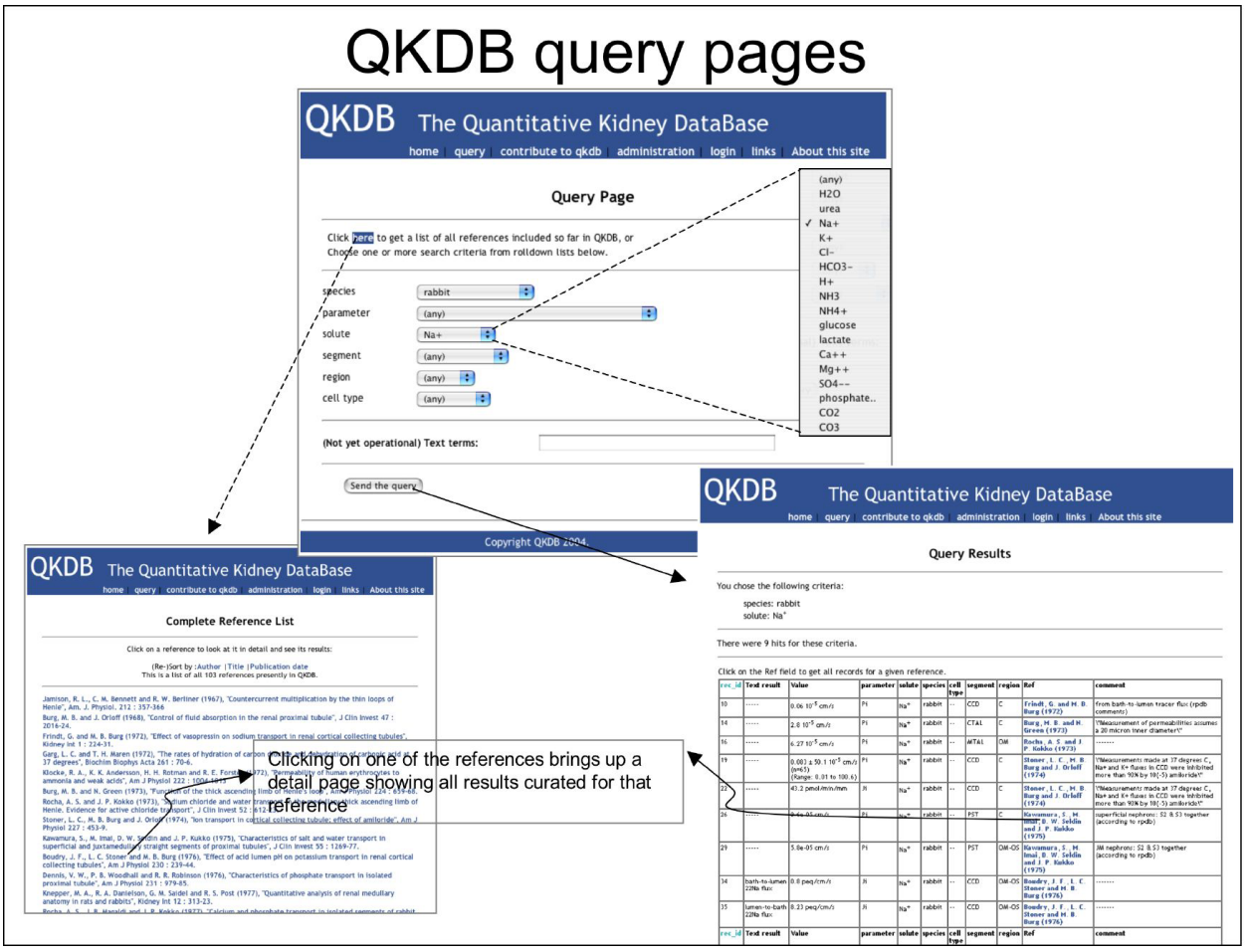


Figure 8. Screenshots of webpages for querying QKDB.

Systems mediation, ontologies and other renal databases

Once the modeling resource and QKDB are both well launched and open to the general community, it will be desirable to establish active links between the two and also to link them with appropriate resources in related fields. For example:

• there exist other databases for renal research to which it would be natural to provide meaningful communication. Among these are:

- the Collecting Duct Database

(http://cddb.nhlbi.nih.gov/cddb/) (Legato et al. 2003) which provides curated information for the

design of microarray targets to scan collecting duct tissues (human, rat, mouse). Exchange between this database and our QKDB may be established via an SBML bridge.
- the Kidney Development Gene Expression Database (http://organogenesis.ucsd.edu/data.html)
- others that exist or may be created in the meanwhile
• 'sibling' databases of QKDB, created for other organ systems using a generic version (QxDB). Already two such are under construction, one for cancer research and the other treating the mechanical properties of cells and tissues.

It will also be essential to develop and share common ontologies among all these resources. This ontology construction is an integral part of the program outlined in the IUPS Physiome Roadmap.

4.2 Two modeling site prototypes

KSim Ñ present state. We have developed a prototype Java applet to display simulation results (http://www.necker.fr/kidneysim/). To ground the development in a realistic context, we implemented this prototype around one of our own models of the renal medulla (Hervy & Thomas 2003). On this website, separate web-pages present: the history leading up to the development of this particular model; a description of the model itself, including a table of the basic parameter values; and finally a window giving entry into a Java applet for presentation of simulation results. The visitor may choose among a small set of previous simulations (typical of those presented in the published article) or may launch a new simulation (server-side execution) based on his/her modifications of a selected set of model parameter values. The applet loads the chosen set of simulation results in the form of an XML file and permits the web visitor to display them as x-y plots or as a color-gradient diagram of the medullary structures. These are illustrated in the screenshots below of the applet GUI, one under Windows XP and the other under Mac OS X.

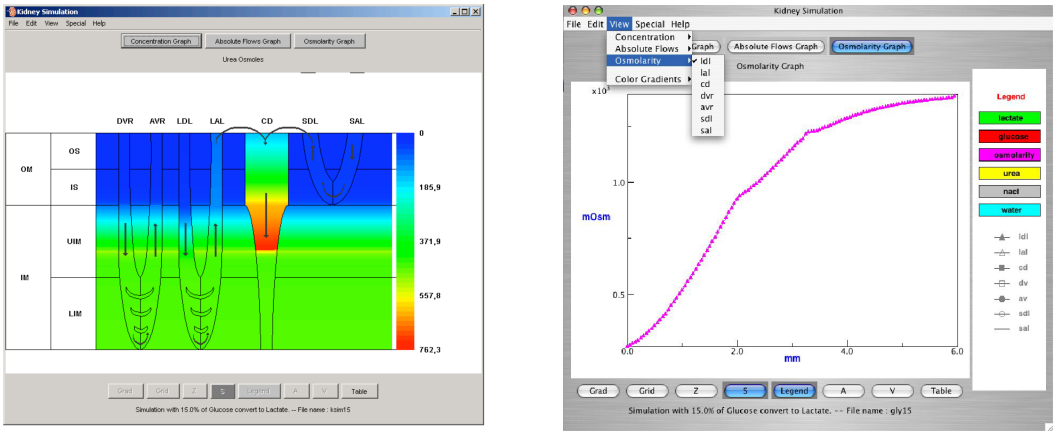


Figure 9. Screenshots from java applet for display of simulation results from the 'flat' medullary model of Hervy & Thomas (2003). Upper panel shows color gradient of urea concentrations in medullary structures. Lower panel shows the graphing interface (total osmolality along collecting ducts).

Development of this demonstration gave us valuable experience concerning the possibilities and limitations of existing graphics packages for Java applets and revealed key development features to focus on in order to develop a more flexible and generic approach for use with the variety of models projected for the Renal Physiome site.

KSimII 3D-Virtual Kidney interface The following screenshots show a first version of our *'virtual kidney' interface* (Lonie et al. 2004, 2005; Thomas et al. 2004). This preliminary version

(funded by the Victorian Partnership for Advanced Computing) was constructed from casts of a rat kidney, but is still far from complete. We propose to use this interface not only for exploring kidney anatomy but also as an interface to the collection of curated, interactive models at various scales. For exploration of kidney anatomy, in the left screenshot of Figure 10, the model in the left panel can be zoomed, rotated and even disassembled using the mouse and its buttons. This is illustrated in the right screenshot, where the capsule, outer medulla, and blood vessels have been removed, revealing an anatomically correct superficial nephron whose path through the outer medulla can be examined by rotating and zooming.

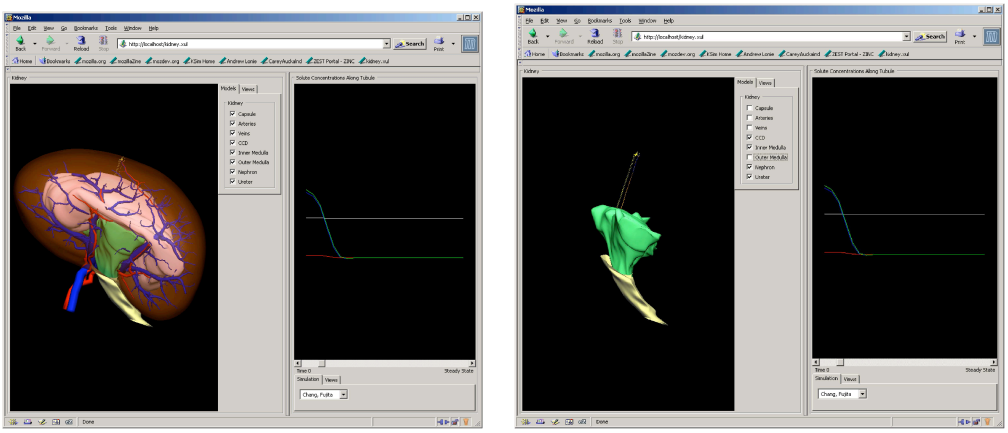


Figure 10. 3D-Virtual Kidney interface. (Left) For exploration of kidney anatomy, the model in the left panel can be zoomed, rotated and even disassembled using the mouse and its buttons. The right panel displays graphs of simulation results from a model selected from a roll-down list. (Right) The capsule, outer medulla, and blood vessels have been removed (checkboxes between the left and right panels), revealing a superficial nephron whose path through the outer medulla can be examined by rotating and zooming.

To illustrate the implementation of Chang & Fujita's (1999) model of transport along the distal tubule, Figure 11 shows a close-up of the nephron, and we draw now attention to the graph in the right panel, where the colored lines represent profiles of ion concentrations (Na^+, K^+, and Cl^-) along the length of the tubule. A slider at the bottom of the panel moves through the simulated time-steps. In the nephron shown in the left panel, at this scale, one can clearly see a spurious 'rod'; this represents the distal tubule in this preliminary version and the colors along its length change to indicate changing concentration of a selected variable. In the next phase of development, this will be mapped onto the distal tubule, instead of this rod, and many additional simulations of the other nephron segments will be implemented in the same manner.

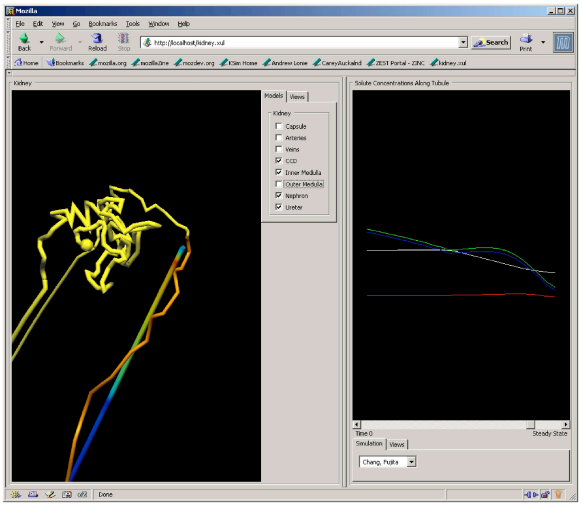


Figure 11. Closeup view of superficial nephron (left panel) and graphs of distal tubule transport model results (right panel); the colored lines represent profiles of ion concentrations (Na^+, K^+, and Cl^-) along the length of the tubule. A slider at the bottom of the right panel moves through the simulated time-steps. In the left panel, at this scale, the spurious 'rod' represents the distal tubule in this prototype, and the colors along its length change to indicate changing concentration of a selected variable.

This GUI is being built using the XUL environment of the Mozilla group (open source development tools) to provide a web front-end for the 3D visualization environment developed by the group of Peter Hunter in Auckland. Their work has long been applied to models of cardiac physiology, and they are now working with us to adapt these tools for the kidney.

Figures 12 and 13 give an idea of the present state of progress on the GUI proposed by the Auckland group for interacting with multi-scale Physiome models. This work is funded by the Wellcome heart physiome project and the CMB CoRE and is aimed at providing a general GUI for physiome models across all organ systems (at all spatial scales) via the ontology database. It is intended both as a means of navigating the model databases and as a means of running model simulations and viewing simulation results.

The overall framework uses the Mozilla browser/Cmgui approach developed within the Auckland group and incorporates their MozCellML cell model integrator. The models are retrieved from a ZOPE CellML/ontology database. The layout within the Mozilla window is defined by XUL. The 2D pathway graphs use SVG.

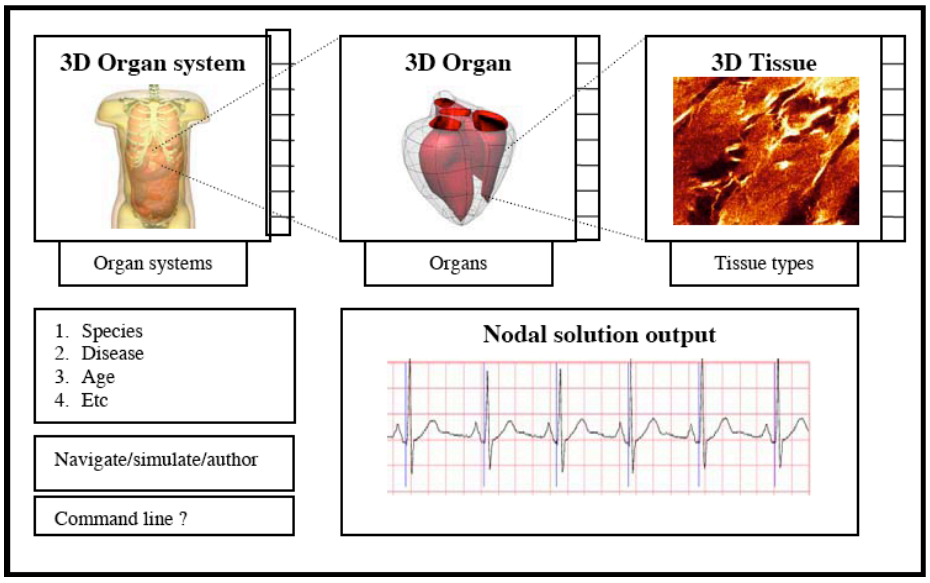


Figure 12. Organ level view of GUI for physiome models. The interface allows zooming to different scales, as shown in the top row, switching between modes (menus in lower left) and display of simulation results (lower tracing) for any selected node in the 3D window displays.

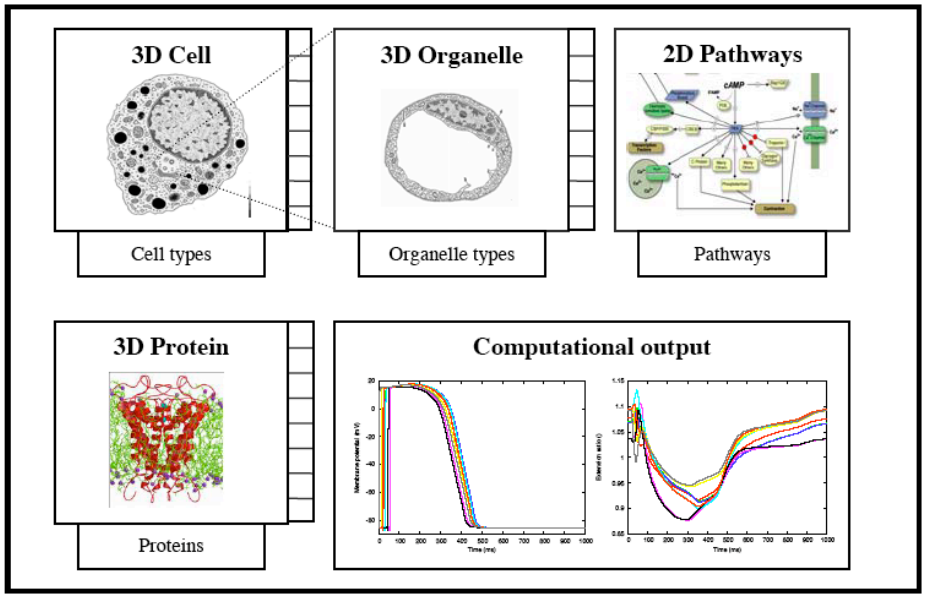


Figure 13. Cell level view of GUI for physiome models. The 3D cell & organelle windows show the spatial distributions of proteins, whereas the right-most panel shows the topological layout of pathways.

5 Conclusion

We have shown several of the methods used for modeling the kidney and illustrated web resources whose purpose is to render kidney models easier to develop (thanks to better accessibility of experimentally measured parameters via QKDB) and to make them accessible to a wider audience. These latter projects are part of the more general IUPS Physiome project and share its goal to develop generic, open source solutions to give access to physiological knowledge at all levels of organization. Our laboratory also participates in two projects in France that share many of the same goals as the Physiome project, namely, the GdR (Groupement de Recherche) STIC/Sant (CNRS Unit 2647) (http://stic-sante.org) and the Institute of Theoretical Medicine (IMTh), based in Lyon.

References

Bankir, L. and C. de Rouffignac (1985). “Urinary concentrating ability: insights from comparative anatomy.” Am J Physiol - Regulatory 249: R643-R666.

Chandhoke, P. S. and G. M. Saidel (1981). ”Mathematical model of mass transport throughout the kidney: Effects of nephron heterogeneity and tubular-vascular organization.” Annals of Biomedical Engineering 9: 263-301.

Chandhoke, P. S., G. M. Saidel and M. A. Knepper (1985). ”Role of inner medullary collecting duct NaCl transport in urinary concentration.” Am J Physiol 249(5 Pt 2): F688-97.

Chang, H. and T. Fujita (1999). “A numerical model of the renal distal tubule.” Am J Physiol 276(6 Pt 2): F931-51.

Chang, H. and T. Fujita (2001). “A numerical model of acid-base transport in rat distal tubule.” Am J Physiol Renal Physiol 281(2): F222-43.

Dzodic, V., S. Hervy, D. Fritsch, H. Khalfallah, M. Thereau and S. R. Thomas (2004). “Web-based tools for quantitative renal physiology.” Cellular and Molecular Biology 50(7): 795-800.

Han, J. S., K. A. Thompson, C. L. Chou, and M. A. Knepper. Experimental tests of three-dimensional model of urinary concentrating mechanism. J Am Soc Nephrol 2: 1677-88., 1992.

Hargitay, B. and W. Kuhn (1951). ”Das Multipikationsprinzip als Grundlage der Harnkonzentrierung in der Niere.” Zeitschrift fr Elektrochemie 55(6): 539-558.

Hervy, S. and S. R. Thomas (2003). “Inner medullary lactate production and urine-concentrating mechanism: a flat medullary model.” Am J Physiol Renal Physiol 284(1): F65-81.

Knepper, M. A., R. A. Danielson, G. M. Saidel, and R. S. Post. Quantitative analysis of renal medullary anatomy in rats and rabbits. Kidney Int. 12: 313 - 323, 1977.

Kokko, J. P. and J. F. C. Rector (1972). ”Countercurrent multiplication system without active transport in inner medulla.” Kidney International 2: 214 - 223.

Latta, R., C. Clausen and L. C. Moore (1984). “General method for the derivation and numerical solution of epithelial transport models.” J Membr Biol 82(1): 67-82.

Layton, A. T. and H. E. Layton (2003). ”A region-based model framework for the rat urine concentrating mechanism.” Bulletin of Mathematical Biology 65(5): 859-901.

Layton, H. E. and E. B. Pitman (1994). “A dynamic numerical method for models of renal tubules.” Bull. Math. Biol. 56: 547-565.

Legato, J., M. A. Knepper, R. A. Star and R. Mejia (2003). “Database for renal collecting duct regulatory and transporter proteins.” Physiol Genomics 13(2): 179-81.

Lemley, K. V. and W. Kriz (1987). ”Cycles and separations: The histotopography of the urinary concentrating process.” Kidney International 30: 538-548.

Lewis, S. A. and N. K. Wills (1981). “Interaction between apical and basolateral membranes during sodium transport across tight epithelia.” Soc Gen Physiol Ser 36: 93-107.

Lonie, A. J., C. Stevens and P. Harris (2004). LB520: Computer modelling of kidney function. Experimental Biology 2004, Washington D.C.

Lonie, A., C. Stevens, P. Harris and S. R. Thomas (2005). Visualization tools for the Renal Physiome. Abstract 4086 at IUPS/Experimental Biology, San Diego.

Lory, P. (1987). “Effectiveness of a salt transport cascade in the renal medulla: computer simulations.” AJP 252 (Renal 21): F1095-F1102.

Mejia, R. and J. L. Stephenson. (1979). “Numerical solution of multinephron kidney equations.” J. Computational Physics 32: 235-246.

Moore, L. C., and D. J. Marsh. How descending limb of Henle’s loop permeability affects hypertonic urine formation. Am J Physiol Renal 239: F57-71., 1980.

Netter, F. H. (1973). Kidneys, Ureters, and Urinary Bladder. Summit, NJ, CIBA.

Pallone, T. L., M. R. Turner, A. Edwards and R. L. Jamison (2003). ”Countercurrent exchange in the renal medulla.” Am J Physiol Regul Integr Comp Physiol 284(5): R1153-1175.

Press, W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling (1986). Numerical Recipes. The Art of Scientific Computing. Cambridge, Cambridge University Press.

Stephenson, J. L. (1972). ”Concentration of urine in a central core model of the renal counterflow system.” Kidney International 2: 85-94.

Stephenson, J. L. (1992). Urinary concentration and dilution: models. IN: Handbook of Physiology. Section 8. Renal Physiology. E. E. Windhager. Bethesda, Md., Oxford University Press. 2: 1349-1408.

Stephenson, J., R. Tewarson, and R. Mejia. Quantitative analysis of mass and energy balance in non-ideal models of the renal counterflow system. PNAS 71(5): 1618-1622., 1974.

Stephenson, J., Y. Zhang, A. Eftekhari and R. Tewarson (1987). “Electrolyte transport in a central core model of the renal medulla.” American Journal of Physiology Renal 253: F982-F997.

Tewarson, R., JL Stephenson, M Garcia, and Y. Zhang. On the solution of equations for renal counterflow models. Comput.Biol.Med. 15(5): 287-295., 1985.

Thomas, S. R. (1991). “Effect of varying salt and urea permeabilities along descending limbs of Henle in a model of the renal medullary urine concentrating mechanism.” Bull Math Biol 53(6): 825-43.

Thomas, S. R. Cycles and separations in a model of the renal medulla. Am. J. of Physiol.

Renal 275: F671-F690, 1998.

Thomas, S. R. Inner medullary lactate production and accumulation: A vasa recta model. Am. J. Physiol. Renal 279: F468-F481, 2000.

Thomas, S. R. and G. Dagher (1994). “A kinetic model of rat proximal tubule transport–load-dependent bicarbonate reabsorption along the tubule.” Bull Math Biol 56(3): 431-58.

Thomas, S. R. (1998). “Cycles and separations in a model of the renal medulla.” Am J Physiol 275(5 Pt 2): F671-90.

Thomas, S. R., H. Layton, A. Layton, P. Harris, A. Lonie and L. Moore (2004). Towards a web resource for quantitative renal physiology. Abstract at: Physiological Society, King’s College London, J. Physiology (in press).

Torrossi, T. and S. R. Thomas (2001). ”The multiplication principle as the basis for concentrating urine in the kidney.” J. Am. Soc. Nephrol. 12: 1566-1586. (translation of ”Hargitay, B. and Werner Kuhn, 1951, ”Das Multiplikationprinzip als Grundlage der Harnkonzentrierung in der Niere”, Zeitschrift fr Elektrochimie und angewandte physikalische Chemie 55(6):539-558.)

Wang, X., A. S. Wexler, and D. J. Marsh. The effect of solution non-ideality on membrane transport in three- dimensional models of the renal concentrating mechanism. Bull Math Biol 56: 515-546, 1994.

Weinstein, A. M. (1994). ”Mathematical models of tubular transport.” Annu Rev Physiol 56: 691-709.

Wexler, A. S., R. E. Kalaba and D. J. Marsh (1991). ”Three-dimensional anatomy and renal concentrating mechanism. I. Modeling results.” American Journal of Physiology Renal 260(3 Pt 2): F368-83.

Yang, H. M., H. E. Layton, C. Clausen and L. C. Moore (2001). ”Mathematical model of NaCl transport in cortical thick ascending limb: Ammonium recycling enhances maximal diluting capacity (abstract).” FASEB J. 15: A143.

Cellular networks morphogenesis induced by mechanically stressed microenvironments

Philippe TRACQUI*, Patrick NAMY and Jacques OHAYON

CNRS, Laboratoire TIMC/DynaCell
Institut de l'Ingénierie et de l'Information de Santé (IN3S),
Faculté de Médecine de Grenoble 38706, La Tronche Cedex France

Abstract

While the Turing reaction diffusion pre-pattern theory of morphogenesis, and the following consideration of cellular chemotaxis models, have been successfully applied to explain the emergence of morphogenetic patterns in different experimental situations, physical properties of extracellular matrices (ECM) are now emerging as unavoidable actors the morphogenetic scenarios which are observed from subcellular to tissues levels. ECM may display adhesive ligands gradients important for anchorage-dependent cells motility and for resulting directional cell locomotion. In addition, ECM stiffness couples cellular traction forces to modulate this haptotactic cellular response.

A theoretical model based on a description of these cell-ECM interactions is proposed here as a basis for understanding how strains propagation within the ECM may define morphogenetic fields which trigger cell aggregation process. We first consider the generic morphogenetic properties emerging in planar cell cultures when cell migration is guided by physico-chemical interactions at the cell-substrate interface. We then illustrated how the dynamical instabilities induced by the balance between cellular forces and ECM compliance generate the different cellular networks morphologies observed when endothelial cells are cultured on biogels, a process called tubulogenesis or *in vitro* angiogenesis. Considering more specifically experiments conducted on fibrin gels with Human Umbilical Vein Endothelial Cells (HUVEC) and EAHy926 endothelial cells, we show that experimental bell-shaped angiogenic index curves can be explained by spatio-temporal evolution of the mechanical stress fields within the biogel. We finally discussed how these results could be included in a multi-scale modelling approach which integrates the cytoskeleton as a structural and dynamical node of the angiogenesis signalling network.

Keywords: mechanical signalling, *in vitro* angiogenesis, mathematical model, extracellular matrix rheology, cellular traction, pattern formation, cellular haptotaxis

1. Introduction

Almost one hundred years after the publication of the classic book "On Growth and Form" by D'Arcy Thompson (Thompson, 1917), understanding biological morphogenesis with the aid of theoretical approaches remains a very challenging and a largely unsolved multi-disciplinary issue. Indeed, considering the incredible increase of biological data and computers performances over these years as rough indicators, this clearly underlines that adequate frameworks and paradigms have still to be found to understand the development of

* Author for correspondence : P. Tracqui, email: Philippe.Tracqui@imag.fr

biological patterns and forms in different contexts. This situation is partly explained by the fact that morphogenesis is fundamentally a dynamical and multi-scale process, and thus a major ftheme of that is called now system biology. Another reason could be that morphogenesis is intimately a matter of growth and form, as well as of form and forces, as presented in another seminal book, "The forces that shape the embryo" by Trinkhaus (1984). Physical forces are indeed omnipresent and necessary at all levels of organization, from DNA decompaction and microtubules assembly (Tabony et al., 2002) to cell cytokinesis or developing embryos (Brouzes and Farge, 2004).

However, theoretical approaches to morphogenesis are still largely influenced by forces-devoid approaches, inspired by the founding work of Turing (Turing, 1952). The Turing's reaction diffusion pre-pattern theory is based on the assumption that diffusible chemicals, called morphogens, can react in a nonlinear way and possibly "self-organise" into various heterogeneous steady-state spatial concentration patterns. Turing's proposition is thus that cells embedded in such a morphogenetic field will respond appropriately to the morphogen concentration pattern, particularly by entering into specific differentiation pathways which specify the cell fate over the epigenetic landscapes depicted by Waddington (1968). This second, cellular interpretation phase, remained largely hypothetical many years after Turing's proposition, since the identification of real morphogens was lacking up to the seventies, where the identification of retinoic acid provides a decisive step for supporting this paradigm. Nevertheless, reaction-diffusion models have been applied widely and with success to a variety of biological problems, including mammals, butterflies and shells coat patterns (Meinhardt, 1995; Murray, 2003) to the interplay of segmentation genes expression during drosophila embryogenesis (Meinhardt, 1989) or the localisation of the division site in E. Coli (Meinhardt and de Boer, 2001).

A somehow extension of this theoretical framework for cell and tissues morphogenesis was to include explicitly the cell response within the model, by considering cell migration in response to local gradients of diffusible chemicals. In this so-called chemotaxis scenario, the spatial pattern is created by local cell aggregation, resulting from directional cell migration toward the chemical stimuli provided by the chemoattractant gradient. Nevertheless, these conceptual frameworks do not include explicitly the physical properties of the extracellular matrices (ECM) which surround cells in tissues and organs and which, with their respective composition, structure and mechanical properties, act as specific and responsive scaffold during the morphogenesis processes.

Thus, the mechanochemical modelling framework for generating biological forms and patterns proposed by Murray and Oster in the early 1980's (Murray and Oster, 1984, for a review see Murray 2003) was a major turning point in the field. The corner stone of this approach, firmly supported by the original experiments of Harris and co-workers (Harris, et al., 1981), is that adherent cells cultured on or embedded into an ECM exert traction forces and thus generate, through the resulting ECM distortion, mechanical morphogenetic signals which will affect cell dynamics. In the early 1980's, this theoretical approach provided a promising alternative to the still rather illusive nature of the morphogens. Today, this original pioneering modelling framework proposition is strongly supported by the increasingly recognized importance of mechanical factors on different aspect of cell dynamics. Indeed, ECM not only displays adhesive ligands important for anchorage-dependent cells, but also presents a wide range of mechanical and structural properties which influence many cell functions such as migration, proliferation, proteolytic activity, and signal transduction, ... Using extracellular substrates with different mechanical stiffness shows that cells organize their cytoskeleton and adhesive contacts differently on soft and stiff surfaces, with substrate stiffness affecting cell spreading and motility (Pelham and Wang, 1997; Lo et al., 2000).

In this paper, the morphogenetic properties emerging from the interplay of cellular forces and ECM-controlled cell dynamical processes will be illustrated in the specific context of *in vitro* angiogenesis. Indeed, in this experimental model, experiments and theories are sufficiently close to be able to shed light on some of the crucial elements and biophysical mechanisms which generate and control the specificity of the structural cell patterns which are observed.
Let us recall briefly that angiogenesis defines the formation of new capillaries from pre-existing vasculature (Folkman and Haudenschild, 1980). Under physiological conditions, angiogenesis is a highly regulated phenomenon controlled by angiogenic stimulators and inhibitors, including soluble factors (various peptides and growth factors like VEGF, Gerhardt et al., 2002) and non-diffusible ECM components. Indeed, different experimental data have shown that ECM mechanical properties and remodelling, resulting from the cell-dependence balance between biosynthesis and proteolysis, defines a biomechanical context which regulates capillary formation *in vivo*. Significant insights into the molecular and cellular biology of angiogenesis have come from *in vitro* assays using cultured endothelial cells (Vailhé et al., 2001). These assays mimics quite well the early morphogenesis of cellular networks of capillary-like structures, or tubulogenesis, during which endothelial cells cultures on different types of ECM may self-organise around meshwork of growing areas devoid of ECM, the lacunae (Figure 1), which can further increase in size and evolve toward the formation of a network of capillary like structures.

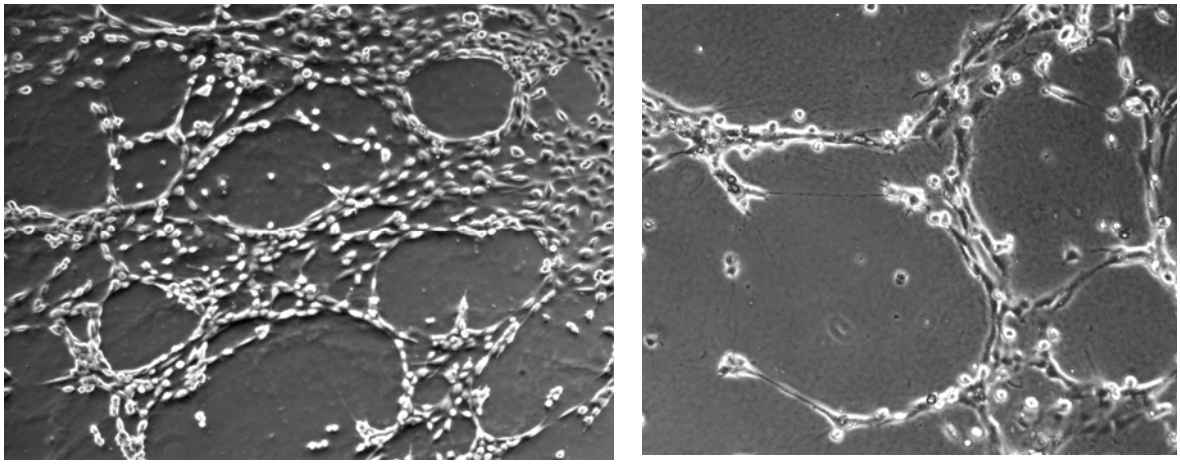


Figure 1: Reorganisation of endothelial EAhy926 cells into lacunae within fibrin gels observed at two different scales (x50, Fig.1.A and x100, Fig. 1.B)

We will focus here on a recent successful application of the mechano-biological framework of morphogenesis in predicting the experimental critical parameters beyond which tubulogenesis and capillary-like networks (CLN) formation may occur (Namy et al., 2004). We first present appropriate forms for the various terms included in the mechanobiological model, focusing on rheological ECM properties and ECM dependant cell migration. Starting from model simulations performed within simple cell-ECM interactions, we then show how the integration of cell dynamics within a mechanically varying micro-environment defines a self-organised dynamical system which promotes capillary-like networks (CLN) formation in specific conditions.

2. Modelling cell migration controlled by extracellular matrices.

Cell migration involves a series of highly coordinated events, including the extension of cell membrane protrusions, the formation of new adhesions, the development of traction forces, and the release of previous adhesion to insure cell body translocation (Sheetz et al., 1998).

At the molecular level, the adhesion process is mediated by transmembrane receptors, mostly of the integrins family. These receptors play an important role since they are physically associated with numerous cytoskeletal and signalling proteins which co-localise at focal contacts, and mechanical properties of the ECM are known to influence the strength of the integrin-cytoskeleton linkages (Choquet et al., 1997). Thus, the integrin $\alpha_v\beta_3$ is largely expressed in endothelial cells engaged in the angiogenic process (Vailhé et al., 1997).

Direction and speed of cellular migration, or taxis, is monitored by environmental stimuli, especially gradients of soluble chemicals (chemotaxis) or gradients of substrate-attached molecules (haptotaxis). Cellular shape, orientation, and migration can also be guided by the topography of the micro-environment. This process, referred to as "contact guidance" or "topographic guidance" is clearly demonstrated by the alignment of cells with micromachined grooves in the substrate (Oakley et al., 1997). Other study has also demonstrated that neutrophils can probe the tension in a three-dimensional ECM and move along the most rigid fibrils (Mandeville et al., 1997).

Dynamical expression of such haptotactic and contact guidance migration processes become more complex to predict when the cells micro-environment continuously varies as a results of the substrate deformation induced by cellular traction forces. At the cell population level, this mechano-stimulating loop may act as both a feedback and a feed-forward control of cell migration: cells are capable of interpreting and respond to the substrate deformation through a tactile exploration process, but they can also exert contractile forces which will determine a preferred direction of motion for the neighbouring cells.

Such involvement of a force-sensing mechanism is confirmed by different experiments: transient mechanical stimuli can induce motility of stationary keratocytes (Verkhovsky et al., 1999), and the direction of fibroblasts migration can be guided by manipulating mechanical strain within collagen-coated polyacrylamide substrates (Lo et al., 2000). However, theoretical models are clearly needed for understanding more precisely how the coupling between ECM mechanical behaviour, cellular traction forces and active cell migration may give rise to specific and self-organised spatial cellular morphologies.

The following paragraphs will precise several ingredients of such mechanobiological models. Starting with a brief illustration of cellular patterning induced by haptotactic cell migration, we will especially focus on theoretical sub-models able to describe the interplay between cell migration processes and constitutive stress-strain relationships derived for ECM rheology.

2.1. Haptotactic cell migration in an unstrained micro-environment

Random cell motility, resulting in cell displacement without any preferential direction of motion, is adequately described by cellular diffusion, in which the diffusive cell flux $\mathbf{J}_d$ is proportional to the local cell gradient:

$$\mathbf{J}_d = -D \nabla n$$

where the parameter D is the cellular diffusion coefficient which can be identified from cell tracking experiments.

In a similar way, active cell migration by haptotaxis, i.e. up to ligand-density gradients (or adhesivity gradient) is modelled by considering an haptotactic cell flux $\mathbf{J}_h$. In a first approximation, this flux is proportional to both the local cell density n(**x**,t) and to the local ligand-density gradient ∇g. Assuming that the sensitivity of the cell response to this gradient is monitored by a coefficient h (haptotactic coefficient), one gets the phenomenological expression:

$$\mathbf{J}_h = h\, n\, \nabla g$$

In absence of cell proliferation and cell death, the spatio-temporal evolution of the cell density is given by the following conservation equation:

$$\frac{\partial n}{\partial t} = -\nabla.[\, \mathbf{J}_d + \mathbf{J}_h \,]$$

As intuitively expect, haptotactic migration will lead to spatially inhomogeneous cellular patterns if the substrate has been coated with an inhomogeneous density of ligands. This is illustrated by the virtual experiment depicted in Figure 2: the extracellular substrat is coated with a uniform ligang density g(**x**)=1, except within a squared frame (grey area in Fig.2B) where the density is 20% higher. If initially cells are uniformly seeded on this patterned substrate (Fig.2A), they will migrate toward the squared area and finally aggregate along the square when the haptotactic motion overcomes cellular diffusion (Fig.2C).

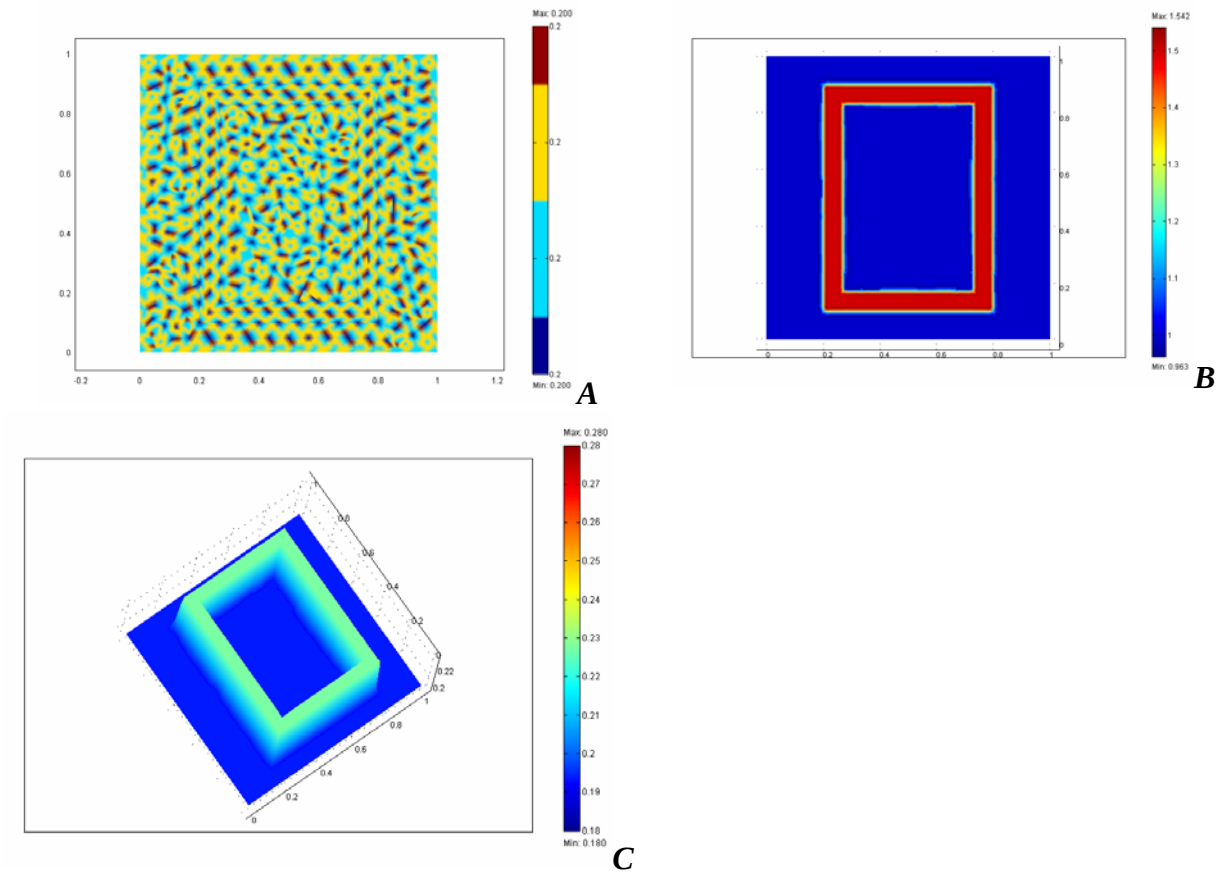


Figure 2: Simulation of haptotaxis-driven cell aggregation on a substrate coated with and heterogeneous density of ligand. The ligand density is 1.2 in the red squared region (Fig.2.B), and equal to 1.0 everywhere else. Initially seeded uniformly over the patterned substrate (Fig. 2.A), the cell then migrate toward higher ligand density regions (Fig. 2.C).

2.2. Cellular migration in a strained micro-environment

In order to couple cell migration processes to deformation of their extracellular environment, it is necessary to propose stress-strain relationships modelling the visco-elastic properties of extracellular matrices.

2.2.a. Constitutive relationships for ECM

Stress-strain relationships defining the mechanical behaviour of the ECM may take several forms, from the simple linear stress-strain relationship (Hooke's law) characterising elastic materials up to nonlinear elasticity and/or viscosity behaviour (Benkherourou et al., 1999). We first considered a linear viscoelastic stress/strain relationship to approximate the biogel mechanical response to cell traction, with the viscous and elastic stresses added linearly. However, the original experiments of Harris et al. (1981) suggest an extension of this stress/strain relationship. Indeed, the merging movement of cells observed when two pieces of tissue explants are plated millimetres apart on distorted collagen gels indicate the ability of cells to interact across long distances. Thus, elastic matrices may represent an effective scaffold for long-range mechanical signalling pathways, those efficiency will partly depend on reticulated fibbers network which characterize ECM architecture. In order to model such long range effects of ECM deformations, second-order derivatives terms have been considered in the elastic stress tensor (Cruywagen et al., 1997). The resulting viscoelastic stress-strain relationship for the ECM is thus obtained from the sum of the following elastic and viscous tensor stresses:

$$\boldsymbol{\sigma}_{\text{elast}} = \frac{E}{(1+\nu)}\left(\boldsymbol{\varepsilon} - \beta_1\nabla^2\boldsymbol{\varepsilon} + \frac{\nu}{(1-2\nu)}(\theta - \beta_2\nabla^2\theta)\,\mathbf{I}\right)$$

$$\boldsymbol{\sigma}_{\text{visq}} = (\mu_1\frac{\partial\boldsymbol{\varepsilon}}{\partial t} + \mu_2\,\frac{\partial\theta}{\partial t}\mathbf{I})$$

where $\boldsymbol{\varepsilon}$ is the strain tensor, θ the dilation and $\mathbf{I}$ the unit tensor. The coefficients β_1 and β_2 are the long range elasticity coefficients, while E and ν are the ECM Young modulus and Poisson ratio, respectively. Parameters μ_1 and μ_2 are the shear and bulk ECM viscosities, respectively.

2.2.b. Modelling cell traction forces

Adherent cells exert substantial traction forces upon the surrounding ECM, mediated by specific adhesion sites (Balaban et al., 2001). In the light of experimental data, cell traction forces are assumed (i) to be proportional to the ligand or gel density (Delvoye et al., 1991; Barocas et al., 1995) and to plateau at high cell density because of cell contact inhibition phenomenon and of cell competition for ECM binding sites (Moon and Tranquillo, 1993; Ferrenq et al., 1997). Thus, we consider that active cellular traction stress $\boldsymbol{\sigma}_{\text{cell}}$ is modelled by:

$$\boldsymbol{\sigma}_{\text{cell}} = \tau\,\rho\,n\,(N_2 - n)\,\mathbf{I}$$

where the parameter τ monitors the cellular traction amplitude, while the positive constant N_2 controls the inhibition of cell traction as cell density increases. The two parameters τ and N_2 can be derived from experimental cell traction curves.

2.2.c Generic mechanobiological models

All the biochemical, mechanical and migrational processes described above are included explicitly within a mechanocellular model incorporating three variables: the cell density $n(\mathbf{x},t)$, the ECM density $\rho(\mathbf{x},t)$ and the displacement $u(\mathbf{x},t)$, (or strain $\varepsilon(\mathbf{x},t)$,) of the ECM at any time t and at any location **x** within the ECM.
The spatio-temporal evolution of cell density is given by the conservation equation:

$$\frac{\partial n}{\partial t} = -\nabla\cdot[\,\mathbf{J}_n + \mathbf{J}_d + \mathbf{J}_h\,]$$

where the flux $\mathbf{J}_n$ of cells passively convected at velocity $\partial\mathbf{u}/\partial t$ has been introduced, while the haptotactic cell flux now reads $\mathbf{J}_h = h\, n\, \nabla\rho$.
Taking into account convection, and neglecting ECM biosynthesis and proteolysis, the ECM conservation equation is simply:

$$\frac{\partial \rho}{\partial t} = -\nabla\cdot[\rho\,\frac{\partial \mathbf{u}}{\partial t}]$$

The cell-ECM mechanical equilibrium results from the sum of the various stresses derived before, in addition to body forces. Inertial terms are small in this context, and *in vitro* experiments are usually conducted by polymerizing biogels (collagen, fibrin, ...) in Petri dishes or wells. ECM gluing on dishes walls corresponds to a non-slip or zero displacement condition in a 3D domain. In a 2D approximation of the ECM medium, the ECM attachment at the dish bottom can be modelled by an elastic restoring force **R**. In this case, the displacement vector $\mathbf{u}(\mathbf{x},t)$ of the cell-ECM composite medium is given by the force balance equation:

$$\nabla\cdot\,[\boldsymbol{\sigma}_{ECM} + \boldsymbol{\sigma}_{cell}] = \mathbf{R}$$

The complete differential system associated to this generic mechanocellular model is given in appendix. In the following simulations of the model equations, we will assume as initial spatial conditions an homogeneous normalised value $\rho_0(\mathbf{x}) = 1$ of the ECM density. Initially, the ECM is in an unstressed state ($\boldsymbol{\varepsilon}(\mathbf{x})=\mathbf{0}$) and cells are randomly seeded, within a range $n_0(\mathbf{x}) = 1\pm0.1$. No-flux boundary conditions were considered for the cells and the ECM at the borders of the 2D domain, together with zero displacement conditions at these boundaries.
The corresponding nonlinear partial differential system was numerically solved on a 2D square by a finite elements method, using the Femlab© software.

2.3. Interactions between cellular diffusion and cellular traction.

In strained extracellular environment, one may expect modifications of the random walk of cells usually described by the standard expression of the cellular diffusion flux $\mathbf{J}_d$ given above. Cook (1995) proposed to account for the influence of ECM fibber orientation by considering a strain-biased diffusion tensor as the factor controlling the cellular diffusive flux, i.e.

$$\mathbf{J}_d\,(\Sigma) = -\nabla\cdot[\mathbf{D}(\Sigma)\,n]$$

where $\mathbf{D}(\Sigma)$ is a strain-biased diffusion tensor defined in a 2D space by:

$$\mathbf{D}(\boldsymbol{\varepsilon}) = D_0 \begin{bmatrix} 1+\dfrac{(\varepsilon_{xx}-\varepsilon_{yy})}{2} & \dfrac{(\varepsilon_{xy}+\varepsilon_{yx})}{2} \\ \dfrac{(\varepsilon_{xy}+\varepsilon_{yx})}{2} & 1-\dfrac{(\varepsilon_{xx}-\varepsilon_{yy})}{2} \end{bmatrix}$$

where D_0 is a scalar parameter and $\sum_{ij}$ are the strain tensor components.
Simulations account quite satisfactorily for cell mechanotaxis, where the extracellular substrate defines preferred directions for cell locomotion (Barocas and Tranquillo, 1997; Girton et al., 2002). In particular, Korff and Augustin (1999) performed experiments on a pre-stressed ECM and quantified the extent of anisotropic cell migration in vitro. These results are the experimental counterpart of strain-dependent diffusion simulated in figure 3.

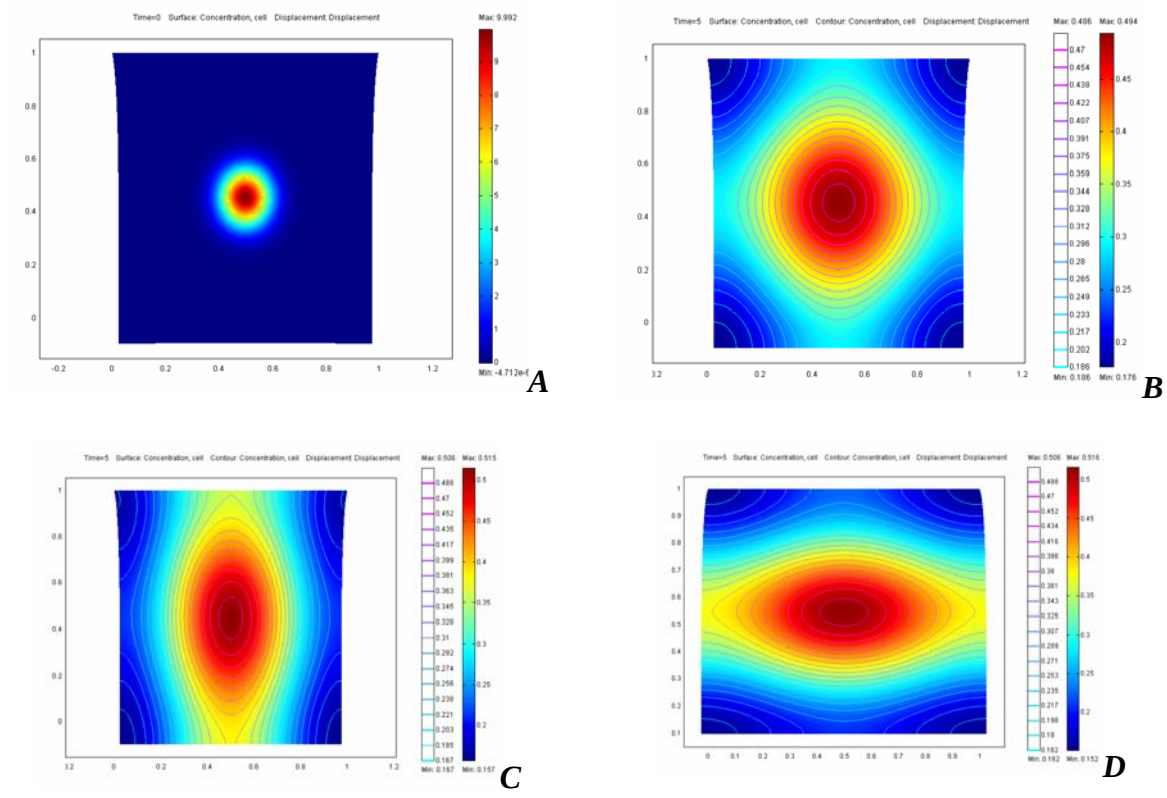


Figure 3: Comparison of cell density patterns obtained at the same given time when cellular diffusion occurs on a stretched (Fig.3B-C) or compressed (Fig.3D) elastic substrate. In figure 3.B isotropic cellular diffusion is assumed, while strain-dependant diffusion is simulated in figures 3.C and 3.D. Figure 3.A shows the initial cell density distribution.

2.4. Interactions between haptotactic cell migration and cellular traction.

As a first example, let us consider the situation depicted in figure 2, in which the inclusion of square-shaped adhesive pattern has been laid down on the elastic 2D substrate. The coupling of cell migration by haptotaxis and cell traction forces induce as cell aggregation on the squared-frame, but now cells are located in a narrower contracted and deformed stripe of extracellular material, with higher cell densities (Fig. 4). However, less obvious patterns could also emerge from instabilities created by the coupling between haptotaxis and cellular forces beyond critical thresholds, as it will be shown in the following paragraph.

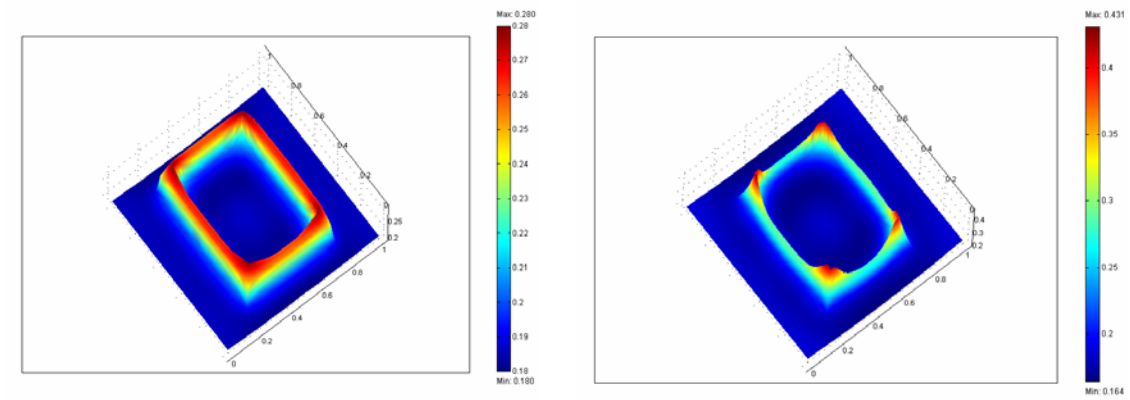


Figure 4: Comparison of cell density patterns obtained with the patterned adhesive substrate of figure 2. In Fig. 4A, elastic substrate properties are as in figure 2, while a 25% decrease of the substrate Young's modulus has been considered in Fig. 3B. In this case, cell aggregation is strongly enhanced in the corners of the more adhesive regions.

3. Mechano-biological models for a qualitative and quantitative analysis of strain-induced cellular network morphogenesis.

Considering the modeling framework defined above, we now illustrate, in the *in vitro* angiogenesis context, how the association of cell motility with the self-enhancement of the local traction fields exerted within the biogel, may explain the pre-patterning of the biogel and subsequent formation of capillary-like networks (CLN).

3.1. Some key experimental features

Figure 5 provides typical views of the CLN formation process during *in vitro* angiogenesis assays conducted with EAhy926 endothelial cells on fibrin gels, an extracellular matrix of special interest for studying angiogenesis (Van Hinsbergh et al., 2001).

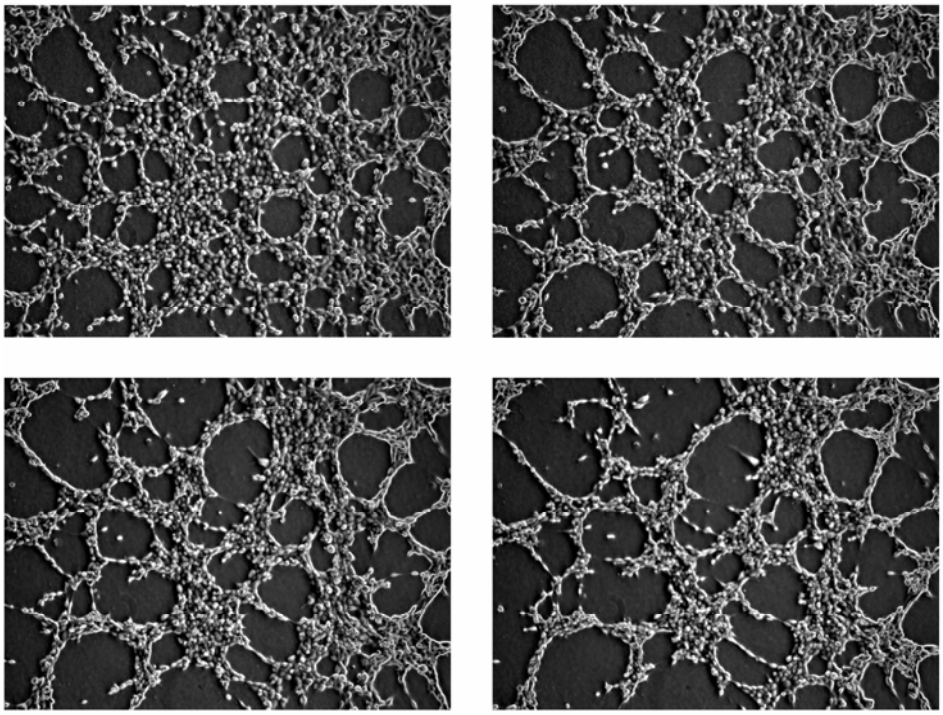


Figure 5: Morphogenesis of endothelial cell networks at successive time as revealed by phase-contrast time-lapse vide- microscopy. EAhy926 endothelial cells were seeded uniformly on a fibrin gel. Time interval between images is of 5h, with image size of 1,8 x 1,3 mm.

Several papers on *in vitro* angiogenesis assays reported that that tubulogenesis of endothelial cells within biogels only appears for critical range of values of the experimental parameters. In rather extensive experiments performed with Human Umbilical Vein Endothelial Cells (HUVEC) cultured on fibrin gels, Vailhé et al. quantified the range of experimental parameters for which CLN formation can be observed (Vailhé et al., 1997, 1998). Briefly, they show that:

-) a critical number of seeded cells is required to induce CLN: in a medium inducing minimal cell growth (2% human serum), CLN formation is strongly dependent on cell number: between 1.5 x 10^5 cells/ml and 2x 10^5 cells/ml, 75% of the surface of the Petri dishes was occupied by a cellular network. Below or above these values, CLN formation was completely inhibited.

-) a critical fibrin gel stiffness is required to induce CLN: when *in vitro* angiogenesis assays were conducted with fibrin concentration increasing from 0.5 mg/ml to 2 mg/ml, i.e. on biogels with decreasing elasticity, the number of lacunae decreased and they neither evolved to CLN after 24 to 48 hours observation.

In order to quantify the extent of the CLN network, these authors defined as an angiogenic or network index the ratio of the area occupied by cellular cords over the total biogel area. Using these data, we extrapolated the experimental bifurcation diagram of figure 6, which exhibits the bell-shaped curve for the angiogenic CLN index.

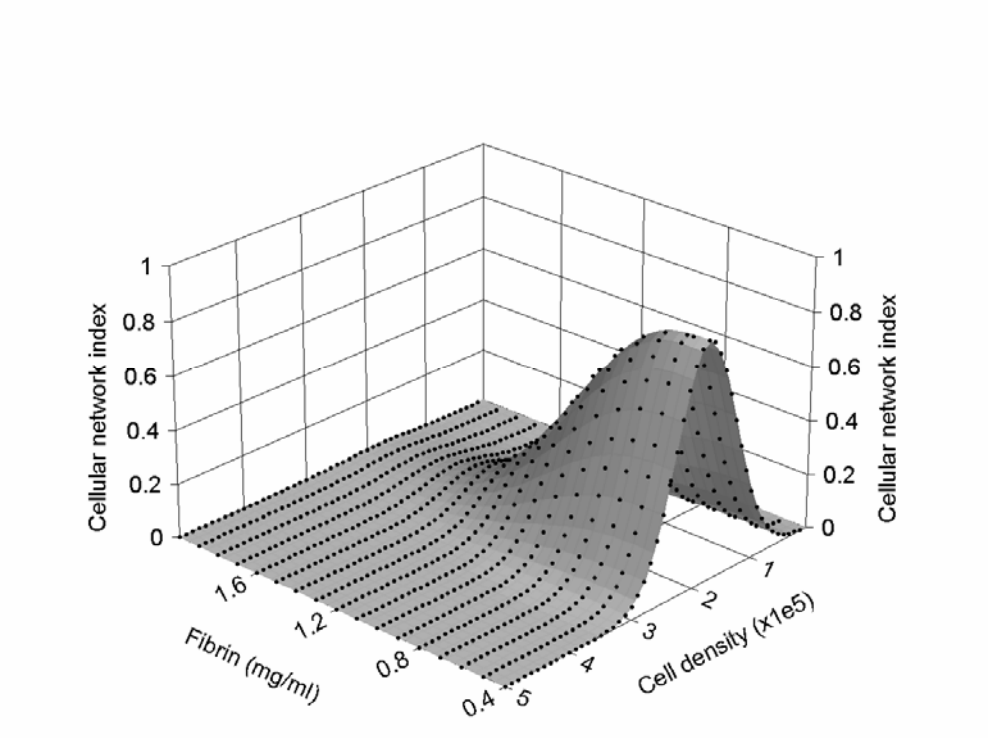


Figure 6: Experimental bifurcation diagramconstructed by extrapolation of the set of data obtained by Vailhé et al. (1997) when fibring gel concentration and seeded cell densities are varied. CLN formation only occurs for a restricted range of these experimental parameters.

3.2. Theoretical foundations for understanding CLN formation process

Following the initial work of Manoussaki et al., (1996), we analyzed more precisely how ECM mechanical properties could be crucial in organizing the collective behavior of endothelial cells in the successive phases of in vitro tubulogenesis. Starting from a one-dimensional model which quantitatively reproduces the range of critical values of cell densities and matrix stiffness for which CLN are experimentally observed (Tranqui and Tracqui, 2000), we developed a more complete 2D theoretical model which accounts for the experimental bifurcation scenario leading to biogel patterning and CLN morphogenesis (Namy et al., 2004). We will review here the main results which validate the proposed model as a candidate for explaining the critical experimental thresholds beyond which a spatial pre-structuration of the fibrin gel can be generated and CLN formed.

3.2.a. Origin of the instabilities

The emergence of non-equilibrium steady-state of cells and ECM densities corresponds mathematically to a bifurcation process, but which originate from different physico-biological processes than those acting in reaction-diffusion (RD) models or in reaction-diffusion-migration (RDM) models of pattern formation. In the original Turing's model, the so-called Turing's instability arises from the coupling between nonlinear reaction kinetics and the relative diffusion of the two competing (inhibitor and activator) chemical species. In reaction-diffusion-migration models (RDM) models, simplest scenarios involve cells which produce their own chemoattractant, leading to an aggregative effect as cells move up concentrations of the chemoattractant (chemotaxis). Considering the two coupled conservation equations for the cells and the chemoattractant, it is intuitively clear how cellular spatial patterns can be formed: since more cells produce more attractant, a small perturbation in an initially

homogeneous distribution of cells can give rise to a gradient in the chemoattractant. The gradient of attractant induces more cells to move towards the higher concentration of cells, thus enhancing the heterogeneity of cell density. On the other hand, the random component of cell migration, i.e. cellular diffusion, and the chemoattractant diffusion tend to eliminate such heterogeneities. Clearly, spatial patterns will be generated if the aggregative process (chemotaxis) is greater than the dispersive effects.
In the mechanobiological framework considered here, in which mechanical cell-ECM interactions are involved, the aggregative effect is monitored by adhesivity gradients (haptotaxis), created by inhomogeneous ECM densities. This aggregative effect is still counterbalanced by cellular diffusion, but since ligands are bounded to the ECM, there is no equivalent for chemoattractant diffusion. It is only the ECM mechanical properties which can act as a stabilizing factor in this ago-antagonistic scenario.

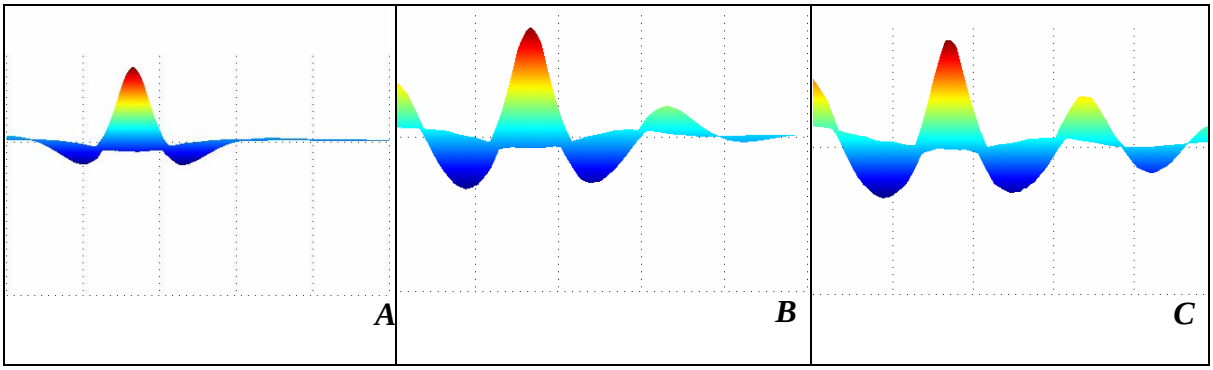


Figure 7: One-dimensional profiles of cellular density at successive normalized times (0.1, 0.3, 0.6) along one space dimension, starting from homogeneous initial cell density seeded on a viscoelastic substrate. Cell aggregation is initiated on the left side by a localised perturbation of the ligands density (Fig. 7A). The two following views show the propagation of the long-range re-arrangement of cell density due to the substrate mechanical deformation (Fig. 7B, 7C).

3.2.b. Critical parameter thresholds for CLN morphogenesis

The above qualitative explanation of the mechanically-induced instability process can be turned into a precise quantitative formulation when undertaking a linear stability analysis of the complete partial differential system associated to the modelling of the CLN formation. The normalized form of this system, derived from the balance equations presented in section XX, is given in appendix.
Using the cell traction force coefficient τ as a bifurcation parameter, the linear stability analysis performed in a neighbourhood of the normalized homogeneous steady state $n(\mathbf{x},0)=1$, $\rangle(\mathbf{x},0)=1$, $\mathbf{u}(\mathbf{x},0)=\mathbf{0}$ establishes that non-homogeneous steady-states can exist when $\tau>\tau_c$, where the critical value $/_c$ of the cell traction amplitude is given by (Namy et al., 2004):

$$\tau_c = \frac{(\lambda+2\mu)+2\sqrt{s(\lambda\beta_2+2\mu\beta_1)}}{\frac{h}{D_0}(N_2-2)+0.5(3N_2-4)}$$

The value $/_c$ appears as the ratio of two terms which highlight the competitive effects linked to the biogel mechanical properties on one hand, to the ratio between haptotactic migration and cellular diffusion on the other hand. Indeed,

-) the numerator contains the mechanical stabilising parameters coming from ECM elasticity moduli (or Lamé coeffcients) λ and μ, in addition to a combination of these moduli with long range elasticity parameters β_1 and β_2, weighted by the value of restoring force parameter s. Thus, pre-pattern formation within the biogel can be reached for lowest cell traction when the elastic mechanical resistance of the biogel is low, i.e. for softer biogels.

-) the denominator contains the stabilising effect of cell traction inhibition, monitored by parameter N_2. Moreover, the balance between cell migration by haptotaxis and diffusion, i.e; the ratio h/D_0, is a key factor for driving the mechanical instability process. Efficient haptotactic cell migration lowers the critical threshold for instability since for positive values of h, a weaker cell traction amplitude is required to induced heterogeneous spatial cell aggregation and modelling of the ECM.

As a whole, the coupling between cellular traction and cell haptotaxis functions as a "mechanical autocatalysis" by inducing a nonlinear type “strain-induced strain increase” (SISI) mechanism.

3.2.c. Simulation of CLN morphogenesis

Figure 8 illustrates the morphogenesis of CLN for a cell traction amplitude τ_1 above the critical bifurcation value τ_c defined above. Other model parameter values are chosen in order to simply get several lacunae within the biogel.

The biogel thickness e(**x**,t) at any location **x** of the 2D domain varies linearly with the biogel density ρ(**x**,t), according to the following relationship, derived from the Hooke's law (Namy et al., 2004):

$$e(\mathbf{x},t) = e_0(\mathbf{x})\left(\frac{(1-3\nu)}{(1-2\nu)} + \frac{\nu}{(1-2\nu)}\frac{\rho(\mathbf{x},t)}{\rho_0(\mathbf{x})}\right)$$

where $e_0(\mathbf{x})$ denotes the initial biogel thickness.

The initially smooth gel density progressively deforms (Fig. 8A), up to the formation of pronounced depressions within the biogel, these lacunae being separated by sharp "walls" of extracellular material (Fig. 8B). According to EC type and fibrin concentration, significant differences in time scales have been reported for the development of *in vitro* tubulogenesis (Vernon et al., 1992; Vailhe et al., 1998, 2001). For HUVEC cultured on fibrin gel, lacunae appear typically in less than 2 hours while cellular network formation is achieved after 24 hours, with lacunae size ranging from 0.2mm to 1.5mm. For the experimentally based parameter values we considered, simulated lacunae morphogenesis takes place within twelve hours, while the complete CLN organization is achieved later, according to a network topology very similar to the experimental observations (Fig. 8C,D).

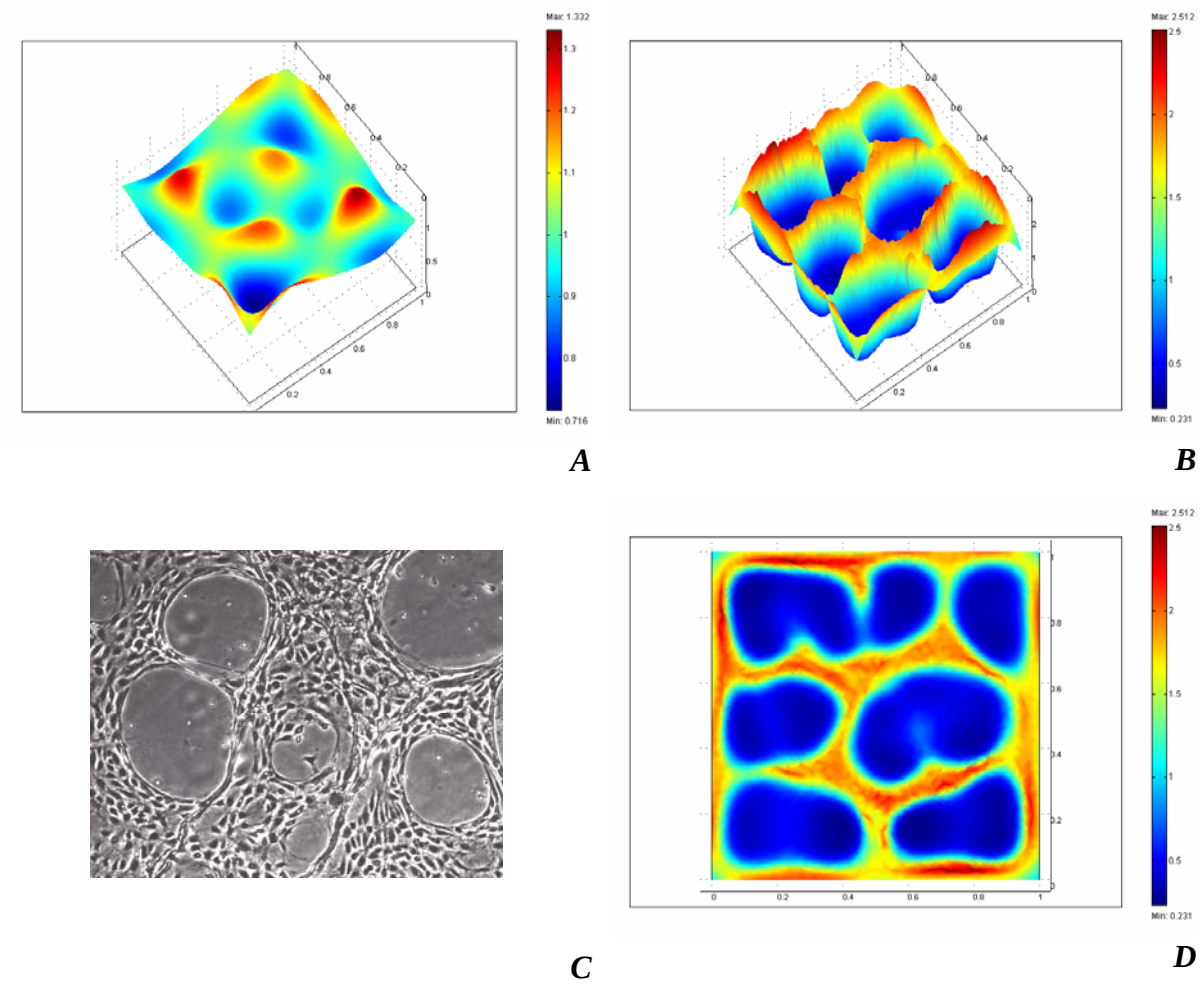
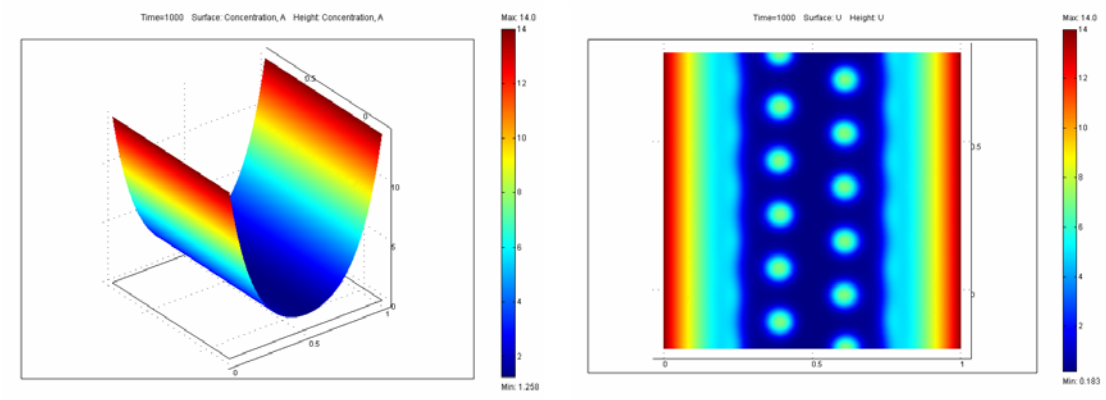


Figure 8: Simulated CLN formation on viscoelastic biogels. The 2D profiles of biogel density, which correlates linearly with the biogel thickness, show the progressive morphogenesis of the lacunae (Fig. 8A) and of the associated CLN pattern (Fig. 8B). The upper view of the simulated biogel patterning (Fig. 8D) compares quite well with EAhy926 endothelial cell organisation on fibrin gels observed by phase-contrast microscopy (Fig. 8C).

3.3. Localized morphogenesis of spatial structures

It is known that morphogenesis based on reaction-diffusion processes can lead to localized spatial structures when some reactants are non-uniformly distributed in space. Considering for example the well known Brusselator model (Nicolis and prigogine, 1977), figure 9.B illustrates the emergence of localized variations of products concentrations when the feeding of the nonlinear reaction is controlled by a non uniform distribution of the substrate (Fig. 9A).

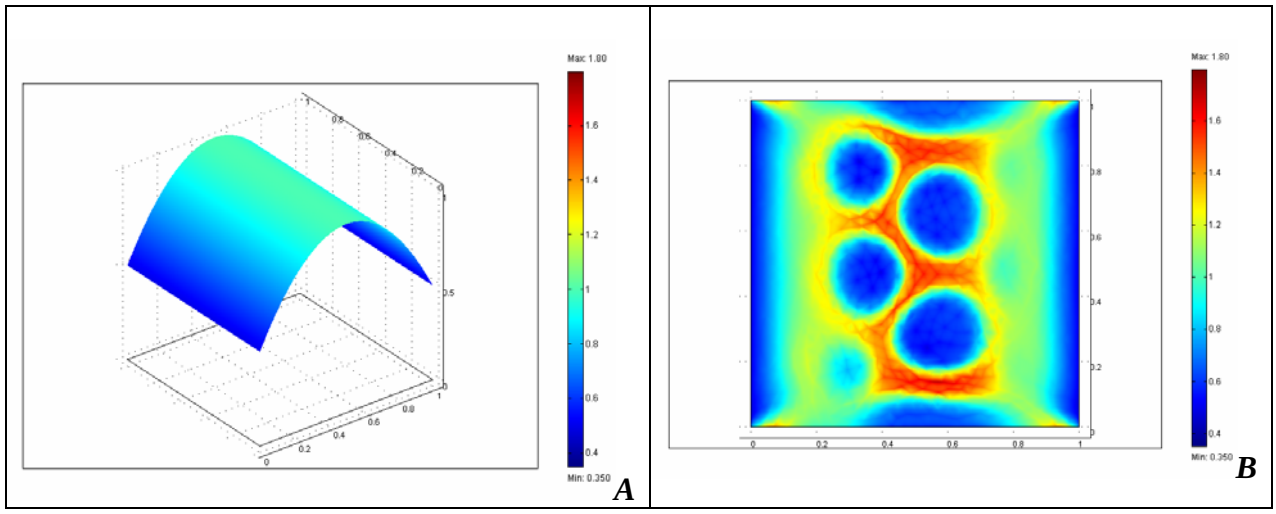


A ***B***

Figure 9: Simulated localized patterns with the Brusselator model with a spatially non uniform influx of reactant.

One may expect similar qualitative results when the pattern formation process is driven by mechanical interactions between cells and ECM. Indeed, the lacunae patterns described in the preceding sections extend throughout the whole domain as a consequence of assuming a uniform medium where ECM density or thickness is initially constant in space. If it is not the case, i.e. if ECM thickness is not initially uniform, this spatial inhomogeneity could result in the localization of the lacunae within a sub-region of the original domain.

In order to analyze how the ECM thickness may contribute to the localized formation of lacunae within the ECM, we consider an initial bell-shaped variation of the ECM thickness along one side of the 2D domain (Fig. 10A). According to our model formulation, this spatial variation of ECM thickness corresponds to a similar variation of ECM density and modulates the elastic restoring body force **R**, whose amplitude decreases with increasing ECM thickness. The simulated spatial organization of the steady-state cell density is shown in Figure 10B: lacunae are only observed in the middle of the ECM, i.e. where the ECM is initially thicker.



A ***B***

Figure 10: Simulated formation of localized lacunae within the ECM (Fig. 10B) when the initial profile of the ECM thickness is the one given in Fig. 10A.

Remarkably, this type of simulated cellular spatial organization reproduced quite well the experimental observations of Vernon et al. (1992), who reported the influence of gradients of ECM thickness on the formation of cellular networks for bovine aortic cells cultured on basement membrane matrix with a thickness gradient. In these experiments, no lacunae appeared in the thinnest ECM regions, while lacunae with increasing size were observed in pace with increasing matrix thickness.

4. Discussion

While the theoretical background for understanding biological morphogenesis is still largely dominated by the reaction-diffusion based Turing's paradigm, considering self-organized morphogenesis driven by cell-ECM mechanical interactions appears clearly as a very promising extended theoretical framework. The pioneering view of J.D. Murray and G. Oster in the eighties is indeed largely supported today by an increasing corpus of experimental data which underline the central role of mechanical factors in the regulation of cell dynamics and cellular organisation processes.

In vivo angiogenesis assays especially established the self-organising properties of endothelial cells which can re-arrange in tubular structures after different spatial organisation stages. Thus, these assays retain most of the complexity of the first stages of the neo-vascularisation process which may take place in an extracellular matrix, while shedding light on the various cell-extracellular matrix interactions (Ingber, 2002). Indeed, both composition and mechanical properties of ECM are important in the initiation of angiogenesis, even if some additional extracellular factors, like VEGF, may come into play (Gamba et al., 2003; Serini et al., 2003).

The combined experimental-theoretical approach presented here explicits the role of each factors within an integrated description of cell-ECM interactions, especially the strength of endothelial cells traction forces. These forces remodel the surrounding ECM through the reorganization of its structure by creating tension lines and structural pathways which can provide migration cues for other cells. However, the effect of strain-induced anisotropic cellular diffusion we investigated appears to be too weak to influence qualitatively the lacunae network pattern formation.

The proposed theoretical model provides estimation of the threshold values of the experimental variables and a direct analysis of the experimental curves defining the limits within which lacunae and CLN formation can take place. Our present analysis concerns the first steps of the lacunae formation within the biogel, in the limit of the small strain formulation of the stress-strain relationships. However, incorporation of new experimental factors within the model (cell mitosis and apoptosis, ECM biosynthesis, …) or sub-integrated within an already modelled process (regulation factors of fibrinolysis or cell migration,…) can be directly considered in this framework.

Because tractions forces concentrate in the cell lamellipodia (Dembo and Wang, 1999; Pelham and Wang, 1999), these actin-riched cell structures are the key elements of an effective mechanically-driven cell locomotion system in which cells send out membrane protrusions to probe the mechanical properties of their micro-environment. However, it remains unclear how cells translate substrate rigidity into downstream responses. Current interpretations suggest that the stronger mechanical feedback on stiff substrates may then lead to the activation of stress-sensitive ion channels (Lee et al., 1999) or conformational changes of other tension-sensitive proteins. These responses in turn may regulate the extent of protein tyrosine phosphorylation (Pelham and Wang, 1997) and the stability and size of focal adhesions (Riveline et al., 2002).

The many different control points in the cell-ECM signalling pathways which can be addressed by the in vitro angiogenesis assay make it an excellent system for investigating how cells process mechanically driven contextual information. The modelling approach presented here can be a guide for further multi-scale approaches which should play a significant role in understanding complex biological processes as integration of signal transduction and developmental patterning.

Acknowledgments

We thank Caroline Rosello, Emmanuelle Planus and Jocelyne Clément-Lacroix for their contributions to the *in vitro* angiogenesis experiments.

Appendix

The mechano-cellular model equations can be non-dimensionalised by re-scaling time and considering the diameter L of the Petri dish as a typical length scale. The following normalisation was used:

$$t^* = \frac{t}{T} \quad ; \; x^* = \frac{x}{L}; \; y^* = \frac{y}{L}; \; u_1^* = \frac{u_1}{L}; \; u_2^* = \frac{u_2}{L}; \quad n^* = \frac{n}{n_0} \quad ; \; \rho^* = \frac{\rho}{\rho_0} \quad ; N_2^* = \frac{N_2}{n_0};$$

$$D^* = \frac{D_0 T}{L^2}; \quad h^* = \frac{h\, T\, \rho_0}{L^2} \quad ; \gamma^* = \gamma T (n_0)^2; \quad \tau^* = \frac{\tau (n_0 \rho_0)^2}{L^2}$$

$$\mu_1^* = \frac{\mu_1 \rho_0}{T\, L^2}; \; \mu_2^* = \frac{\mu_2 \rho_0}{T\, L^2}; \; \beta_1^* = \frac{\beta_1}{L^2}; \; \beta_2^* = \frac{\beta_2}{L^2};$$

$$E^* = \frac{E(1-\nu)\rho_0}{L^2(1+\nu)(1-2\nu)}; \; \lambda^* = \frac{E\nu\rho_0}{L^2(1+\nu)(1-2\nu)}; \; \mu^* = \frac{E\rho_0}{2L^2(1+\nu)}$$

The mechano-cellular model we analysed is defined by the following set of conservation and force balance equations, where the asterisks have been dropped for convenience:

$$\left|\begin{array}{l} \dfrac{\partial n}{\partial t} + \nabla.\left[-\nabla.\left(\boldsymbol{D}(\boldsymbol{\varepsilon}) n + h n \nabla\rho + n \dfrac{\partial \boldsymbol{u}}{\partial t} \right) \right] = 0 \\ \\ \dfrac{\partial \rho}{\partial t} + \nabla.\left[\rho \dfrac{\partial \boldsymbol{u}}{\partial t} \right] = 0 \\ \nabla.\left[\left\{ 2\mu\left(\boldsymbol{\varepsilon} - \beta_1 \nabla^2 \boldsymbol{\varepsilon}\right) + \lambda\left(\theta - \beta_2 \nabla^2 \theta\right)\boldsymbol{I} \right\} + \left\{ \mu_1 \dfrac{\partial \boldsymbol{\varepsilon}}{\partial t} + \mu_2 \dfrac{\partial \theta}{\partial t} \boldsymbol{I} \right\} + \left\{ \tau\rho n\left(N_2 - n\right)\boldsymbol{I} \right\} \right] = s\, \boldsymbol{u}\, /\rho \end{array}\right.$$

where bold face characters denote vectors or tensors, $\boldsymbol{I}$ being the unit tensor and $\boldsymbol{\varepsilon}$ the strain tensor. In the last force balance equation, the divergence operator is applied to three terms (between brackets) which correspond to – the elastic ECM response and – the viscous ECM response to the – active cellular stress, respectively.

REFERENCES

Balaban, N. Q., Schwarz, U. S., Riveline, D., Goichberg, P., Tzur, G., Sabanay, I., Mahalu, D., Safran, S., Bershadsky, A., Addadi, L. & Geiger B. (2001) Force and focal adhesion assembly: a close relationship studied using elastic micropatterned substrates. Nat. Cell Biol. 3,466-472.

Barocas, V. H. & Tranquillo, R. T. (1997) An anisotropic biphasic theory of tissue-equivalent mechanics: the interplay among cell traction, fibrillar network deformation, fibril alignment, and cell contact guidance. J. Biomech. Eng. 119, 137-145.

Barocas, V. H., Moon, A. G. & Tranquillo, R. T. (1995) The fibroblast-populated collagen microsphere assay of cell traction force--Part 2: Measurement of the cell traction parameter. J. Biomech. Eng. 117, 161-170.

Benkherourou, M., Gumery, P. Y., Tranqui, L. & Tracqui, P. (2000) Quantification and macroscopic modeling of the nonlinear viscoelastic behavior of strained gels with varying fibrin concentrations. IEEE Trans. Biomed. Eng. 47, 1465-1475.

Brouzes E, Farge E.,(2004) Interplay of mechanical deformation and patterned gene expression in developing embryos. Curr Opin Genet Dev.14(4):367-74.

Choquet D., Felsenfeld D. P., Sheetz M. P., Extracellular matrix rigidity causes strengthening of integrin-cytoskeleton linkages, Cell 88 (1997) 39-48.

Cook, J. (1995) Mathematical models for dermal wound healing: wound contraction and scar formation. Ph.D. thesis (University of Washington, Seattle) pp. 98-133.

Cruywagen, G. C., Maini, P. K. & Murray, J. D. (1997) Biological Pattern Formation on Two-Dimensional Spatial Domains: A Nonlinear Bifurcation Analysis. SIAM J. Appl. Math. 57, 1485-1509.

Delvoye, P., Wiliquet, P., Leveque, J. L., Nusgens, B. V. & Lapiere C. M. (1991) Measurement of mechanical forces generated by skin fibroblasts embedded in a three-dimensional collagen gel. J. Invest Dermatol. 97, 898-902.

Dembo M, Wang YL.,(1999) Stresses at the cell-to-substrate interface during locomotion of fibroblasts. Biophys J. 76(4):2307-16.

Ferrenq, I., Tranqui, L., Vailhe, B., Gumery, P. Y. & Tracqui, P. (1997) Modelling biological gel contraction by cells: mechanocellular formulation and cell traction force quantification. Acta Biotheor. 45, 267-293.

Folkman, J. and Haudenschild, C. (1980) Angiogenesis in vitro. Nature (London) 288, 551-556.

Gamba, A., Ambrosi, D., Coniglio, A., de Candia, A., Di Talia, S. Giraudo, E., Serini, G., Preziosi, L. & Bussolino, F. (2003) Percolation, morphogenesis and burgers dynamics in blood vessels formation. Phys. Rev. Lett. 90, 118101-1-118101-4.

Gerhardt, H., Golding, M., Fruttiger, M., Ruhrberg, C., Lundkvist, A., Abramsson, A., Jeltsch, M., Mitchell, C., Alitalo, K., Shima, D. & Betsholtz, C. (2003) VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. J. Cell Biol. 161, 1163-1177.

Girton, T. S., Barocas, V. H. & Tranquillo, R. T. (2002) Confined compression of a tissue-equivalent: collagen fibril and cell alignment in response to anisotropic strain. J. Biomech. Eng. 124, 568-575.

Harris A K, Stopak D, Wild P. Fibroblast traction as a mechanism for collagen morphogenesis. *Nature*1981 **290** : 249-251

Ingber, D. E. (2002) Mechanical signaling and the cellular response to extracellular matrix in angiogenesis and cardiovascular physiology. Circ. Res. 91, 877-887.

Korff, T. & Augustin, H. G. (1999) Tensional forces in fibrillar extracellular matrices control directional capillary sprouting. J. Cell Sci. 112, 3249-3258.

Lo CM, Wang HB, Dembo M, Wang YL. (2000) Cell movement is guided by the rigidity of the substrate. Biophys J. 79(1):144-52.

Mandeville JT, Lawson MA, Maxfield FR (1997) Dynamic imaging of neutrophil migration in three dimensions: mechanical interactions between cells and matrix. J Leukoc Biol. 61(2):188-200.

Manoussaki, D., Lubkin, S. R., Vernon, R. & Murray, J. D. (1996) A mechanical model for the formation of vascular networks in vitro. Acta Biotheor. 44, 271-282.

Meinhardt H. (1989); Models for positional signalling with application to the dorsoventral patterning of insects and segregation into different cell types, Development, (1989 Suppl.), p. 169-180,

Meinhardt H. (1995), The algorithmic beauty of seashells, Springer Verlag,

Meinhardt H, de Boer PA. (2001) Pattern formation in Escherichia coli: a model for the pole-to-pole oscillations of Min proteins and the localization of the division site. Proc Natl Acad Sci U S A. 98(25):14202-7.

Murray, J. D. & Oster, G. F. (1984) Cell traction models for generating pattern and form in morphogenesis. J. Math. Biol. 19, 265-279.

Murray, J. D. (2003) Mathematical Biology, Vol. I, II New York: Springer.

Namy P., J. Ohayon and P. Tracqui (2004) Critical Conditions for Pattern Formation and in vitro Tubulogenesis Driven by Cellular Traction Fields, *J. Theor. Biol.* 227:103-120.

Nicolis G. et I. Prigogine ; Self Organization in Nonequilibrium Systems, J.Wiley & Sons, 1977.

Oakley C, Jaeger NA, Brunette DM. (1997)Sensitivity of fibroblasts and their cytoskeletons to substratum topographies: topographic guidance and topographic compensation by micromachined grooves of different dimensions. Exp Cell Res. 234(2):413-24.

Pelham RJ Jr, Wang Y.(1997) Cell locomotion and focal adhesions are regulated by substrate flexibility. Proc Natl Acad Sci U S A. 94(25):13661-5.

Pelham RJ Jr, Wang Y.(1999) High resolution detection of mechanical forces exerted by locomoting fibroblasts on the substrate. Mol Biol Cell. 10(4):935-45.

Riveline D, Zamir E, Balaban NQ, Schwarz US, Ishizaki T, Narumiya S, Kam Z, Geiger B, Bershadsky AD.(2001) Focal contacts as mechanosensors: externally applied local mechanical force induces growth of focal contacts by an mDia1-dependent and ROCK-independent mechanism. J Cell Biol. 153(6):1175-86.

Serini, G., Ambrosi, D., Giraudo, E.,Gamba, A., Preziosi, L. & Bussolino, F. (2003) Modeling the early stages of vascular network assembly. EMBO J. 22, 1771-1779.

Sheetz MP, Felsenfeld DP, Galbraith CG. (1998) Cell migration: regulation of force on extracellular-matrix-integrin complexes. Trends Cell Biol.8(2):51-4.

Tabony J, Glade N, Papaseit C, Demongeot J. (2002) Microtubule self-organisation and its gravity dependence. Adv Space Biol Med. 8:19-58.

Thompson D.W., (1917) On growth and Form, Cambridge University press

Tranqui, L. and Tracqui, P. (2000) Mechanical signalling and angiogenesis. The integration of cell-extracellular matrix couplings. C. R. Acad. Sci. 323, 31-47

Trinkhaus J. P., (1984) Cells into organs. The forces that shape the embryo, 2nd edition, Prentice-Hall, Englewood.

Turing A. M.(1952) ; The chemical basis for morphogenesis, Philos. Trans. R. Soc. Lond. B, 237, p. 37-72, 1952.

Vailhe, B., Lecomte, M., Wiernsperger, N. & Tranqui, L. (1998) The formation of tubular structures by endothelial cells is under the control of fibrinolysis and mechanical factors. Angiogenesis 2, 331-344.

Vailhe, B., Ronot, X., Tracqui, P., Usson, Y. & Tranqui, L. (1997) In vitro angiogenesis is modulated by the mechanical properties of fibrin gels and is related to $\alpha_v\beta_3$ integrin localization. In Vitro Cell. Dev. Biol. Anim. 33, 763-773.

Vailhe, B., Vittet, D. & Feige, J.-J. (2001) In vitro models of vasculogenesis and angiogenesis. Lab. Invest. 81, 439-452.

Van Hinsbergh, V. W., Collen, A. & Koolwijk, P. (2001) Role of fibrin matrix in angiogenesis. Ann. N. Y. Acad. Sci. 936, 426-437.

Verkhovsky AB, Svitkina TM, Borisy GG. (1999) Self-polarization and directional motility of cytoplasm. Curr Biol. 9(1):11-20.

Vernon, R. B., Angello, J. C., Iruela-Arispe, M. L., Lane, T. F. & Sage, E. H. (1992) Reorganization of basement membrane matrices by cellular traction promotes the formation of cellular networks in vitro. Lab. Invest. 66, 536-547.

Vernon, R. B., Lara, S. L., Drake, C. J., Iruela-Arispe, M. L., Angello, J. C., Little, C. D., Wight, T. N. & Sage, E. H. (1995) Organized type I collagen influences endothelial patterns during "spontaneous angiogenesis in vitro": planar cultures as models of vascular development. In vitro Cell. Dev. Biol. 31, 120-131.

Waddington C. H. (1968) Theoretical biology and molecular biology, dans : Towards a Theoretical Biology . I., edited by C. H. Waddington, Edimburg University Press, p. 103-108.

Part 2

LECTURES / ABSTRACTS

The evolutionary forces that shape genetic networks

Andreas Wagner*

*Associate Professor of Biology, Univ. of New Mexico External Faculty, The Santa Fe Institute

Abstract

We know that natural selection has influenced many features of living organisms, both on the level of individual genes and on the level of whole organisms. However, we know little about the influence of natural selection on the intermediate level of genetic networks. I here ask on what scale of organization natural selection molds the structure of molecular networks, on the largest scale, such as that of genome-scale molecular interaction or regulatory networks, on the intermediate scale of small regulatory circuits, or on the smallest scale, that of individual network parts? I discuss this question in the context of three biological networks, metabolic networks, transcriptional regulation networks, and protein interaction networks.

Part 3

POSTERS

Automatic Tuning of Agent-Based Models using Genetic Algorithms

Benoît Calvez & Guillaume Hutzler

LaMI Université d'Évry Val d'Essonne & CNRS UMR 8042, Tour Évry 2, 523 Place des terrasses de l'agora, 91000 Évry
{bcalvez,hutzler}@lami.univ-evry.fr

Abstract

When developping multi-agent systems (MAS) or models in the context of agent-based simulation (ABS), the tuning of the model constitutes a crucial step of the design process. Indeed, agent-based models are generally characterized by lots of parameters, which together determine the global dynamics of the system. Moreover, small changes made to a single parameter sometimes lead to a radical modification of the dynamics of the whole system. The development and the parameter setting of an agent-based model can thus become long and tedious if we have no accurate, automatic and systematic strategy to explore this parameter space.
That's the development of such a strategy that we work on, suggesting the use of genetic algorithms. The idea is to capture in the fitness function the goal of the design process (efficiency for MAS that realize a given function, realism for agent-based models, etc.) and to make the model automatically evolve in that direction. However the use of genetic algorithms (GA) in the context of ABS raises specific difficulties that we develop in this article, explaining possible solutions and illustrating them on a simple and well-known model: the food-foraging by a colony of ants.
We apply the method to a more complex example. We work on the simulation of the glycolysis and the phosphotranferase systems in Escherichia coli. In this work, we are interested in testing the hypothesis of hyperstructures, which are believed to improve the behavior of a cell. We try to determine under what conditions this may be true, and how these hyperstructures may function.

1 Introduction

Agent-based simulation (ABS) is interested in the modelling and the simulation of complex systems. Its aim is to reproduce the dynamics of real systems by modelling the entities as agents, whose behavior and interactions are defined. A first validation of such models is obtained by comparing the resulting dynamics, when the model is simulated, with that of the real system (measured thanks to experimental data). Similarly, Multi-Agent Systems (MAS) are designed so as to accomplish a given function in a collective and decentralized way. The validation of the system is thus given by the fact that the function is realized and that it is efficient. In both cases, one of the crucial aspects of the design process lies in the tuning of the model. Indeed, this kind of model is generally characterized by lots of parameters which together determine the global dynamics of the system. The search space is thus gigantic. Moreover, the behavior of these complex systems is often chaotic: on the one hand small changes made to a single parameter sometimes lead to a radical modification of the dynamics of the whole system; on the other hand some emergent phenomena are only produced in very specific conditions and won't occur if these conditions are not met. The solution space can thus be very small. As a consequence, the development and the parameter setting of an agent-based model may become long and tedious if we have no accurate, automatic and systematic strategy to explore the parameter space.
The approach that we suggest is to consider the problem of the development and the validation of ABS or MAS models as an optimization problem. The validation can thus be reformulated as the identification of a parameter set that optimizes some function. The optimization function for ABS would be the distance between the artificial model that we simulate and the real system. The

optimization function for MAS would be the efficiency in the realization of the function. Given the large dimensionality of the problem, optimization techniques such as Genetic Algorithms (GA) can then be used to explore the parameter space and find the best parameter set with respect to the optimization function. However the use of genetic algorithms in this context is not so simple, as we will explain.
In section two we present the problematics related to the parameter tuning of an agent-based simulation. Then in section three we present the general framework of genetic algorithms and show the difficulties that arise from the application of these techniques to agent-based simulation.

2 Parameter tuning

2.1 Parameters of agent-based models

In the context of agent-based simulation, a model and the simulator with which it is executed include lots of parameters. These parameters can be of different natures. Some parameters are peculiar to the simulator: the discretization step for the modeling of time and space for instance can be a fixed feature of the simulator. As a consequence, these parameters can generally not be modified by the user. For this reason, we do not include this type of parameters in our parameter space. We only include the parameters that are specific to the model. Some of them can be extracted from the knowledge of the field (either experimental or theoretical) and can thus be associated to fixed values. Other parameters have to be kept variable, which can be for different reasons: on the one hand, the knowledge of the field is generally not exhaustive (which is the reason why we build a model and simulate it); on the other hand, this knowledge may not be directly compatible with the model. In this case, a common approach can be to try some values and simulate the model to see how it behaves globally. What we propose is to have a general approach to automate this long and tedious process.

2.2 Objective

Depending on the motivation of the modeling work, the criteria used to explore the parameter space will also be different. This motivation may be to model and simulate a real system, but it can be to study the discrete models that may produce a given emergent phenomenon. Finally, the motivation may be to propose models that perform best in the realization of a specific function.
In the first case, we want to check if the simulated model correctly grasps the behavior of the real system. The validation of the model will thus be to have a behavior identical to (as close to as possible) experimental knowledge. The search problem can be seen as the search of the parameter set that minimizes the distance between real and simulated data.
Having a similar behavior can also mean that specific emergent phenomena known to occur in a real system can be observed in the simulation. Emerging ant lines for example, will only occur if the chemical trails leaved by the ants behind them (see next paragraph) have specific properties, as we will see in next section. The emergence of this phenomenon will thus be associated to specific parameter values, and the search problem will consist in searching the different ranges of parameters where an emergent phenomenon is observable. In some cases, choosing slightly different values may lead to completely different results during the simulation, which complicates a manual exploration of the parameter space and justifies the development of automatic techniques.

2.3 Example

We will present the parameter setting of an agent-based model with the example of ant foraging (search for food), in which ants leave chemical trails behind them when coming back to the nest

	Model 1	Model 2	Model 3
Diffusion rate	40	50	60
Evaporation rate	15	15	20

Model 1 Model 2 Model 3

Figure 1: Models with slightly different parameters.

with food (we use the multi-agent programmable modeling environment `NetLogo` [1] and its "Ants" model).
In this model, two parameters condition the formation of chemical trails. The first one is the diffusion rate of the chemical, which corresponds to the fact that a given proportion of the chemical will be diffused to the neighboring patches (regions of the environment) at the next time step. This is used to simulate the diffusion of the chemical in the atmosphere. The second parameter is the evaporation rate of the chemical, which corresponds to the fact that a given proportion of the chemical will disappear from the patch at the next time step. This is used to simulate the evaporation of the chemical in the atmosphere.
For example, we can be interested more precisely in the dynamics of ant lines. Table 1 shows three models with small modifications for the two parameters. We can obtain different dynamics: the difference lies in the way that food sources are exploited. In model 1, food sources are exploited in turn while in model 3, they are all exploited at the same time. As a result, we observe one, two or three ants lines.

2.4 Previous work

Different methods have already been proposed to explore automatically the parameter space of discrete models. In the `NetLogo` platform for instance, the "BehaviorSpace" [1] tool allows to explore automatically and systematically the parameter space. This space is a Cartesian product of values that each parameter can take. However when we have lots of parameters (real-valued parameters for example), the parameter space becomes huge and the systematic exploration becomes impossible.
Other methods have been proposed, which differentially explores the whole parameter space, focusing on the most interesting areas. That's the case of the method developed by Brueckner and Parunak [2]. They use a "parameter sweep infrastructure", which is similar to the "BehaviorSpace" tool of `NetLogo` . However, to avoid a systematic exploration, they use searcher agents and introduce the fitness notion. The aim of a searcher agent is to travel in the parameter space to look for the highest fitness. Starting from a given location in the parameter space, searcher agents have two choices: move or simulate. Each agent chooses according to the confidence of the fitness estimate (proportional to the number of simulations at this point) and the value of the fitness. If it chooses to move, it heads for the neighboring region with highest fitness. A disadvantage of this method is that searcher agents may head for local fitness maxima.

3 Use of genetic algorithms

As the tuning of the parameters of a model is a strongly combinatorial problem, we propose to use genetic algorithms, which generally provide good results on problems of this kind.

3.1 Choice of the fitness function

If we consider the exploration of the parameter space as an optimization problem, we need to define very carefully the function that will have to be maximized by the algorithm. This fitness function is of fundamental importance since the models that will be selected are the one that perform best with respect to this function. In the context of agent-based simulation, the choice of the fitness function is problematic for several reasons: as a first thing, it is not the result of a computation but the dynamics of a process that has to be assessed; secondly, emergent phenomena may be difficult to characterize quantitatively since they are often related to a subjective interpretation by a human observer.

3.1.1 Quantitative vs. qualitative.

Validating an agent-based model by assessing the distance between the simulation and the real system can be done either quantitatively or qualitatively.
In the quantitative case, data are measured in the simulation and compared to data measured in similar conditions in the real system. The distance between the simulation and the real system is then the Euclidean distance between the two data vectors. If we try to select models that are optimized for the realization of some function, the fitness function can also be directly measured by the performance of the system for that function.
In the qualitative case, what is important is that a given emergent phenomenon be present in the simulation: for example, the ants line. The difficulty is then to translate this observation into a quantitative measure (the fitness function). In some cases, the characterization of such emergent phenomena may not be so simple since it may be the result of a subjective interpretation by an observer, which cannot be captured easily by a quantitative measure.

3.1.2 A dynamic process.

In classical optimization problem, the fitness function corresponds to the result of a computation. Therefore, the question of the time at which the measure should be made doesn't make sense: the measure is done when the computation has ended. On the contrary, agent-based simulations are dynamic processes that evolve along time and generally never end.
We can clearly see in the example given in the previous section that the evaluation of the fitness function generally has to be done at a given time-step of the simulation. The choice of this time-step is not neutral and may greatly influence the performance of the genetic algorithm and the resulting model.
Figure 2 shows the foraging simulation at five different time-steps. We can see that the ant lines are not present during all the simulation. This example shows the difficulties to choose the time-steps for the evaluation.

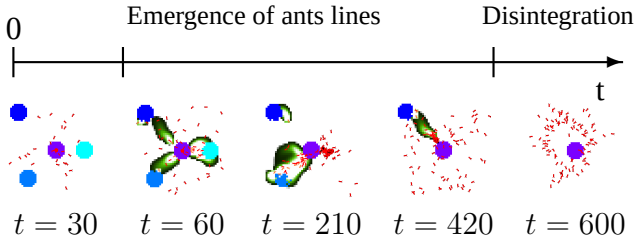


Figure 2: Ant foraging at different time-steps

3.2 Computation of the fitness function

3.2.1 Time.

Since no mathematical model can anticipate the dynamics of an agent-based model without executing it, the computation of the fitness function requires one or even several simulations. This means that the time required to compute the fitness function will be significant. We must therefore find methods to reduce either the number of chromosomes, the time to converge towards an optimum or the time to compute the fitness function. We mainly studied the last possibility through distributed computation and fitness approximation.
Since the different models are independent from each other, the evaluation of their fitness is also independent. Therefore each evaluation of the fitness (that is to say each agent-based simulation) can be done on a different computer. We can thus have several computers to simulate the models and use the master-slave parallel genetic algorithms [3], which improves the performance as compared to standard GA.
Fitness approximation comes to approximate the result of the simulation by a mathematical model, such as a neural network or a polynomial for instance. We tried this approach by training a neural network with test data. After the learning phase, we used it to compute the fitness, with the generation-based control approach, in which the whole population of η generations is evaluated with the real fitness function in every λ generations [4]. The results however were not so good and this approach has been temporarily abandoned. We suspect in that case that the approximation was not good enough to obtain satisfying results but this has to be explored in more details.

3.2.2 Stochasticity.

Two agent-based simulations can generally bring slightly different results even if the underlying model is exactly the same due to the stochasticity of the model and of the simulator. One simulation is not enough to evaluate the fitness function: it can only be considered as an estimate for the fitness.
In such noisy environments, a first solution is to increase the size of the population [5]. To multiply the number of the simulated models reduces the effect of the stochasticity. A second solution is to simulate each model several times to improve the evaluation of the fitness function. Both solutions greatly increase the number of simulations, thus the time, of the genetic algorithm.
Another solution is to use the same technique as with fitness approximation. A solution to the stochasticity problem is then to estimate the fitness of each model with one simulation, and each n generations of the GA (n to choose according to the stochasticity of the model and the desired quality of the estimation), to estimate the fitness of each model with x simulations.
We use the elitism genetic algorithm [6] that is to say we keep the best chromosomes during the algorithm, which allows to continuously improve the solution. Our implemented genetic algorithm replace only 25 % of the population at each generation. Every 3 generations, we estimate the fitness of the models with more simulations.

4 Discussion & Conclusion

We applied the method to some simple examples: the ant foraging with different fitness functions (both quantitative and qualitative). We do not show the results because of the lack of space. As we could already see with the study of the stochasticity, we obtained very different results depending on the choice of the fitness function. The models are strongly optimized for a specific fitness function and may not perform so well with another one. The optimization creates a loss of the flexibility of the dynamics of the agent-based model. A possible solution would be to use several

different initial conditions to evaluate the fitness function. This would however increase again the time necessary to run the algorithm.
The optimization by the genetic algorithm also depends on the constraints imposed to the agents in the model. If a model has lots of constraints (fewer resources for example), it is necessary that it optimizes its global functioning. On the contrary, if the resources are abundant, the pressure on the model to adapt and optimize its functioning will be weaker. As a result, the use of our approach will be mostly beneficial when constraints on the model are high.
The next step is to apply the method to a more complex example. We began a work for the simulation of the glycolysis and the phosphotranferase systems in Escherichia coli. In this work, we are interested in testing the hypothesis of hyperstructures [7]. The hyperstructures are dynamic molecular complexes, enzyme complexes in the case of this work. These complexes allow to improve the behavior of a cell : more flexibility, quicker adaptation. In our study, we have 25 kinds of molecules (or agents). There are altogether about 2200 agents in the simulation. We want to study the potential interest of hyperstructures for the cell. To do this we make the rates of enzymes association and dissociation variable. In this context, the simulation of a model lasts about 10 minutes, which imposes to use the methods described in this article like the distributed computation. To explore this complex example, we will need to develop additional strategies to reduce the parameter space (e.g. by introducing coupling between parameters), to accelerate the evaluation of the fitness function (e.g. by developing approximation methods), and to accelerate the convergence of the algorithm (e.g. by using interactive evolutionary computation). Finally, another important perspective is to explore the effect of varying dynamically the simulation conditions so as to produce more versatile models.

References

[1] Seth Tisue and Uri Wilensky. Netlogo: Design and Implementation of a Multi-Agent Modeling Environment. *Proceedings of Agent 2004*, October 2004.

[2] S. Brueckner and H. Van Dyke Parunak. Resource-Aware Exploration of the Emergent Dynamics of Simulated Systems. *AAMAS 2003*, pages 781–788, 2003.

[3] Erik Cantú-Paz and David E. Goldberg. Efficient parallel genetic algorithms: theory and practice. *Computer Methods in Applied Mechanics and Engineering*, 186, 2000.

[4] Yaochu Jin, Markus Olhofer, and Bernhard Sendhoff. A Framework for Evolutionary Optimization with Approximate Fitness Functions. *IEEE Transactions on Evolutionary Computation*, 6(5):481–494, 2002.

[5] David E Goldberg, Kalyanmoy Deb, and James H. Clark. Genetic algorithms, noise, and the sizing of populations. *Complex System*, 6, 1992.

[6] Tuvik Beker and Lilach Hadany. Noise and elitism in evolutionary computation. In *Soft Computing Systems - Design, Management and Applications*, pages 193–203, 2002.

[7] Patrick Amar, Gilles Bernot, and Victor Norris. Modelling and Simulation of Large Assemblies of Proteins. *Proceedings of the Dieppe spring school on Modelling and simulation of biological processes in the context of genomics*, pages 36–42, May 2002.

A birth-death branching process and its applications in population biology.

G Caron-Lormier[1], JP Masson[2], N Ménard[1] & JS Pierre[1]

[1]UMR CNRS 6552 "Ethologie éVolution Ecologie", Université de Rennes 1, 263 av. Gal Leclerc, F-35042 RENNES Cedex, France
[2]Agrocampus Rennes, 65 rue de Saint-Brieuc - CS 84215 - 35042 Rennes Cedex - France

Abstract

Many mammal species live in groups such as marmots, macaques, bats or wolves. These groups are often socially structured, and in the case of Barbary macaques (*Macaca sylvanus*) the social structure is matriarchal lineages, that is a structure of related individuals. A lineage is constituted of a live female (the matriarch) - whose all ancestors are dead - and all its living female offspring. In this case, males disperse. The dynamic of this structure - simple at first sight - increase the complexity of the groups and populations dynamics, and therefore complicate the understanding of a such system. For instance, in the Barbary macaques species, if the group reaches a critical size, its splits into new groups. The fission does not occur at random but follows the lineages existing in the group, and the newly formed groups are constituted of a combination of these lineages. This dynamic induces strong genetic differentiation among groups, which is a way to speciation. A theoretical approach seemed necessary to understand the evolution of the characteristics of lineages.
We therefore built a mathematical model based on the branching process theory. We use a birth-death branching process constituted of two distinct processes. First, the birth process is a Galton-Watson branching process and generates individuals in the group. We assume that the process begins with one individual at time zero. Second, the death process is a stochastic law, independent from the birth process, and only depends on the generation of the individuals. This process kills individuals and fragments the group in different lineages. The time goes in generation; at time n, both processes stop. Thus, at time n, the group includes individuals with parental link between them and a live-dead status for all of them.
This model allowed us to link demographic parameters such as natality and mortality with characteristics of lineages. Indeed, we obtained an analytical expression of the number of live lineages in a group and we checked the influence of life history parameters (natality and mortality) on the outcomes of the system. First, we note that the system is highly variable, then for identical initial conditions the system presents many different types of outcomes (from extinction to persistence). Second, we show that the number of lineages depends on both the natality and mortality sub-processes. Both an increase of mortality or natality in the process increases the number of lineages. More particularly, the mortality of the oldest individuals in the group presents stronger influence on the number of lineages.
This theoretical approach of group dynamics of a such system can help to understand the influence of life history traits of mammals on the characteristics of social structures. Then, we will be able to investigate the influence of these structures on genetic populations. This model could be adapted easily to many species that live in socially structured groups such as marmots, bats and many primates species - even humans. We could use this model to test biological hypothesis or to predict behavior of the system. Therefore, this approach - including further development - could be used in a conservation biology purpose.

Hybrid Model from fibroblast crawling guided by fibronectin remodeling

J W Liu, O Collin, P Tracqui, E Planus & A Stéphanou

Laboratoire de techniques de l'imagerie, de la modélisation et de la cognition CNRS 5525, Université Joseph Fourier , 38706 La Tronche, France

Abstract

The mechanical properties of the extracellular microenvironnement influence the cell behaviour. It is thus essential to identify their role, in particular in the assembly of the extracellular matrix components (MEC) like fibronectin (FN), for better understanding the mechanisms which control the processes of cellular migration and adhesion. We used mouse fibroblasts (3T3-YFP-FN), transfected in a stable way for the expression of the FN fused with a yellow fluorescent protein (YFP). Experimental datas were obtained by tracking with video microscopy the cell crawling along FN fibers during wound healing experiments. We then develop an hybrid model to investigate the morphogenesis of the FN network synthetised by the cells and the way this network influence the guided movement and migration of cells. In this model, cell migration is described by a isotrope cellular diffusion and an anisotrope haptotactic migration on an oriented FN network, this network beeing synthetised by the cell during their migration and degrade by proteases. The migratory behavior of each cell is depicted on a discrete 2D matrix, using finite elements method. Results show that during wound healing, early migratory cells that go in the wound, organize a new FN network. This network morphogenesis is critical in the process, making highways for later cells to move, thus increasing the speed of obturation of the wound.

Modeling systems biology for *in vrtuo* experiments

G Desmeulles[1] & L Misery[2] & V Rodin[1]

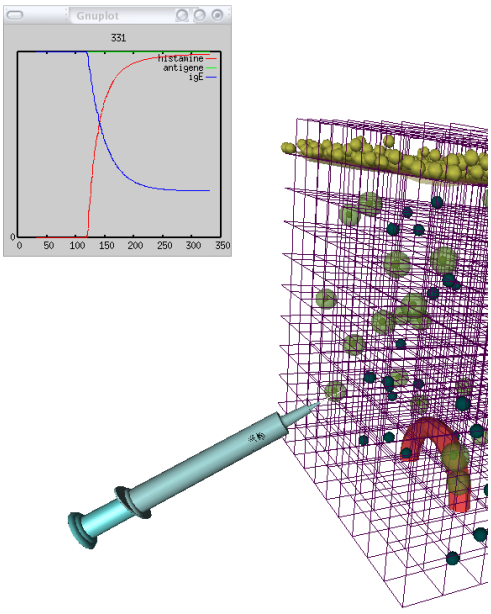
[1]Centre Européen de Réalité Virtuelle, Brest, France.
[2]Service de Dermatologie et Laboratoire de Neurobiologie Cutanée, Brest, France.

Abstract

Nowadays, the Virtual Reality (VR) makes feasible the simulation and the (*in virtuo*) experimentation of complex phenomena, to complete the *in vivo* or the *in vitro* investigations. We apply this alternative method to the study of a complex pathology: the allergic urticaria. We have created a model of virtual skin.

The VR methods differ from classical analytical methods for the simulation in biology. They are based on the autonomy of the entities that populate the virtual universes and are focused on the user/simulation interaction. This aspect allows the coexistence of data having very different natures. It also permits to different disciplines (biology, computer science, dermatology, mathematics...) to collaborate. Thus, we use multiagent systems, the AReVi plateform and algorithms based on chaotic and asynchronous iterations. We propose a generic model for systems biology based on reification of the interactions, on a concept of organisation and on a multimodelling approach. By "reification" we understand that interactions are taken as autonomous agents

In this framework, we have built a skin cube of 600μm length that we can experiment. Chemical reactions and diffusion have been inserted in this model. The obtained results are preliminary to the constitution of a wider biological model and allow envisioning *in virtuo* experiments on skin allergy and its treatment.



Résumé

L'usage de la réalité virtuelle (RV) permet, aujourd'hui la simulation et l'expérimentation (*in virtuo*) de phénomènes complexes, pour compléter la recherche *in vivo* ou *in vitro*. Nous appliquons cette méthode alternative à l'étude d'une pathologie complexe: l'urticaire allergique. Pour cela nous avons créé un modèle de peau virtuelle.

Les méthodes de RV diffèrent des méthodes analytiques classiques de simulation en biologie. Elles sont basées sur l'autonomie des entités qui peuplent les univers virtuels et mettent l'accent sur l'interaction entre utilisateur et simulation. Cet aspect permet la cohabitation de données ayant des natures très différentes et permet la collaboration entre plusieurs métiers (biologistes, informaticiens, dermatologues, mathématiciens,...). Ainsi, nous utilisons les systèmes multi agents, la plateforme ARéVi et des algorithmes basés sur des itérations chaotiques et asynchrones. Nous proposons un modèle pour l'étude des systèmes complexes, basée sur la réification des interactions, sur le concept d'organisation et sur une approche multi-modèle. Le terme "reification" signifie que les interactions sont les objets actifs des simulation.

Dans ce cadre, nous avons construit un cube de peau de 600μm de coté que nous pouvons expérimenter. Ce modèle met en oeuvre des mécanismes de de reaction chimique et de diffusion. Les résultats obtenus sont préliminaires à la constitution d'un modèle biologique plus étoffé pour envisager des expériences *in virtuo* sur l'allergie cutanée et son traitement.

RNA Secondary structure prediction

S.Engelen[1] & F.Tahi[1]

[1]LaMI UMR 8042 Université d'Evry - CNRS, Genopole, 91000 Evry, France

1 Introduction

The concept of secondary structure was introduced by Doty and Fresco [1]. The secondary structure is composed of all the Watson-Crick pairings, AU, GC, and the Wobble pairing GU. The knowledge of secondary structure is essential to understand the relations between structure and function of the RNA. There are two main approaches to predict the secondary structure of RNA: the thermodynamic approach and the comparative approach. The thermodynamic approach is based on the idea that the structure of minimal energy is the real structure. This approach consists to calculate the structure with the lowest energy using a set of thermodynamic parameters experimentaly defined [2]. An algorithm has been proposed by Zuker [3] in $O(n^3)$ time complexity and $O(n^2)$ space complexity. A disadvantage of the thermodynamic approach is that the real structure is not necessarly the lowest free energy structure. The comparative approach is used when several number of aligned homologous RNA sequences which are assumed to have the same secondary structure but different sequences are available. The idea is to find pairs which covary to maintain Watson-Crick and Wobble complementarity [4]. At the beginning, the comparative approach was used manually to predict the structure of large RNA sequences [4]. The first algorithm was designed by Han and Kim [5] and has an $O(m \times n^2 + n^3)$ time complexity, where m is the number of homologous sequences involved and n the size of the largest sequence. The comparative approach gives better results than the thermodynamic approach. In the first part of the paper, we propose an original implementation of the comparative approach, *P-DCFold*, which yields, in the worse case, an $O(log_4 n * n^2)$ time complexity where n is the sequence length. The main problem of the comparative approach is that there are many possible combinations of homologous sequences to use for the prediction, and only a few gives correct structure predictions. So, in the second part, we expound an algorithm, called $SSCA$, to select the best homologous sequences to use with the comparative approach. In the last part, we apply the two algorithms on two known RNA structure.

2 P-DCFold description

2.1 Principle of DCFold

DCFold is an algorithm for RNA secondary structure prediction based on the comparative approach [6]. The goal of the algorithm is to predict the secondary structure of one sequence, called the "*target sequence*", using informations from the other aligned homologous sequences, called the "*test sequences*". The search for helices is done in two steps: first, we search for helices in the *target sequence*, then we check their conservation in the *test sequences*.

Two palindromes (p, p') and (q, q') are compatible when they are disjoint (*...p...p'...q...q'...*) or embedded (*...p...q...q'...p'...*). Interleaved palindromes (*...p...q...p'...q'...*) form pseudoknots and are not compatible. A palindrome is said structural if it defines a helix of the secondary structure. It is said conserved if it appears in each sequence at the same aligned position. A palindrome is not necessarily conserved with the same pairs of bases. Some mutations may occur. They are called compensatory mutations when the pairing still remains possible. As a conserved palindrome is not always structural, one of the central problems is to determine a heuristic criterion to select the ones that are structural. Two criteria are combined: the length of the palindromes and the number of compensatory mutations. The palindromes are selected only if their length is greater than $log_4 n$, where n is the target sequence length, and when they present at least one compensatory mutation per site. We also attribute scores to palindromes according to thermodynamic parameters. The

first parameter is on pairing stabilities. The second parameter is function of pairings which close helices. The palindrome are searched using the "divide and conquer" approach (Figure 1).

Algorithm $DCfold$

Begin

- Search for palindromes that satisfy the length criterion in the target sequence S.
- Select among them the conserved ones satisfying the mutation number criterion.
- Select among them the compatible palindromes.
- Iterate the process on sub-sequences of S (deduced from the subdivision of S by the compatible palindromes).

End

Figure 1: $DCfold$ Algorithm

As the length criterion is function of the sequences and sub-sequences length, the subdivision allows to search palindromes from the most significant to the less significant. The "divide and conquer" approach allows to reduce the space of research for the less significant palindromes, by forcing the structure with the most significant palindromes. The obtained palindromes constitute a set of structural palindromes, defining the secondary structure. The subdivision of the sequence is possible because only compatible palindromes (so no pseudoknots) are searched in $DCFold$.

The complexity in time of $DCFold$ in the worse case is equal to $O(log_4 n * n^2)$, when n is the sequence length. Indeed, the search for palindromes in a sequence can be done without effort in $O(n^2)$ and the number of iterations is less than $log_4 n$.

2.2 Principle of the pseudoknot searching

The search for pseudoknots is done after finding the secondary structure without pseudoknots. Given a sequence S, $DCFold$ finds a list $L1$ of all compatible palindromes that satisfy our selection criteria. Re-launch $DCFold$ on S without palindromes of L (S') allows to find another list $L2$ of all compatible palindromes which are not compatible with palindromes of $L1$. Therefore, a palindrome of $L2$ will form a pseudoknot with a palindrome of $L1$. Now, relaunch again $DCFold$ on S' without palindromes of $L2$ allows to find a third list $L3$ of compatible palindromes which are not compatible with palindromes of the list $L1$ and with palindromes of the list $L2$. And so on, until no palindromes are found. Therefore, the principle of the algorithm is to search for pseudoknots in several steps, each helix of the pseudoknot being selected in a different step.

3 Selection of the homologous sequences

To predict the secondary structure of a given RNA sequence, the comparative approach needs well aligned homologous sequences. There are three main problems with alignments of sequences under structure constraints. Firstly, these sequences are often quite variable in regions which are located at the periphery of the structure. So in these regions, the comparison of sequences is difficult or even impossible. Secondly, in opposite, the heart of the structure often has catalytic activity and is less variable. So, in this region, no information on compensatory mutations can be used. Thirdly, because of structure constraints, the mutations are selected more rapidly by evolution in helical regions. So these helical regions are more variable than the remainder of the structure (single strand regions). The result is that in these regions, alignments are often of low quality and helices are often shifted. To answer the first two points, one must select "adequate" homologous sequences that are variable enough to have compensatory mutations but close enough to be compared. This can be made using criteria on homology between the homologous sequences

and the target sequence. The third point is crucial because to be efficient, the comparative approach needs to use homologous sequences with correct helices alignment without shifted helices. So, one must find criteria to differenciate homologous sequences with shifted helices from homologous sequences that are correctly aligned, precisely in helices, with the target sequence.

3.1 Algorithm for homologous sequences selection

We have build an algorithm, called $SSCA$, to select the interesting sequences which have adequate variability and correct helices alignment. In order to evaluate these parameters, for each test sequence, our algorithm calculates a substitution matrix: it contains all the substitution rates between this test sequence and the target sequence for the four bases A,C,G and U. To modelize an ideal test sequence having "adequate variability" and "correct helices alignment", we also build a model consisting on constraints on the substitution matrices of the sequences. A score of interest is then calculated for sequences according to the model and their substitution matrices. The sequences closest to the ideal test sequence are the most interesting for the structure prediction (Figure 2).

Algorithm $SSCA(Aln)$

Begin

- Select a target sequence S_t in the alignment Aln
- Build a model M according to contraints
- For each homologous sequences S_i of the alignment Aln:
  - Calculate the substitution matrix M_i between S_t and S_i
  - Calculate a score for S_i according to constraints of the model M and the matrix M_i
- Rank the sequences S_i according to their score
- Select the adequate number of sequences to predict the structure of S_t

End

Figure 2: Algorithm for homologous sequences selection

In fact the constraints of the model M are a whole of formulas. They are divided into two groups. The first group concerns the selection of homologous sequences according to their variability and the second the selection of homologous sequences according to their helices alignments.

3.2 Constraints on variability of homologous sequences

The constraints concerning the homologous sequences selection with the "adequate variability" are applied on squares of the diagonal and on deletion ("-") and undeterminate ("N") columns (Figure 3). These constraints favorize homologous sequences which have homology closer to 75 percent, 2 percent of deletion and none undeterminate bases (0 percent). The model is as follows:

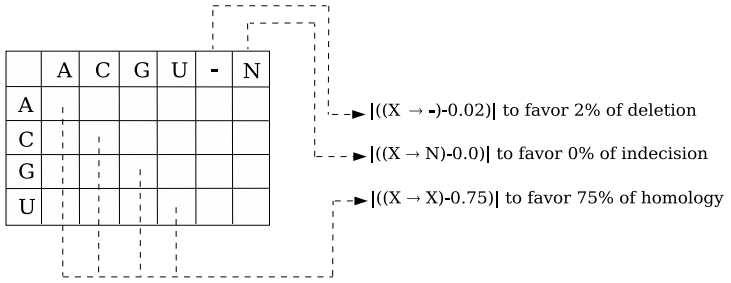


Figure 3: Homology formulas of the model M for selection of four homologous sequences

3.3 Constraints to modelize homologous sequences with adequate helices alignment

We expound two possible influences in helices which will allow us to select the homologous sequences according to their helices alignments. These influences are on the stability of base pairs in helices and on intermediate states in double mutations.

3.3.1 Stability of base pairs

Due to the number of hydrogen bonds, the GC base pair is more stable than the AU base pair which is more stable than the GU one. Because of these differences in stability, when a helix is important to maintain the global structure, GC base pairs are favorized while GU base pairs are disfavorized. So, we must observe a deviation in substitution matrices of helices due to this high stability of GC pairs. Consider all possible mutations beetween base pairs without GU base pairs which are less represented. We can eliminate AU↔UA and CG↔GC mutations because these mutations do not change the stability of helices. For the four other mutations, if GC pairs are prefered in helices, the equilibrium between base pairs will be deviated to GC base pairs (Figure 4). The result will be a deviation of the nucleotide substitution rates too. As we see in Figure 4, under this hypothesis, we observe an increase of substitutions A→C and A→G compared to A→U, and an increase of U→C and U→G compared to U→A and a discrease of C→A and C→U compared to C→G, and a discrease of G→A and G→U compared to G→C. So, the constraints to underline the influence of the GC stability are applied on squares which show a deviation (Figure 4).

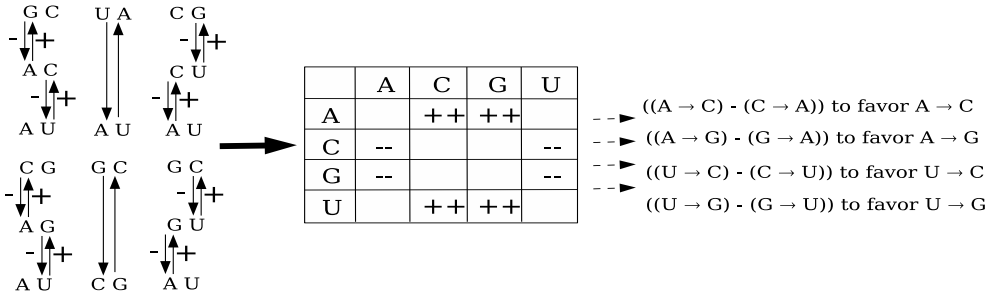


Figure 4: Deviation of base pair substitutions in helices

3.3.2 Stability of intermediate states

Due to the low rate of mutations, the double mutations between the different base pairs can not appear simultaneously, so double mutations use an intermediate state (for example: AU→UU→UA). In relation to the stability of this intermediate state, the double mutations are favorized or disfavorized. This intermediate state may be a very deleterious unpair state or a GU pairing state which is slightly deleterious. As we can see in Figure 5, the double mutations which use this state are AU↔GC and UA↔CG. So, in helices, an increase in the mutation rate of AU↔GC and UA↔CG must be observed.

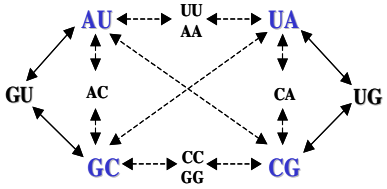


Figure 5: Base pair substitutions in helices

Thus, differences between the stability of GU intermediate state and the stability of the other states deviate the base pair mutation rate matrix to an increase of AU↔GC and UA↔CG in helical

regions. As these base pair mutations imply substitutions A↔G and C↔U, the result must be an increase in nucleotide substitutions A↔G and C↔U in relation to the other mutations. So, the substitution matrix will be deviated (Figure 6).

	A	C	G	U	
A		-	+	-	- - - ► ((A → G) - (A → C)) to favor A → G
C	-		-	+	- - - ► ((C → U) - (C → A)) to favor C → U
G	+	-		-	- - - ► ((G → A) - (G → U)) to favor G → A
U	-	+	-		- - - ► ((U → C) - (U → G)) to favor U → C

Figure 6: Deviation matrix due to the GU intermediate state

So, the constraints to underline the influence of the GU intermediate state are applied on gray squares of the Figure 6. Indeed, for these squares the influence of the GC stability is in the same direction. For example we compare A→G with A→C because for these two squares the influence of the GC stability is positive.

4 Results

We have used $SSCA$ algorithm on structure of Escherichia coli RNAseP and Escherichia coli tmRNA using P-DCfold. For these RNA, P-DCfold must use four homologous sequences interesting for the comparative approach. To measure the effectiveness of P-DCfold and $SSCA$, we do structure predictions with P-DCfold for each possible combination of four homologous sequences among all the homologous sequences of the alignments and among the ten best sequences selected by $SSCA$. As we know the secondary structure of the RNA, we attribute a score to each prediction and we plotte the distribution of the prediction scores. The more the curves (with the ten best homologous sequences) are deviated towards high scores compared with the general curve (with all homologous sequences), the more the predictions are good. The general curve and the curve for the homologous sequences selected with $SSCA$ are in Figure 7.

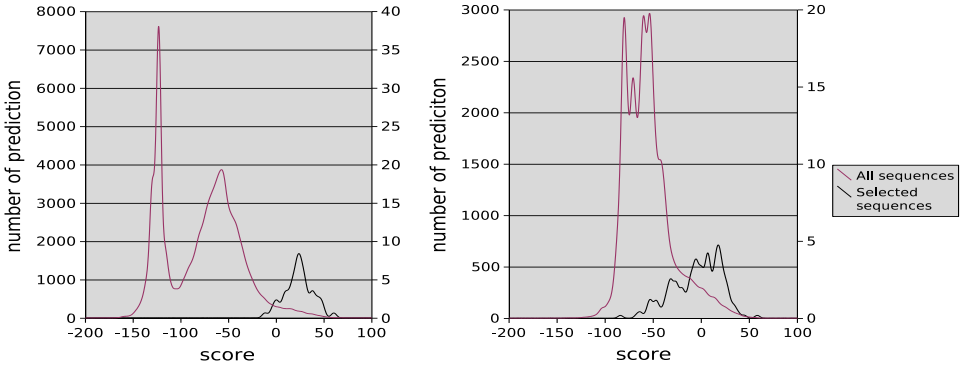


Figure 7: Score distribution of tmRNA and RNAseP prediction. Left: RNAseP. Right: tmrna

In this graphic, we see that the curves for the score predictions using the ten best homologous sequences is deviated to the higher scores. This means that our algorithm allows to select combinations of homologous sequences which give correct predictions. The results are better for the RNAseP than for the tmRNA because the average is more deviated to the higher score, from -80 to 16. This means that almost 50 percent of the selected combinations give acceptable predictions (above a score of ten). While without criteria of selection we have two chances per thousand to obtain a good prediction (score above 30), with $SSCA$ we have now two chances per ten.

In the Figure 8 we show the prediction of the Escherichia coli tmRNA which give the best score. In this prediction there are three ommitted helices (in bold) and none false positives.

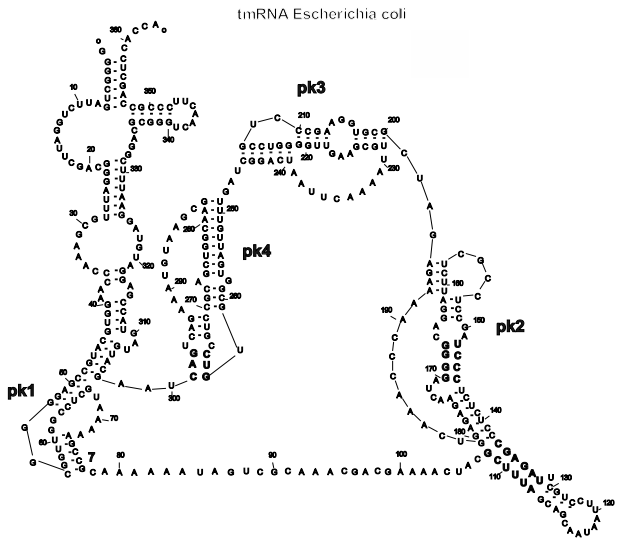


Figure 8: Best prediction of the Escherichia coli tmRNA

5 Conclusion

We have presented here results on two RNA but we have tested other RNA like ribosomal RNA (16S RNA and 23S RNA). P-DCfold presents several strong points. The first one is the palindrome selection criteria used. Indeed, we have set length and mutation criteria that proved their highly efficiency. Therefore, our algorithm insures a good prediction of the secondary structure including any kind of pseudoknots. Another strong point is the fact that very few sequences are necessary for our algorithm. Finally, the high performance of our algorithm is its ability to correctly predict a secondary structure in record time. Indeed, the search for pseudoknots do not increase the algorithm complexity, since it is just a re-launching of $DCFold$. $SSCA$ allows to improve all the results of P-DCfold selecting the best homologous sequences for the comparative approach. Indeed, without $SSCA$ a user of P-DCfold had two chances per thousand to obtain a good prediction, while with $SSCA$ he have now two chances per ten. The use of $SSCA$ does not increase the complexity time since it is in *O(m×n)* with n the length of the target sequence and m the number of homologous sequences in the alignment.

References

[1] P. Doty, H. Boedtker, J.R. Fresco, R. Haselkorn, and M.Litt. Secondary structure in ribonucleic acids. *Proc. Natl. Acad. Sci.*, 1959.

[2] D.H. Matthews and al. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, 1999.

[3] D.H. Mathews, D.H. Turner, and M. Zuker. RNA secondary structure prediction. *Current Protocols in Nucleic Acid Chemistry*, 2000.

[4] C.R. Woese and N.R. Pace. Probing RNA structure, function, and history by comparative analysis. *The RNA Worl Eds.: R.F. Gesteland and J.F. Atkins (Cold Spring Harbor Laboratory Press) NY*, 1993.

[5] K. Han and H.J. Kim. Prediction of common folding structures of homologous RNAs. *Nucl. Acid Res.*, 1993.

[6] F. Tahi, M. Gouy, and M. Regnier. Automatic RNA secondary structure prediction with a comparative approach. *Computers and Chemistry*, 2002.

Immersive Exploration by Visualization of Factual and Textual Genomic Data

Nicolas FEREY, Pierre-Emmanuel GROS, Joan HERISSON and Rachid GHERBI

Laboratoire d'Informatique et de Mécanique pour les Sciences de l'Ingénieur (CNRS, Université Paris-Sud XI, BP 133, 91403 Orsay CEDEX, France
{ferey, gros, herisson, gherbi}@limsi.fr

Abstract

Biologists are leading current research on genome characterization (sequencing, alignment, transcription). However, genomic information shows some characteristics that make it very difficult to exploit. These data are heterogeneous, huge in quantity, geographically distributed, recorded within public or private databanks, and constitute an important factual data source (GenBank, SwissProt and Decrypthon), but genome knowledge is not limited to DNA or annotated protein sequences. Indeed, there is a significant quantity of information relating to these genes recorded in an unstructured format within many publications (Medline). This paper presents Genome3DExplorer, a new modeling and software solution to explore textual and numerical genomic data based on an adapted federator description language. Genome3DExplorer offers biologists a user-friendly visualization of data, within a virtual reality environment, and using a well-adapted graphical paradigm. This solution allows biologists to explore huge sets of genomic data, and it could be applied to other fields. This kind of graph-based exploration has the advantage of displaying global topological characteristics, which are not easily visible using traditional exploration tools. Finally, some results produced by Genome3DExplorer software from various sets of biological data will be presented.

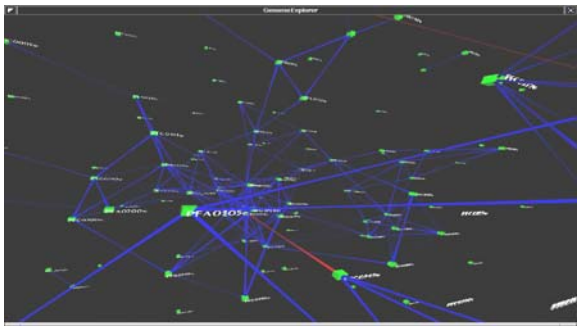


Figure 1: Plasmodium correlation expression profiles network

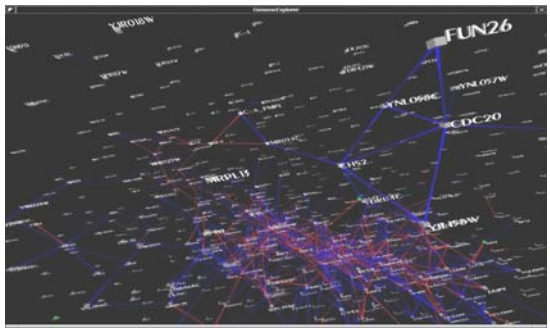


Figure 2: Yeast correlation expression profiles network

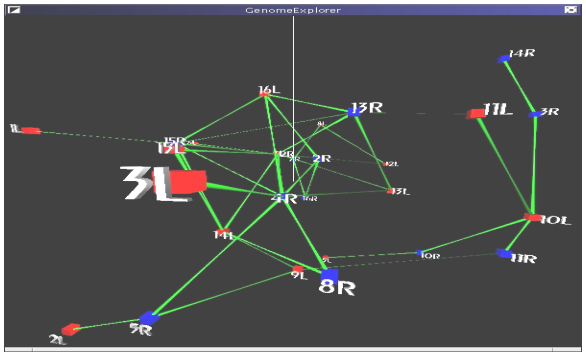


Figure 3: Visualization of gene duplications of yeast chromosome

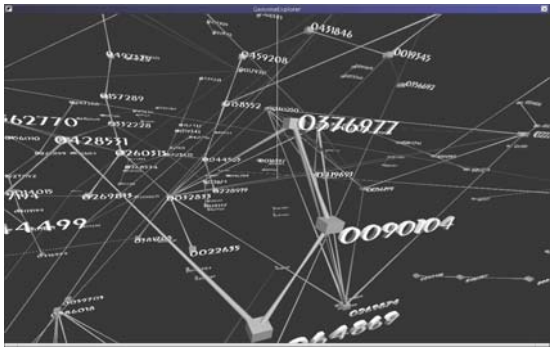


Figure 4: Exploration of a protein sequence alignments network (*Decryphton* data)

Period distribution of regulatory networks

Y Grondin[1] & D J Raine[1]

[1]Department of Physics and Astronomy, University of Leicester, Leicester, LE1 7RH, UK

Abstract

We have designed a simple regulatory model presenting the basic features of a biological regulatory system. In our model, the nodes correspond to genes that interact on each other by a set of positive or negative directed links. A node responds to regulatory signals to become active or inactive in accordance with the net number of activators minus inhibitors acting. If there are no signals, the node becomes inactive.

The state of a small random subset of nodes is specified and fixed, representing, for example, external inputs to the network. The proportion of negative links, μ, turns out to be a major factor in determining the behaviour of the network. Here, we focus especially on the range of μ for which the dynamics of the network is periodic. For a given network, the variation in the random subset of inputs, leads to a set of attractors with a range of periods. We show that the distribution of periods of the resulting cycles follows a power-law. This feature is independent of the architecture of the networks. At any one time, a subset of nodes does not take part in the regulation and the activity of the network.

Several parameters of the model affect the characteristics of the distribution of periods, including the size of the network, N, the proportion of negative links, μ, and the mean connectivity. We show that the period of the longest cycle varies approximately as $\propto \ln N$ implying, in this model, that the number of attractor states grows slowly with the size of the genome.

GenoMEDIA, a Multimedia Distributed Genomic databases

Gros Pierre-Emmanuel, Gherbi Rachid

[1] Bioinformatics team, Human-Computer Communication Department LIMSI-CNRS, University of Paris-Sud, BP 133, 91403 0rsay cedex, France, genoteam@limsi.fr

Abstract

The central dogma of biology states that DNA is expressed and transcribed into RNA, which is translated into an amino acid sequence, which then folds into its 3D structure. High-throughput experimental methods produce masses of data, so that the whole of biology has changed from a data-light science into a data-driven science. The field of bio-informatics relates to the collection, organization and analysis of this large amount of biological data using networks of computers and databases (usually with reference to the genome project and DNA sequence information).
Many databases and software tools are currently available for accessing these genomic data. In general, each tool uses a different type of data in exchange protocols, and usually they offer specific services. Our system approach, named GenoMEDIA [2], addresses the problem of integrating these heterogeneous databases and software utilities within a generic distributed platform. The main GenoMEDIA interest is its possibility to offer users to focus more on their aims rather than loosing time and energy with different and barbaric technical considerations. This could help users to perform a lot of studies and analysis on large-scale genomic sequences in a reasonable human time. In order to achieve his aims, GenoMEDIA provides two specific middleware, named Lydia and Antje, for this integration. Lydia offers facilities for working simultaneously with a variety of Services and Databases; Antje is designed to allow users to manage multiple heterogeneous remote databases in a uniform vision. The aim of this paper is to present GenoMEDIA and how heterogeneous databases and remote services are integrated, in particular how Antje was designed, implemented and tested with various genomic databases. Lydia is used to call a service through a RDBMS and Antje allows users to manage multiple remote databases through a single interface. The GenoMEDIA architecture allows a user to send an augmented unified query through standard SQL that involves remote heterogeneous databases and services. Unlike GeneWeaver, GenoMEDIA uses this SQL approach for managing existing DBMS technologies (PHP, ODBC,..). Like in Oracle's distributed database architecture, Antje's goal is to facilitate access to databases on multiple computers by making them appear as if they were a single file available through a user-friendly interface. Antje uses the notion of remote view instead of the notion of database link. Several applications are now available through GenoMEDIA; most of them are focused on 3D modeling and processing of genomes [1]. The platform integrates also a genomic database that informs and responds many users' and applications requests, and on which various operations could be performed.

References:

1. P.-E. Gros, J. Hérisson, N. Férey, and R. Gherbi, "Combining Applications and Databases Integration Approaches in a Common Distributed Genomic Platform". The First IEEE International Workshop on Information Networking and Applications (INA'2005), 28-30 march 2005, Tamkang University, Taiwan.
2. P.-E. Gros, N. Férey, J. Hérisson, and R. Gherbi, "A Distributed Multimedia Database Visualization within an Immersive Environment for Bioinformatics". International Conference on Bioinformatics and its Applications, ICBA2004, 17-19 december 2004, Lauderdale, FL, USA

Chemical Reaction Network Generation Using a Strategy Language in Term Rewriting

Liliana Ibanescu

LORIA-Université Henri Poincaré Nancy 1 F-54506 Vandoeuvre-lès-Nancy Cedex, France

Abstract

In the context of an industrial application, we have explored rule-based programming in the field of automated generation of chemical reaction mechanisms. In this application, called GasEl, the representation of chemical species uses the notion of molecular graphs, encoded by GasEl terms inspired from linear SMILES notation. A solution of chemical species is encoded by an Associative Commutative operator.

Chemical reactions are expressed by rewriting rules on molecular graphs, encoded by a set of conditional rewriting rules on GasEl terms. The primary mechanism is a specific chemical reaction network and the strategies language of the ELAN system is used to express reactions chaining in primary mechanism generator. Our approach is illustrated by coding in ELAN ten generic reactions of oxidizing pyrolysis.

ELAN has some good properties for the generation of kinetic mechanisms: chemical reactions are naturally expressed using conditional rules; ELAN matching power allows for retrieving patterns in chemical species, thanks to the capability of handling multiset structures through the use of associative and commutative functions; ELAN provides a strategy language and strategy constructors to define control on rules, which appears as essential for designing generation mechanisms in a flexible way; thanks to its efficient compiler, ELAN can handle a large number of rules application (several thousands per second) and is thus well-suited to the computational complexity of the application.

Computational methods for *in virtuo* experiments of biochemical reactions on the microscopic scale
Application to blood coagulation

S. KERDÉLO, P. REDOU, J.F. ABGRALL, J. TISSEAU

Centre Européen de Réalité Virtuelle (CERV), 25 rue Claude Chappe, F-29280 Plouzané

Abstract

Like *in vitro* and *in vivo* experiments, *in silico* experiments are nowadays commonly run for the investigation of a complex biological system. Our work tends to demonstrate that *Virtual Reality* is also a useful tool for the investigation of such systems; using the main tools and concepts of this discipline, *in virtuo* experiments are then available.

The investigation of a complex biological system always involved the study of biochemical reactions; these are commonly represented by their kinetics that can be described at 3 different scales:

- the macroscopic scale: each species is represented by its concentration and the system evolution is given by a set of *differential equations*,
- the mesoscopic scale: each species is represented by its number of molecule and the system evolution is given by a *master equation*,
- the microscopic scale: each species is represented individually (position, orientation) and the system evolution is given by the *brownian motion laws* and by *productive molecular collisions*.

In this study, we focus on the microscopic scale and we propose computational methods, based on multiagent systems, that allowed *in virtuo* experiments of a complex biological system composed of biochemical reactions.

We then apply our computational methods to the particular case of blood coagulation, and especially on prothrombin activation. This activation is a key process during blood clotting, and it is due to an enzymatic complex assembled on a phospholipid membrane. It is well-known that the phospholipid membrane enhances prothrombin activation but the underlying mecanism is an area of discrepancy. We have thus used our approach in order to simulate this reaction, and as a result we propose a mecanism by which the phospholipid membrane increases the rate of prothrombin activation.

Integration of high-throughput functional genomics approaches in Arabidopsis

C Lurin, C Andres, A Avon, N Dautrevaux, A Falcon de Longevialle, L Heurtevin, A Marmagne, C Schmitz-Linneweber, I Small

Unité de Recherche en Génomique Végétale (INRA/CNRS/UEVE), 2, rue Gaston Crémieux, CP 5708, 91057 Evry cedex

The initial analyses of the Arabidopsis genome sequence described less than 25000 genes. Experimental evidence of gene structure (from full-length cDNAs) and expression was available for only a small proportion of these putative genes and experimental evidence of function available for even fewer. In the 4 years since the completion of the genome sequence, giant strides have been made in accumulating cDNA sequences and expression data from microarrays. Together with improvements in automatic annotation routines and better use of expert manual annotation, these data have enabled us to have a much better view of the coding potential of the Arabidopsis genome. The current estimates of the number of protein-coding genes in Arabidopsis are now around 28000. However, the improvements in structural annotation have not been matched by equivalent improvements in functional annotation. Attribution of experimentally proven functions to Arabidopsis genes/proteins remains a long and painstaking task. The principal reason for this lag is that the genomics approaches that have so accelerated the mapping, sequencing and expression studies by permitting parallel analyses on hundreds or thousands of genes at once have been largely inapplicable to functional studies. This situation is starting to change and new “massively parallel” clone-based proteomics and reverse genetics approaches promise to bring the same benefits to functional annotation. Our group is pioneering the systematic use of RNAi [1] to ascertain gene functions and the use of high-throughput ORFeome-based protein function assays including protein localisation, protein-protein and protein nucleic-acid interactions [2]. A major challenge that remains to be solved is the integration and visualisation of these data obtained by varied approaches.

[1] Hilson, P., Allemeersch, J., Altmann, T., Aubourg, S., Avon, A., Beynon, J., Bhalerao, R.P., Bitton, F., Caboche, M., Cannoot, B., Chardakov, V., CognetHolliger, C., Colot, V., Crowe, M., Darimont, C., Durinck, S., Eickhoff, H., Falcon de Longevialle, A., Farmer, E.E., Grant, M., Kuiper, M.T.R., Lehrach, H., Léon, C., Leyva, A., Lundeberg, J., Lurin, C., Moreau, Y., Nietfeld, W., PazAres, J., Reymond, P., Rouzé, P., Sandberg, G., Segura, M.D., Serizet, C., Tabrett, A., Taconnat, L., Thareau, V., Van Hummelen, P., Vercruysse, S., Vuylsteke, M., Weingartner, M., Weisbeek, P.J., Wirta, V., Wittink, F.R.A., Zabeau, M. and Small, I. (2004) Versatile genespecific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications. Genome Res, 14, 2176-2189

[2] Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M-L., Mireau, H., Peeters, N., Renou, J-P., Szurek, B., Taconnat, L., Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell, 16, 2089-2103

Games networks applied to Plasminogen Activation system

M. Manceny[1], C. Chettaoui[1], M. Malo[2], G. Barlovatz-Meimon[2] & F. Delaplace[1]

[1]Laboratoire de Méthodes Informatiques (LaMI), UMR 8042 CNRS-Université d'Evry, FRANCE
[2]LaMI, Equipe DYNAMIC, Université Paris XII, UMR 8042 CNRS-Université d'Evry, FRANCE

1 Introduction

Cell behaviors is the result of complex interplays between molecules, transduction pathways and biochemical "systems". Understanding the complexity of these interactions is an important but challenging problem. This put the emphasis on the necessary modeling to describe, explain and predict the dynamics of such systems.

To model complex interplays, we propose in section 2 a game-theory-based framework named *theory of games networks*. Section 3 presents an application of games networks theory on a biological functional network: *the Plasminogen Activation system*. We conclude in section 4.

2 Game Theory and Theory of Games Networks

The theory of games networks is an extension of game theory. Section 2.1 presents some results about game theory, and section 2.2 presents the extension.

2.1 Game Theory

Game theory allows us to model a set of interacting agents. Each agent chooses the strategy that maximizes its payoff. The payoff depends on the strategies played by other agents. The reader can refer to the books [4, 6, 7, 8] for a complete overview of game theory and its applications.

2.1.1 Strategic games

Strategic game is a model of interplays where each agent chooses its plan of action (or strategy) once and for all, and these choices are made simultaneously. Moreover, each agent is rational and perfectly informed of the payoff function of other agents. Thus, they aim at maximizing their payoffs while knowing the expectation of other agents.

Definition 1 (Normal or Strategic Representation)
A strategic game Γ is a $3-$uple $\langle A, C, u\rangle$ where:

- *A is a set of players or agents;*
- *$C = \{C_i\}_{i\in A}$ is a set of strategy sets where each C_i is a set of strategies available for the agent i, $C_i = \{c_i^1, \cdots, c_i^{m_i}\}$*
- *$u = (u_i), i \in A$ is a vector of functions where each $u_i : C \mapsto \mathbb{R}, i \in A$ is the payoff function of the agent i.*

2.1.2 Nash Equilibrium

Nash Equilibrium is a central concept of Game Theory ([7]). This notion captures the steady states of the play for a strategic game in which each agent holds the rational expectation about the other players behavior. A *Pure Nash Equilibrium* (**pne**) corresponds to a *strategic profile* c (or vector) where c_i is the strategy "chosen" by the player i.

Definition 2 (Pure Nash Equilibrium of a Strategic Game)
Let $\langle A, C, u\rangle$ be a strategic game, a Pure Nash Equilibrium *is a profile of strategies $c^* \in C$ with the property that:*

$$\forall i \in A, \forall c_i \in C_i, u_i(c^*_{-i}, c_i) \leq u_i(c^*_{-i}, c^*_i)$$

In other words, *no agent can unilaterally deviate of a* **pne** *without decreasing its payoff.*

2.2 Theory of Games Networks

In game theory, every agents are interacting together. In theory of games networks, a modular description of the dynamics is possible. It enables us to describe local interactions between agents. Thus, games networks allow situations where an agent can be involved in several different games at the same time and with different agents. The theory of games networks is based on strategic games.

Games networks make the representation of *modular interactions* possible, each one is supported by a subset of agents. The agents involved in local interactions are participating to the same game, i.e. the same module. The payoffs of the agents define the interaction rules. An agent is shared between several modules, but its strategies remains the same whatever the game. The reader may refer to [3] to have a complete overview of theory of games network.

2.2.1 Definition

The definition of a games network mainly consists of defining a set of agents connected to a set of games. The normal representation of a games network is as follows :

Definition 3 (Games Network)
A games network is a $3-$uple $\langle \mathcal{A}, C, \mathcal{U}\rangle$ where

- *$\mathcal{A}$ is a set of agents or players.*
- *$C = \{C_i\}_{i\in\mathcal{A}}$ is a set of sets of strategies.*
- *$\mathcal{U} = \{\langle A, u\rangle\}$ is a set of* game nodes *where each $A \subseteq \mathcal{A}$ is a set of agents and $u : A \times C_A \mapsto \mathbb{R}$ is a set of payoff functions such that $u = \{u_i : C_A \mapsto \mathbb{R}\}_{i\in A}$.*

A games network offers a synthetic representation to define the different interplays between several players. The structure $\langle A, u\rangle$ totally determines a game played by a subset of agents since it useless to include the strategies which are the same for any agent of the network. With this representation, game nodes can be viewed as dynamic modules whose agents are locally interacting.

A games network is represented by a bipartite graph $\langle \mathcal{A}, \mathcal{U}, E\rangle, E \subseteq \mathcal{A} \times \mathcal{U}$ where an edge $(i, \langle A, u\rangle)$ is a member of E if and only if $i \in A$. Graphically, agents are represented by circles, and game nodes by rectangles (See figure 1 for an illustration).

2.2.2 Global Equilibria

Considering biological systems, the steady states are considered as the characteristic observable states of the system dynamics. In game theory, the computation of Nash equilibria gives us the steady states. A similar notion has been developed in games networks: *Pure Games Network Equilibria* (**PGne**). To define **PGne**, we equip the theory with the *restriction operator* which allow us to focus on an arbitrary sub game.

Restriction

Definition 4 (strategy Profile Restriction)
Let $\mathcal{A} = [1 : n]$ be a discrete interval representing a set of agents, let $C = \{C_i\}_{i\in\mathcal{A}}$ be a set of strategy sets. Given a strategy profile $c \in C$, we define its restriction to a subset $A \subseteq \mathcal{A}$, denoted by $c{\downarrow_A}$, as follows:

$$\forall i \in \mathcal{A},\ (c{\downarrow_A})_i = \begin{cases} c_i & \text{if } i \in A \\ \bot & \text{otherwise} \end{cases}$$

We extend the restriction operator by removing bottom elements ($\bot$) of the profile, but the order of the other values is conserved in the resulting profile. We note the composition of the removals and restriction operation as follows: $c \Downarrow_X$

Example 1 *Let $\mathcal{A} = [1 : 4]$ and $c = (c_1, c_2, c_3, c_4)$. Let $A = \{1, 3\}$, we have $c{\downarrow_A} = (c_1, \bot, c_3, \bot)$ and $c \Downarrow_A = (c_1, c_3)$.*

Pure Games Network Equilibrium corresponds to a compatible association of local equilibria. We assume that agents follow the *single played strategy* rule, i.e. an agent plays the same strategy for every connected games.

Definition 5 (Pure Games Network Equilibrium)
Let $\Gamma = \langle \mathcal{A}, C, \mathcal{U} \rangle$ be a games network, let $c^* = (c_1, \cdots, c_n)$ be a strategy profile of every agent. c^* is a pure games network equilibrium of a subset $U \subseteq \mathcal{U}$ (noted $c^* \in \mathbf{PGne}_\Gamma(U)$) iff:

$$\forall \langle A, u\rangle \in U, c^* \Downarrow_A \text{ is a pure Nash equilibrium of the game} \langle A, (C_i)_{i\in A}, u\rangle$$

3 Application to the Plasminogen Activation system (*PAs*)

The theory of games networks has been used to model biological systems. In such systems, agents correspond to genes or proteins for example. In particular, we have modeled here a part of the Plasminogen Activation system (*PAs*), which is involved, among others, in the migration of cancer cells. *PAs* is composed of 3 molecules (*uPAR*, *uPA* and *PAI-1*). Understanding the regulation of the *uPAR/uPA/PAI-1* complex formation appears to be central to analyze the migration.

3.1 Cell migration and the PA system (PAs)

Cellular migration is a complex process which can be described like a succession of stages: *adhesion*, *contraction*, *de-adhesion* [5]. We are interested in the *PAs* system which participates to the establishment of a molecular bridge between the cell and the extra-cellular matrix. This bridge leads to the migration of the cell [1] . *PAs* system is composed of a receptor *uPAR* (Receptor of urokinase), a protease *uPA* (urokinase Plasminogen Activator), and a specific inhibitor *PAI-1* (Plasminogen Activator Inhibitor-1)[9]. The sequence of interactions implied in the promigratory process is as follows: *PAI-1* can blind to *VN* (the Vitronectin, a protein of the extra cellular matrix) which stabilizes *PAI-1* in its activation form. Once *PAI-1* is activated, it clings to a complex formed

by *uPAR* and *uPA*. The complex is internalized by a receptor $\alpha 2$ *M-LRP* (Low-density lipoprotein receptor-related protein) in the cell. Then *uPAR* is recycled at the front of the cell. The signaling molecule *PAI-1* induces modifications of cell morphology including changes in cytoskeleton of actin, necessary to the migration. These modifications imply the regulation of the activation of GTP-ases of the *Rho* family [10].

3.2 Games network modeling

Figure 1 shows the games network modeling of PAs. Circles represent the different agents involved in PAs (c_i agents are intermediary complexes), rectangles are games which define the rules of interactions between agents.

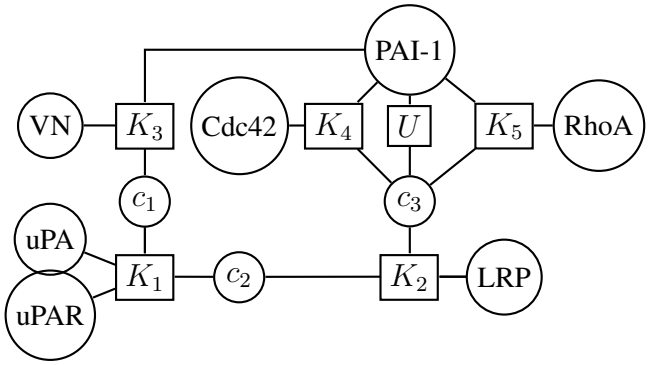


Figure 1: Games network modeling of PAs

3.3 Results

Once we have modeled the structure of the PAs network, we have to define the payoffs of each game. Theory of games networks allows us to determine that the K_i games (which represent complexation) are not central in the determination of **PGne**. The main game is U, whose payoffs are shown in figure 2. It is to be noted that for biological reasons, not explained here, the complex c_2 has 2 strategies, and *PAI-1* has 3 strategies ([2]). In the table, the first number represents the payoff of c_2, and the second number the payoff of *PAI-1*.

$c_3/pai-1$	0	1	2
0	**0,0**	$1,0$	$2,0$
1	$-,-$	**2,0**	$3,-1$

Figure 2: Payoffs for game U from figure 1

With the payoffs defined, we are able to compute **PGne**, i.e. is steady states of the *PAs*. Some tools of theory of games networks, not described here, enable us to affirm that **PGne** are exactly composed of Nash equilibria from game U. These equilibria are written in bold in figure 2.

According to experiments, both Nash equilibria correspond to characteristic biological states. The first one $(0,0)$ corresponds to a non migratory state and the last one $(2,0)$ corresponds to a promigratory state.

4 Conclusion

In this paper we have presented a theoretical framework based on game theory which is suitable to help explain complex biological interplays. Theory of games networks has been used to model a part of the Plasminogen Activation system. The modeling confirms the existence of two characteristic states which correspond to physiological configurations (i.e. non and pro migratory states).

The definition of a games network allows the combination of several games into a single network. This puts the emphasis on the way that the network structure is determined, because different structures can be proposed to model the same situation. In order to compare them, it is necessary to identify the equivalence between games networks. We based the equivalence on the equilibria. Informally, two games are equivalent if their equilibria are the same.

We have also defined the notion of reorganization of a games network based on the idea of *dependence* (an agent is dependent on another if its choices of strategies are altered by the strategies of the other). This reorganization enables us to compute a normal form which emphasizes the *basic "building" modules* of the network. According to preliminary experiments, these modules appear to be closed to the notion of "*biological modules*". Indeed a reorganization of the normal form removes steady states of the system. Normal form is the smallest network which has identical dynamics with that of the initial biological system.

References

[1] B. Chazaud, R. Ricoux, C. Christov, A. Plonquet, RK. Gherardi, and G. Barlovatz. Promigratory Effect of Plasminogen Activator Inhibitor-1 on Invasive Breast Cancer Cell Populations. *American Journal of Pathology*, 160(1):237–246, January 2002.

[2] F. Delaplace, C. Chettaoui, M. Manceny, and M. Malo. Games Networks & Application to PAs System. In *IPCAT (Submited)*, 2005.

[3] F. Delaplace and M. Manceny. Games network. Technical Report 101-2004, Laboratoire de Méthodes Informatiques (LaMI), CNRS-UMR 8042, Université d'Évry-Val d'Essonne, 2004.

[4] R. Gibbons. *Game Theory for Applied Economists*. Princeton University Press, 1992.

[5] DA. Lauffenburger and AF Horwitz. Cell migration : a physically integratedmolecular process. *Cell*, 84:359–369, 1996.

[6] R. B. Myerson. *Game Theory : Analysis of Conflict*. Harvard University Press, 1991.

[7] J. F. Nash. *Essays on Game Theory*. Edward Elgar Publishing Limited, 1996.

[8] M. J. Osborne. *An introduction to game theory*. Oxford University Press, 2003.

[9] KM. Providence and PJ. Higgins. PAI-1 Expression Is Required For Epithelial Cell Migration In Two Distinct Phases Of In Vitro Wound Repair. *Journal of cellular physiology*, pages 200:297–308, 2004.

[10] AJ. Ridley. Rho GTPases and cell migration. *Journal of cell science*, 114:2713–2722, August 2001.

Human genome-wide protein interacrions

P Mazière & C A Ouzounis

EMBL-EBI
Computational Genomics Group
Welcome Trust Genome Campus
Hinxton, Cambridge, CB10 1SD, UK

Abstract

The determination of protein interactions is essential for the understanding of molecular function and is the basis for the contstruction of cellular networks. With the development of high-throughput methods that allow the acquisition of numerous experimental data, the dataset of identified protein-protein interactions is growing quickly. Whereas model organisms such as yeasts are well suited for this kind of methods, experimental data related to human protein interactions remain relatively poor. This is yet a key factor to improve understanding of diseases and pharmaceutical developments.

In order to fill this gap, several predictive computational methods have been implemented. They rely on different characteristics that can be used to infer human protein interactions from other kind of information: protein interactions in other species are used to predict human protein interactions on the basis of orthologs [1]; co-expressed proteins are grouped in functional clusters; gene fusion events may reflect the possibility of interactions between two proteins that were fused on one transcript in organisms located upward the evolution tree; phylogenetic profile analysis attempt to gather proteins that seems to be linked along the evolutionnary process to perform specific biological processes [2].

All these methods have high rates of false positives and false negatives that should be decreased by crossing their results [3]. By representing these different protein-protein interaction networks as graphs, we have the possibility to analyse the differences existing between the results provided by these methods. Confronting this information with sequence similarity data, structural domain composition and functional annotation should allow us to validate or to eliminate interactions. Another interesting possibility would be to differentiate the various kinds of interactions that are expressed by the term 'interaction', such as stable interactions (complexes), transient interactions (i.e, enzymatic reaction) or involvement in the same pathways.

References

[1] Lehner B. and Fraser A.G., *A first-draft human protein-interaction map*, Genome Biology 2004, 5:R63.
[2] Eisenberg D., Marcotte E.M. and Yeates T.O., *Protein function in the post-genomic era* , Nature 2000, 405(6788);823-6.
[3] von Mering C., Krause R., Snel B., Cornell M., Oliver S.G., Fields S. and Bork P., *Comparative assessment of large-scale data sets of protein-protein interactions*, Nature 2002, 417(6887);399-403.

Validation of an Agent Based System Using Petri Nets

Thomas Moncion[1] & Guillaume Hutzler[1] & Patrick Amar[2]

[1]Laboratoire de Methodes Informatiques, Equipe SyDRA, Tour Evry 2, 523 Place des terrasses de l'agora, 91000 Evry Cedex.
[2]Laboratoire de Recherche en Informatique, Bât 490 Université Paris-Sud, 91405 Orsay Cedex France.

Abstract

Multi-agents systems are powerful tools used to model complex systems. The validation of such tools is usually carried out by running simulations of a model and by comparing with data from the real system. Since this process can be very long, it would be interesting to be able to validate a model, at least partially, without having to run simulations. To achieve this, our approach consists in transposing the model in an abstract framework where spatial and temporal constraints no longer apply and check in this abstract framework that some of the well-known properties of the system are preserved by the model.
We developed our approach in the context of the Hsim platform, which is an agent-based simulator used to model the interactions of macromolecules in a 3D virtual cell surrounded by a membrane. This simulator is not dedicated to a specific model of macromolecular interactions but allows the modelling of any kind of interactions, described using a modelling language representing reactions between the various molecules, including the formation of complexes.
The aim of this work was to explore systematically the molecules and complexes potentially produced and consumed by a given model. To this end, we build a Petri net (PN) in which places represent the chemical species that are created and transitions represent the corresponding reactions. Whenever a reaction may apply to one or several molecules, producing new molecules, the reaction and molecules are added to the PN, along with the corresponding edges. Starting with the initial molecular species present in the system, the process ends with a complete network of all the potential reactions and molecules respectively triggered and produced by the simulator.
On the one hand the construction phase of the Petri net allows us to determine all the possible macromolecular assemblies, and on the other hand, the analysis of some properties of these Petri nets enables to exhibit properties of the model without needing to run a simulation (e.g. detection of invariants, loops. . .)

1 Multi-agent model

To describe the movement and the association and dissociation phenomena within a cell, a simulation program has been developed in C++ and OpenGL[1]. This program simulates a virtual cell as a three dimensional space surrounded by a membrane.
The simulator is designed to be independent of a particular model. A language has been developed to describe the interactions rules which can exist between the various agents involved in the simulation.
This language describes four possible types of interaction between two molecules S and T.

- Association: the molecule S binds to the molecule T to form the complex S-T.
  Example: `{C}A + {~D}B -> A(1)* B(1) [0.6]`
- Reaction: the molecule S reacts with the molecule T to produce two new molecules S' and T'.
  Example: `A + B -> C + D [0.5]`

- Dissociation: a complex S-T can dissociate and release the molecules S and T
  Example: {˜C}A * B -> A + B [0.01]

- Catalysis: a complex S-T can be transformed into another complex S'-T'
  Example: A * B -> C * B [0.9]

The numbers between brackets in the right part of the association rule mean that molecule A (respectively B) cannot bind to more than one molecule B (respectively A). It is also possible to restrict the scope of a reaction to already bound (or unbound) molecules. We can take again the reaction of the association example and trigger the reaction only if the molecule of type A is already bound to a molecule of type C, and if the molecule of type B is not bound to a molecule of type D. Each rule is assigned an execution probability which corresponds, in the long-range, to the kinetic of the reaction.

2 Petri net construction and properties

2.1 Petri net

Let $G(P, T, A, M_0)$ a Petri Net (PN) where:

- $P = \{p_1^{S_1}, \ldots, p_i^{S_i}, \ldots, p_t^{S_t}\}$ is a set of places where each place is labelled by an assembly S_i In our case, these places correspond to the resources of the system. A place can thus contain several tokens.

- $T = \{t_1^r, \ldots t_j^r, \ldots t_u^r\}$ is a finite set of transitions where each transition is labelled by a reaction $r \in R$, R being the set of all the reactions.

- $A = \{a_1, a_2, a_3, \ldots, a_n\}$ is a finite set of directed edges where $A = \{A^- = P \times T \mapsto \mathbb{N}\} \cup \{A^+ = T \times P \mapsto \mathbb{N}\}$ corresponding to the set of directed edges from places to transitions or from transitions to places.

- $M_0 = \{m(p_1), m(p_2), \ldots, m(p_l)\}$ is an initial marking. This marking specifies the number of tokens in each place for the initial state of the system.

It is possible to build a matrix representation of a Petri net. We consider two applications Pre and $Post$ such as:

- $Pre = P \times T \mapsto \mathbb{N}$ is the pre-incidence application, where $Pre(p_i, t_j)$ is the weight of edge (p_i, t_j). $Pre(p_i, t_j) > 0$ if the edge exists, $Pre(p_i, t_j) = 0$ otherwise.

- $Post = T \times P \mapsto \mathbb{N}$ is the post-incidence application, where $Post(p_i, t_j)$ is the weight of edge (t_j, p_i). $Post(p_i, t_j) > 0$ if the edge exists , $Post(p_i, t_j) = 0$ otherwise.

We then have a matrix C which we call the incidence matrix, which is defined as follows:

$$C = Post - Pre$$

2.2 Petri net construction algorithm

Since there is a language describing the biochemical reactions between the molecules, we can compute the complete set of all the assemblies which can possibly be made. For this purpose, we have designed an algorithm to build a such *Petri net*.
To understand how this algorithm works we will explain it using a simple example: a cascade of two enzymes, $e1$ which catalyses the transformation of the initial substract $s1$ to $s2$, and $e2$ which catalyses the substract $s2$ giving the final product $p3$. We will suppose that the whole reaction is *channelised* by the self assemblied complex $e1-e2$. Here are the rules used to model this example:

```
r1: s1 + e1 -> s1(1) * e1(1) [0.6];
r2: {s1}e1 + e2 -> e1(1) * e2(1) [0.9];
r3: {~s1}e1 * e2 -> e1 + e2 [0.0001];
r4: {e1}s1 + {e1}e2 -> s1(1) * e2(1) [1.0];
r5: {e2}e1 * {e2}s1 -> e1 + s2 [1.0];
r6: s2 * e2 -> p3 + e2 [0.9];
```

The places p_1, p_2 and p_3 which represent the initially provided single molecules $e1$, $s1$ and $e2$ are inserted into the vector P. The initial Petri net is shown in figure 1 while the final vector P showing the mapping between the places and the chemical species is in figure 2 (at this point of the explanation we consider only the three first entries). The algorithm will build step by step both the vector P (the places) and the set A (the transitions), giving the Petri net.
The main process is done by sweeping the vector P from the first to the last entry. Starting from the initial set of places, the vector is built by appending new entries at the end. So let's consider the first entry p_1 labelled by $e1$; The reaction r_1 can be applied using the molecule $s1$ labelling the place p_2. This reaction builds a new complex $e1s1$, as this is a new molecular species a new entry p_4, labelled by $e1s1$, is appended at the end of the vector. The new node $t_1^{r_1}$ is put into the set of transitions and the corresponding arcs are added to update the Petri net (cf. figure 1).

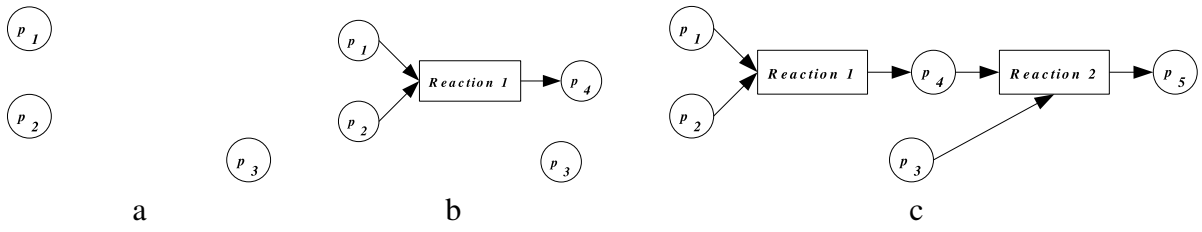
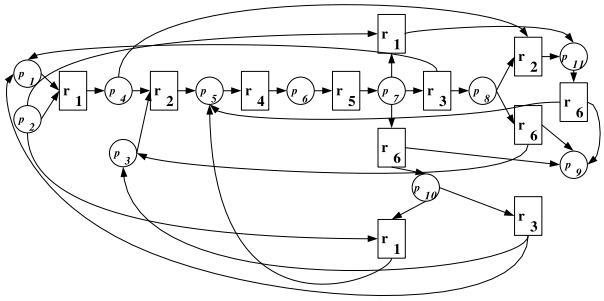


Figure 1: (a)Initial Petri net (b)First transition of the Petri net (c)Second transition of the Petri net

As no other reaction can be applied to the molecular species labelling the place p_1, the algorithm goes to the next entry of the vector P i.e. p_2 labelled by $s1$. The only possible reaction with $s1$ has already been used, r_1 giving the complex $e1s1$; As both this molecular species labels an entry in the vector, and this entry has been created using this same reaction, nothing more is done.
The process continues with the third entry of the vector, p_3 labelled by $e2$. The reaction r_2 can be applied to $e2$ and $e1e2$ building a new complex $e1s1e2$, this complex labelling a new entry p_5 in the vector. As previously, a new node is put into the set of transitions and the corresponding arcs are added to update the Petri net (cf. figure 1).
The process continues with the next entry, the place p_4 labelled by $e1s1$, which will lead to nothing more as the only possible reaction, r_2 has already been used with this assembly. Then we go to the next entry, the place p_5 labelled by $e1s1e2$. At this point, the only possible reaction is r_4 which applies *inside* the complex $e1s1e2$ leading to a third binding, between $s1$ and $e2$, and thus to a new place p_6 and a new transition $t_3^{r_4}$.
This algorithm continues until the last entry in the vector P has been processed, the Petri net being incrementally built (cf. figure 2).



P_1	P_2	P_3	P_4	P_5	P_6
e1	*s1*	*e2*	*e1* | *s1*	*e1*—*e2* | *s1*	*e1*—*e2* | *s1*
P_7	P_8	P_9	P_{10}	P_{11}	
e1—*e2* | *s2*	*e2* | *s2*	*p3*	*e1*—*e2*	*e1*—*e2* | | *s1* *s2*	

Figure 2: Final Petri net.

2.3 Properties of the Petri net

By simply observing the final Petri net, we can notice that the place p_2, labelled by the molecule $s1$, has only outgoing arcs while the place p_9, labelled by the molecule $p3$, has only incoming arcs. This shows that the molecules of type $s1$ are consumed (and can only be consumed) while molecules of type $p3$ are produced (and only produced) by the whole system.
This simple observation however is not sufficient, which justifies the formal study of the properties of the PN. This study allow us to confirm the validity of a model, to show the principal properties of the reaction network and to observe the dynamics of the system.
Several works [2] [3] show the interest of this type of study within the framework of metabolic networks.

2.3.1 P-invariants

P-invariants are vectors of places, which we note y. The multiplication of the transpose of y with any marking is identical to the multiplication with the initial marking ($y^T.M = y^T.M_0$). Vector y thus describes a conservation relation of markings. By taking into account this conservation relation, it comes:

$$y^T.C = 0$$

These relations imply that, for any two place invariants that we note I_1 and I_2, we have:

$$I_1 + I_2 = const \implies c_1 I_1 + c_2 I_2 = const \quad (c_1, c_2\text{: natural integer})$$

The essential property of P-invariants is that the weighted sum of the tokens associated to the vector is constant whatever the evolution of the PN.
If we come back to the example of the PN shown in section 2.2, it is then possible to show that all the places containing either $e1$ or $e2$ are invariants. Translated into biological vocabulary, this means that whatever the evolution of the system, the enzymes will never be consumed. If we hadn't detected these P-invariants for the enzymes $e1$ and $e2$, this would have been the sign of a problem in the model.

2.3.2 Dynamics of the Petri net

The study of the dynamics of the PN corresponds to the observation of the evolution of the tokens in this PN. This allows to determine the molecules that are produced and consumed and to observe the consequences in the evolution of the system when some resources are suddenly missing.

Let's take the example of the PN of section 2.2 with the initial marking $M_0 = \{m(p_1) = 1, m(p_2) = 4, m(p_3) = 1\}$. The study of the evolution of the PN shows us that at the end (i.e. when no more transitions are enabled), all the tokens initially present in p_2 can be found in p_9, with one token left in each of places p_1 and p_3.
By considering the simple case where $M_0 = \{m(p_1) = 1, m(p_2) = 4, m(p_3) = 0\}$, the evolution of the system will stop at the level of p_4. Indeed, the lack of the enzyme $e2$ prevents the production of p_3.

3 Discussion

The *Hsim* simulator is able to modelling and simulate any kind of macromolecular assemblies described as the result of elementary biochemical reactions between molecules.
We demonstrated in this work that, even if the main interest of agent-based modelling lies in the simulation of emerging spatio-temporal structures, it still may be useful to develop validation tools that allow to explore some properties of models without even running the corresponding simulations.It is necessary to check the validity of a model defined by the modeller. We proposed a method based on the construction of a Petri net of macromolecular assemblies and reactions. This enabled to verify that well known properties of a given system may be verified (conservation of enzymes for example).
In the example shown in the article, the biological properties that we exhibited were simple and could have been noticed by the simple observation of the reaction network. As the number of reactions increases however, the number of possible structures can become very important. It then rapidly becomes impossible to study visually the properties of the model, which justifies the automated techniques that we proposed.
In addition, our method has proven to be useful for the validation of models but it may also be useful to characterise the dynamics of agent-based models and automatically identify emergent phenomena such as spatial and temporal structures. The first step would be to automatically identify, in the Petri net, such interesting properties as P-invariants, loops, consumed or produced assemblies etc. In our example, we defined the P-invariant that we wanted to verify (corresponding to the knowledge that enzymes are not transformed by a reaction). It would be best if such P-invariants could be found automatically by the validation tool, which would enable the modeller to see directly all the invariants in his or her model. The second step would then be to identify, in the reaction network, subnetworks that may be considered as higher-level structures in the simulation. To be able to take into account such higher-level structures, it is necessary to extend the modelling language so as to allow the definition of reactions between structures at different levels. It will also be necessary to study the behaviour of these higher-level structures in the dynamic context of the simulation, so as to reintegrate space and time factors in the characterization of their behaviour. By integrating our validation tool in the simulator, the long-term objective is thus to enable the dynamic characterization of emergent structures, which may lead to the dynamic design of multi-level agent-based models and simulations.

References

[1] Patrick Amar, Gilles Bernot, and Victor Norris. Hsim: a simulation programme to study large assemblies of proteins. *Journal of Biological Physics and Chemistry*, 4:79–84, 2004.

[2] K. Voss, M. Heiner, and I. Koch. Steady state analysis of metabolic pathways using petri nets. *In silico Biology*, 2003.

[3] I. Zevedel-Oancea and S. Shuster. Topological analysis of metabolic networks based on petri net theory. *In silico Biology*, 2003.

Comparison of Multiscale Models of RNA Secondary Structures

Aïda Ouangraoua[1]

[1]LaBRI - Université de Bordeaux 1, 351 Cours de la Libération, 33405 Talence cedex, France

Abstract

RNA molecules consist of a succession of nucleotides A, C, G and U representing the *primary structure*. By folding back onto itself, an RNA molecule forms a secondary structure, stabilized by the forces of hydrogen bords between certains pairs of bases, and dense stacking of neighbouring base pairs.

For a few years, based on an analogous approach to the philosophy of sequence homology programs, several studies propose to show how structural similarity is related to functionnal similarty in RNA (see [1]). Thus, the ability to compare Seconday Structures can potentially led to a better understanding of the function of RNA molecules (see [2]).

RNA secondary structures are usually represented by different labeled ordered trees. According to different biological goals, these tree representations can be rough (considering only structural patterns) or refined until an exact coding of the structure is obtained.(see [3, 4]). RNA secondary comparison can then be reduce to tree comparison. However each comparison can use only one type of tree representation.

In order to take into account every tree representation of a given RNA secondary structure during comparison algorithms, we propose to integrate them in a uniq model called *multiscale tree graph*. A multiscale tree is a tree defined with additional equivalent relations on vertices and such that all the quotient graphs are also trees. We then present a dynamic programming algorithm for comparing multiscale models of RNA secondary structures. The core of the method relies on an adaptation of an algorithm proposed by Zhang and Shasha for comparing ordered rooted trees (see [5]).

References

[1] W.Fontana, D.Konings, P.Stalder, P.Schuster: Statistics of RNA secondary structures. Biopolymers, 33: 1389-1404. (1993)

[2] I. Hofacker, W.Fontana, P.Stalder, L.Bonhoeffer, M.Tracker, P.Schuster: Fast folding and comparison of RNA secondary structures. Monatshefte fur Chemie, 125: 167-188. (1994)

[3] M. Vauchaussade, X. Viennot: Enumeration of RNA secondary structures by complexity. Lecture Notes in Biomathematics no. 57: 360-365. (1985)

[4] B.A. Shapiro, K. Zhang: Comparing multiple RNA secondary structures using tree comparisons Cabios (6): 309-318. (1990)

[5] K. Zhang, D. Shasha: Simple fast algorithms for the editing distance between trees and related problems: SIAM Journal on Computing, 18(6): 1245-1262. (1989)

A structured approach to store metabolism modeling data

N. Parisey[1], M. Beurton-Aimar[1,2],C. Nazaret [3],JP. Mazat[1]

[1]Laboratoire de physiologie mitochondriale - Université Bordeaux 2 - 146 rue Léo Saignat, 33076 Bordeaux - France.
[2]LaBRI UMR 5800, Université Bordeaux 1 - 350 Avenue de la libération, 33405 Bordeaux - France.
[3]ESTBB, Université Victor Segalen - Université Bordeaux 2 - 146 rue Léo Saignat 33076 Bordeaux - France

Abstract

Metabolic pathways are often described through static descriptions of the different biochemical reactions that compose them (like in KEGG[1] databank) or with sets of equations which describe the activity of the networks (often time dependent like Ordinary Differential Equations (ODE)).
There are many tools to interact and to use this kind of information but integration of the semantic of the data into a whole model is still not very widespread. The major project in this domain : E-Cell[6] has focused on the description of processes more than on the description of the biological objects. SBML[2] (linked to the E-Cell project) or CellML[3] are XML-like languages [2] that describe biological processes. They posses a semantic squeleton but they are not at the moment fully capable of expressing the core of the biological data.
A possible answer to this problem is to structure information with different types. These types will define characteristics (attributes) of each entity and behaviors (functions) attached to them. Such methodology is largely used in software engineering in many domains like economy, manufacturing, computer-aided tasks.

So, we have defined a model following the **object oriented methodology** to build a generic model of metabolic pathways [4]. An Object Oriented Model (**OOM**) was designed to structure data extracted from three recent models of mitochondrial pathways(Cortassa[1], Wright [5] and Yugi[6]). These models design metabolic pathway behaviors using sets of ODE.
The first one (Cortassa) has three main components: *TCA cycle, oxidative phosphorylation and* Ca^{2+} *dynamics*. Each of these three sets of reactions describe a *"module"* in the whole mitochondrial metabolic pathway. They can be considered specific pathways identified by a name and the list of their reactions (linked here to ODE equations).Our OOM can easily store this modules in an entity named metabolic pathway which can be composed into a larger module named biological description. So any parameter that concerns the composition is stored at this level whereas specific module parameters can be stored inside the composition. We have observed that some parameters are dependent of the biological modeling hypothesis. They must be stored in biological description for they express data that are true for all the modules . Some of them are module dependent (for example constant value of a reaction) and must be stored into components.
The second one (Wright) is a *TCA cycle* model. After encoding it through the OOM, we have tested the possibility to integrate this *TCA cycle* data into the previous model. This operation of mixing models is possible through our OOM structure.
The more recent model is a whole mitochondria model (Yugi) composed from several publications, some quite old. This model was converted into an SBML document. By doing so, it was no more possible to use this model has a sub-model within another SBML structure. Moreover, the SBML structure is not fully able to capture the biological hypothesis linked to the biological modeling. With our OOM, all these information are kept.
The structure of our OOM permit to automatically check if model composition is relevant using pieces of other model. These other models biological hypothesis and model constructions (for

[1]http://www.kegg.org

example kinetic parameters taken from the same species) should be compatible as expressed by the application. Our OOM should also permit to annotate the mitochondrial metabolism using information as heterogeneous as kinetics parameters, metabolic pathway involvement, model simulation results, experimental results, link with biological theory about pathologies and so on.

References

[1] Sonia Cortassa, Miguel A Aon, Eduardo Marban, Raimond L Winslow, and Brian O'Rourke. An integrated model of cardiac mitochondrial energy metabolism and calcium dynamics. *Biophys J*, 84(4):2734–2755, Apr 2003. Evaluation Studies.

[2] M Hucka, A Finney, H M Sauro, H Bolouri, J C Doyle, H Kitano, A P Arkin, B J Bornstein, D Bray, A Cornish-Bowden, A A Cuellar, S Dronov, E D Gilles, M Ginkel, V Gor, I I Goryanin, W J Hedley, T C Hodgman, J-H Hofmeyr, P J Hunter, N S Juty, J L Kasberger, A Kremling, U Kummer, N Le Novere, L M Loew, D Lucio, P Mendes, E Minch, E D Mjolsness, Y Nakayama, M R Nelson, P F Nielsen, T Sakurada, J C Schaff, B E Shapiro, T S Shimizu, H D Spence, J Stelling, K Takahashi, M Tomita, J Wagner, and J Wang. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, Mar 2003. Evaluation Studies.

[3] Catherine M Lloyd, Matt D B Halstead, and Poul F Nielsen. CellML: its future, present and past. *Prog Biophys Mol Biol*, 85(2-3):433–450, Jun 2004.

[4] Beurton-Aimar M. S. Pérès N. Parisey J.P. Mazat. How to lock a model from a miscomposition of objects ? In *13th European Conference on Object-Oriented Programming*, june 2003.

[5] B E Wright, M H Butler, and K R Albe. Systems analysis of the tricarboxylic acid cycle in Dictyostelium discoideum. I. The basis for model construction. *J Biol Chem*, 267(5):3101–3105, Feb 1992.

[6] Katsuyuki Yugi and Masaru Tomita. A general computational model of mitochondrial metabolism in a whole organelle scale. *Bioinformatics*, 20(11):1795–1796, Jul 2004.

Classification of elementary flux modes in mitochondrial metabolism

Sabine Pérès[1], Marie Beurton-Aimar[1] & Jean-Pierre Mazat[1]

[1]Physiopathologie mitochondriale, INSERM U688 and Universite Bordeaux 2,
146, rue Leo-Saignat, 33076 Bordeaux-cedex, France.
sabine.peres@etud.u-bordeaux2.fr; Marie.Aimar@phys-mito.u-bordeaux2.fr; jpm@u-bordeaux2.fr

Abstract

We describe the bioenergetic mitochondrial metabolism (Krebs cycle, β-oxidation of fatty-acids, oxidative phosphorylation, etc.) with 45 enzymatic reactions and carriers and 31 metabolites. Using Fasttool, FluxAnalyzer or Metatool, 435 324 elementary flux modes are derived from the stoichiometry matrix of the network. This huge number of elementary flux modes is due to the great number (20) of carriers and exchangers of intermediate metabolites between mitochondrial matrix and cytosol, and to the conservation of "currency metabolite". This decomposition in elementary flux modes demonstrates : the importance of some reactions in the mitochondrial metabolism (pyruvate carboxylase) the importance of carriers favouring different metabolic pathways. However it is difficult to tackle this huge number of elementary flux modes. In order to gain insight into their relevant physiological properties, we investigate several methods of classification (clustering) with the hope to let come out classes of elementary flux modes and to derive their physiological properties, which could in some instances be attached to specific type of mitochondria in different tissues or different organisms. We compare our results with previous results obtained by others on mitochondrial metabolism.

Acknowledgements: The authors would like to thank Steffen Klamt for performing some calculations.

Computer simulation of Multiple Myeloma in the context of Systems Biology

G Querrec[1], R Bataille[2], V Rodin[1], JF Abgrall[1] and J Tisseau[1]
[1]Centre Européen de Réalité Virtuelle, Brest, France, contact : gquerrec@enib.fr
[2]Centre Anti-cancéreux René Gauducheau, Nantes, France

Cancer is complex adaptive system. We assume that multiple myeloma (MM) can be study in the context of complexity using Systems Biology (SB). SB is a new field in biology aims at understanding biology at systems level. We develop the "In Virtuo" experimentation for SB using virtual environment. Multiagent systems provide an attractive computer framework for SB and "In Virtuo" approachs.

We implement a computer simulation supporting the maturation model of myeloma cells based on the CD45 expression. One has shown that CD45 expression decreases with the maturation of the cells and CD45 annihilation is a critical pronostic for patient survival. CD45 expression is also necessary for IL6 proliferation signal but inhibits IGF-1 proliferation signal. Moreover, CD45 expression is stimulate by IL6 and its activity inhibited by dimerization . The kinetic of CD45 dimerization depends of the molecule isoform.

So, we have developed a computer simulation based on a multiagent system which integrates the model of myeloma cell previously described in their micro-environnement.

The simulation shows that kinetic of MM tumor presents an exponential shape (Figure 1) which is characteristic of the first stages of tumor growth. Next, we can observe the lost of tumor connectivity, which was previously shown to occur with tumor evolution. We assume that CD45 isoform is an important feature for the proliferation signal choice (data not shown).

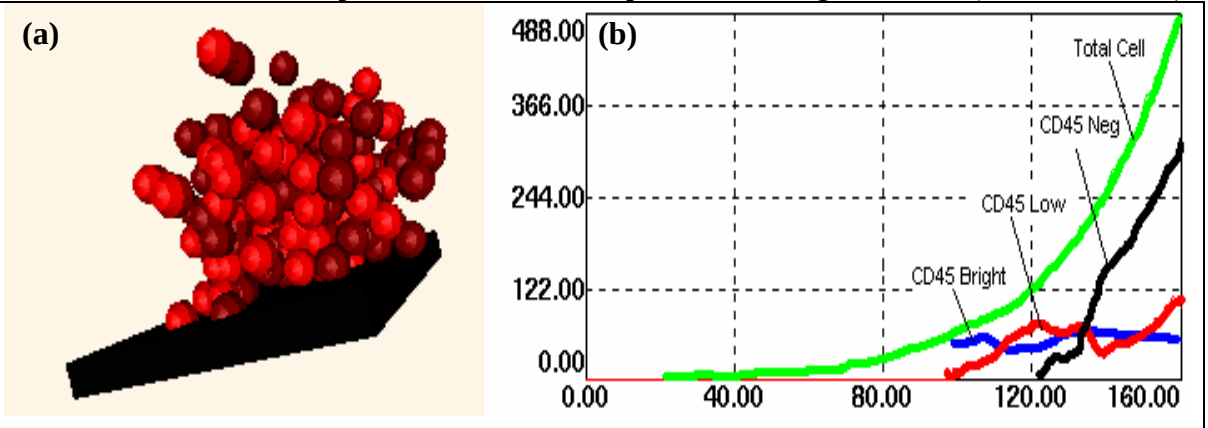


Figure 1: (a) 3D visualization ; (b) Kinetic of total cell (total cell), cell with hight expression of CD45 (CD45 Hight), cell with few CD45 (CD45 Low) cell with no CD45 (CD45 neg)

Holistic study, by the way of "In Virtuo" simulation enables to simulate emergent behaviour of MM. The simulation shows the importance of the microenvironement and the CD45 isoform for tumor evolution. Moreover, we assume that our original approach may allows research of new therapeutical targets and therapies.

Mathematical modelling of the apoptosis

N. Reymond[1], E. Grenier[1] & J.-P. Boissel[2]

[1]Unité de Mathématiques Pures et Appliquées, UMR CNRS 5669, ENS Lyon, France
[2]Service de Pharmacologie Clinique, EA 643, Faculté de médecine Laënnec, Université Claude Bernard, Lyon, France

Abstract

Programmed cell death or apoptosis is a highly conserved pathway that is functional in all higher organisms. This mechanism eliminates defect cells without damaging neighbouring tissue cells and is therefore recommended for tissue maintenance. However failure in apoptosis regulation is a major cause of diseases, like stroke (increased apoptosis) and cancer (decreased apoptosis). A mathematical model of apoptosis integrating the presently distributed and heterogenous knowledge about apoptosis in an integrated model would be of great benefit, since it allows the identification of most sensitive signaling molecules and predictions on the systemic behavior of apoptotic signaling, upon stimulation by different molecules or through interaction of chemotherapeutics. In fact, the question of a threshold for induction of apoptosis plays a central role in the understanding of the sensitivity and resistance cells toward various chemotherapeutic agents.

The aim of this work is to take into account the two main caracterisitics of the apoptosis. First, to address the complexity of apoptosis signaling, the system is subdivided into subsystems of different information qualities. And second, major molecules involved in apoptosis pathways shown to be initially distributed in different compartments (cell membrane, cytosole, nucleus and mitochondria) and to dynamically change their localization during apoptosis. To address this spatial aspect, we have developed a diffusion-reaction model of apoptosis that is numerically described by a system of partial differential equations (PDE).

Towards a Topological Game of Life

Samir Saidani & Ludovic Guegan

Laboratoire GREYC, Université de Caen, France

Abstract

Cellular automata was first introduced by Stanislaw Ulam and John Von Neuman to study the self-reproduction process through a distributed model of computation. Indeed, cellular automata share with biological processes some key features, like their intrinsic distributed nature and emergence of complexity from very simple rules, as exhibited by John Conway's Game of Life. Basically, the Game of Life is a 2D cellular automata, *i.e.* an infinite grid where a node or cell becomes alive (black) when three living cells are around it (reproduction), and stays alive when two or three living cells are in its neighborhood (survival).

However, the underlying graph of a cellular automata is static and recent research in both natural and artificial networks have shown the crucial importance of dynamic networks, *i.e.* networks whose topology changes over time, present. Topological dynamicity, called here topodynamic, seems to be also a key feature of the emergence of complex structures in biological networks.

Thus we are especially interested in modeling the morphodynamic of systems,*i.e.* systems able to change their own shape. For this goal, let us consider the set $\mathcal{M}$ of all topologies on a given set X as the phase space of a dynamical system. Then a trajectory on this phase space represents the change of the shape of X and designing such a dynamical system in an appropriate manner could help in modeling and understanding the morphodynamic and morphogenesis of a real entity. Notice that recent works on topology theory have extended the notion of topology in several ways, so topological space should be understood as a space where a notion of topology can be defined on this space.

So we will call *topodynamical system* on the set X a dynamical system $(\mathcal{M}_X, f)$ where the phase space $\mathcal{M}_X$ is the set of all "topologies" on a given set X, and $f : \mathcal{M}_X \mapsto \mathcal{M}_X$ a map called *topodynamic*, which expresses the morphodynamic of this system. When we consider the phase space up to an "homeomorphism", we will call the map a *metamorphism*.

Here, we would like to focus on the topodynamic of a network. So let us now consider a discrete topodynamical system on a finite graph. We call dynamic graph a sequence of graphs consisting of an initial graph and a function which transforms its topology to a new topology. Nevertheless, we would like to have dynamic graph vertices active by themselves, namely able to change their own degrees by accepting, keeping or removing its adjacent edges, according to local transition rules inducing the global graph topodynamic. For this purpose, we introduce the dynamic graph principle as follow : the next *neighborhood* of a *node* depends on its current *neighborhood* and of the *neighborhood* of its neighbors.

The dynamic graph, defined here in the sense of discrete dynamical system, is a purely topological model of computation, and combine graph theory expressness with richness of cellular automata dynamic [1]. We show here by a simulation tool specially designed for dynamic graph [2] how it is possible to exhibit complex topodynamic structure with simple rules.

References

[1] Samir Saidani. Self-reconfigurable robots topodynamic. In *Proceedings, IEEE Int. Conf. on Robotics & Automation (ICRA'04)*, pages 2883–2887, New Orleans, Louisiana, USA, 2004.

[2] Samir Saidani and Michaël Piel. Dynagraph : a smalltalk environment for self-reconfigurable robots simulation. In *European Smalltalk User Group (ESUG'04), Köthen, Germany*, 2004.

Understanding organelle biogenesis by integration of high-throughput functional genomics approaches

C Andres, A Avon, N Dautrevaux, A Falcon de Longevialle, L Heurtevin, C Lurin, A Marmagne, C Schmitz-Linneweber, I Small

Unité de Recherche en Génomique Végétale (INRA/CNRS/UEVE), 2, rue Gaston Crémieux, CP 5708, 91057 Evry cedex

The major organelles in green plant cells are the chloroplast (the site of photosynthesis) and the mitochondrion (the site of respiration). Most of the mitochondrial and chloroplast genomes have been transferred to the nucleus during evolution, and the nucleus now coordinates key aspects of their activity. In fact, up to one-quarter of plant nuclear genes encode proteins that are located to these three organelles, yet the function of the majority of these proteins are unknown or poorly defined. Organelle biogenesis requires (a) replication and maintenance of the organellar genomes, (b) correct expression of the organellar genes and (c) import of nuclear-encoded proteins required for organelle function. All three of these processes are important control points requiring nuclear/organellar coordination. This coordination is known to rely on retrograde regulation signals that act on nuclear gene expression and nucleus-encoded factors that regulate organellar gene expression. Our aim is to understand this regulatory network that coordinates nuclear and organellar gene expression as a function of the needs of the plant. Organellar gene expression is controlled in a post-transcriptional manner by nucleus-encoded RNA-binding factors required for processing, splicing, editing and translating transcripts. The majority of the factors probably belong to a single protein family characterized by tandem arrays of pentatricopeptide repeats (PPRs). We have undertaken a detailed bioinformatic analysis of 441 members of the Arabidopsis PPR family including genomic and genetic data on the expression (microarray data), localization (green fluorescent protein and red fluorescent protein fusions), and general function (insertion mutants and RNA binding assays) of many family members [1]. The results suggest that PPR proteins are transcript-specific factors that control expression of single organellar genes.
Further progress will be dependant on our ability to integrate data from several multiparallel approaches to gain insights into the functioning of the system as a whole.

References

[1] Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M-L., Mireau, H., Peeters, N., Renou, J-P., Szurek, B., Taconnat, L., Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell, 16, 2089-2103

Modelisation of early Hematopoiesis with Hybrid Functional Petri Nets

David Campard[1], Janine Guespin[2], Fariza Tahi[3],
Sylvie Troncale[3] and Jean-Pierre Vannier[1]

[1]Groupe de Recherche « Micro-Environnement et Renouvellement Cellulaire » EA 2122
[2]Université de Rouen
[3]Laboratoire de Mathématiques et d'Informatique- UMR 8042
[3]Université d'Evry Val d'Essonne- CNRS- Génopole

I. Introduction

The cellular production constitutes one of the essential characters of hematopoiesis. The set of blood cells are continuously renewed from a small pool of totipotent cells in bone marrow, called hematopoietic stem cells (HSCs). The maintenance of HSCs population is enabled thanks to a mechanism of self-renewal *in vivo*. Nevertheless our knowledge of hematopoietic regulation does not allow maintaining these cells *in vitro* without loss of the long-term multilineage growth and differentiation properties required for their clinical utility [1].

In order to understand which regulation entitles keeping the HSCs pool, we concentrate on early hematopoiesis. A hypothesis is that a stem cell enters in differentiation thanks to an epigenetic modification controlled by an autocrine substance: the receptor of Interleukin-6 (IL-6R) [2]. Indeed this molecule is a candidate for a positive retrocontrol loop which would enable the epigenetic modification of HSCs.

Nevertheless, the study of this hypothesis is complex, since the difficulty of biological experiments. In this way, we work on the conception of a model simulating regulation at early time of hematopoiesis. For our project, we first use Hybrid Functional Petri Nets [3][4], which are Petri nets permitting discreet and continuous modelisation, as well as production and consumption of resources. All these aspects are primordial in order to model respectively, cellular evolution, molecular evolution and concentration variation.

II. Hematopoiesis regulation

1) Molecular factors regulation

Hematopoiesis has to be extremely regulated in order to maintain the number of blood cells, whatever cellular lineage or environmental conditions. The regulation of hematopoiesis is controlled by a large number of molecular factors, which can be inhibitive or activator. Secretion of growth factors is insured by cell-matrix. However recent studies demonstrate that HSCs can also secrete molecular factors intervening in autocrine or paracrine regulation of cellular mechanisms such as multiplication, differentiation, quiescence or apoptosis.

2) Regulation of IL-6

Thus HSCs constitutively secrete a receptor of a cytokine called Interleukin-6 (IL-6). First IL-6 specifically binds with its receptor (IL-6R) existing under shape soluble (IL-6Rs) or membranar (IL-6Rm) *(Figure 1).* This complex of low affinity (IL-6|IL-6R) recruits the transducer subunits gp130. The complex gp130|IL-6|IL-6R internalizes and activates the transduction signal. The configuration gp130|IL-6|IL-6R enables the fixation of JAK *(Janus Kinase)*, which leads to a set of phosphorylations. The phosphorylized receptors recruit the STAT *(Signal Transducer and Activator of Transcription).* These STAT act as transcription factors and activate the synthesis of gp130, IL-6Rm as well as self-renewal. The synthesized IL-6Rm are thus truncated to IL-6Rs. Moreover it exists a negative retrocontrol, which

inhibits the effects of the JAK and the STAT, via presumably a SOCS pathway *(Suppressor Of Cytokine Signaling).*

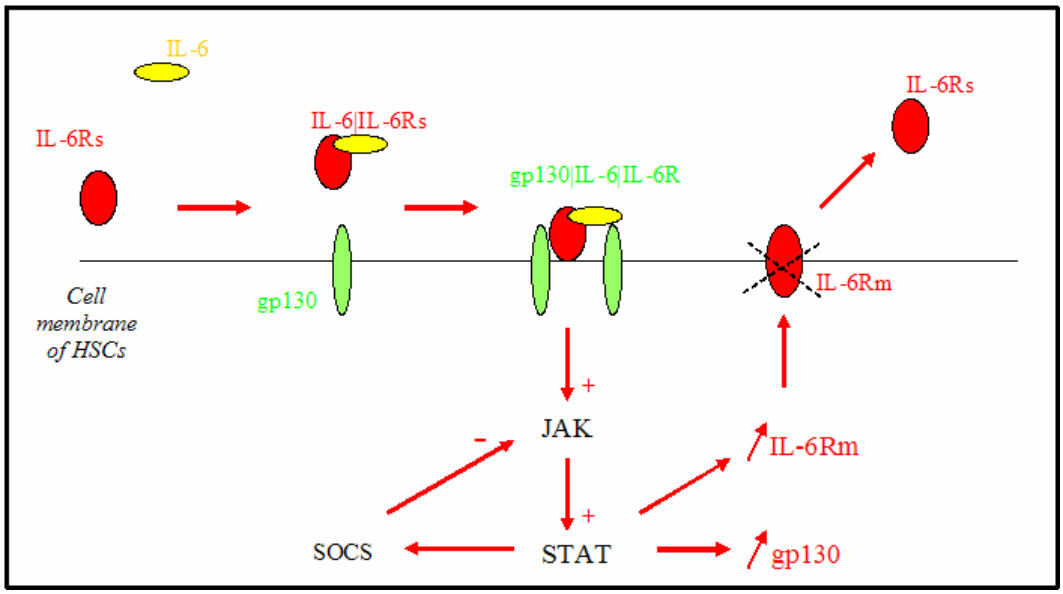


Figure 1: *Signalling pathway of the cytokine Interleukin-6.*

III. Hybrid Functional Petri Nets (HFPN)

To model this regulation several aspects are indispensable. First we need nets which model discreet time for cellular evolution as well as continuous time for molecular concentration. Moreover, the notion of consumption and production of resources is necessary to model variation of molecular and cellular concentration. In a first time we use Hybrid Functional Petri Nets which conciliate all these points. These nets are Petri nets which bind the notion of discreet and continuous and which enable inserting functions in connectors and processes.

The Figure 2 presents notation used in HFPN:

- Discreet places represent entities whose the associated number fits in with the number of tokens.
- Continuous places represent entities which we associate real number, such as concentration.
- Discreet transitions which use delay of time.
- Continuous transitions which continuously fire to a speed v(t).

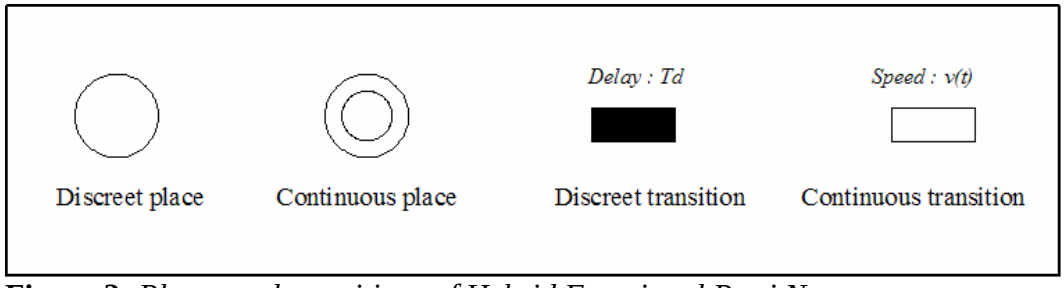


Figure 2: *Places and transitions of Hybrid Functional Petri Nets.*

IV. Cellular model

Only HSCs have capacity of self-renewal. It's know that among these primitive cells, 75% are quiescent (at the G0 state of the cellular cycle), and 25% are permissive (in division). Given *Pq* the quiescent cells, *Pp* the permissive cells and *C* the mature cells. We will present

all evolution pathways of these cells [5][7]. The cellular evolution is in discreet time since we enumerate a number of cells *(Figure 3).*

- A *Pq* cell can enter in the cellular cycle to become a *Pp* cell and inversely (Transitions T1 and T2)
- A *Pp* cell can renew itself :
  - ✓ with symmetric division, when the two daughter cells are identical (T3)
  - ✓ with asymmetric division, when the two daughter cells are different. One is the same as the mother and the other is differentiated. (T4)
- A *Pp* can differentiate :
  - ✓ without division (T5)
  - ✓ with division (T6)

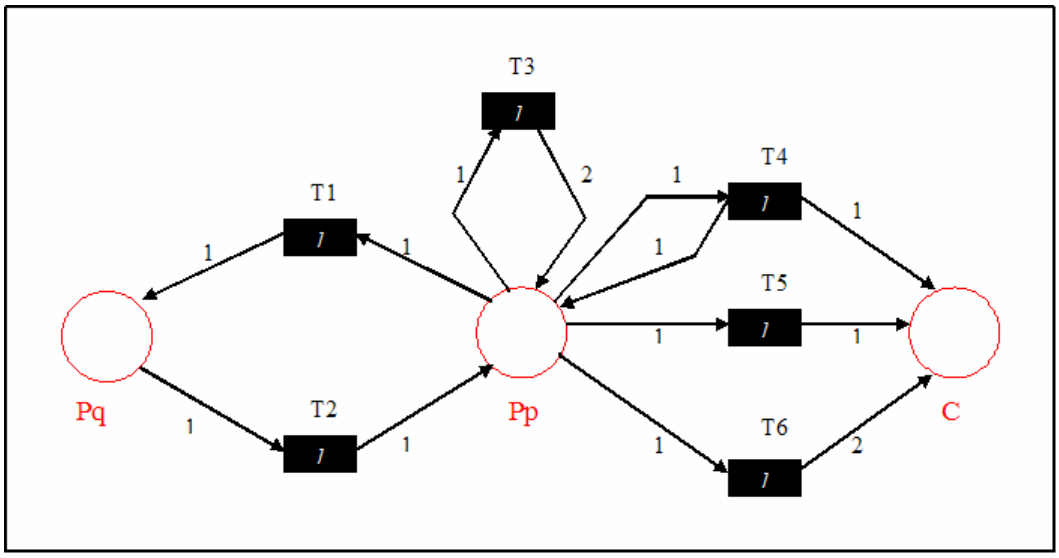


Figure 3: *Cellular model.*

On the connectors we can read the number of cells which go to the transition as well as the number which is obtained after the division. Finally, the delays on the transitions have as unity the day.

V. Molecular model

The modelisation of molecular interaction requires continuous time in order to represent molecular concentration *(Figure 4).* We need places for all molecules, IL-6, IL-6Rs, IL-6Rm, gp130 and the complexes IL-6|IL-6R and gp130|IL-6|IL-6R. The transitions T7, T8, T9 and T10 mean the formation of the complexes whereas T11, T12 and T13 show dissociation. Thus we represent these interactions :

- IL-6 + IL-6Rs ⟷ IL-6|IL-6Rs
- IL-6 + IL-6Rm ⟷ IL-6|IL-6Rm
- IL-6|IL-6Rs + 2gp130 ⟷ gp130|IL-6|IL-6Rs
- IL-6|IL-6Rm + 2gp130 ⟷ gp130|IL-6|IL-6Rm

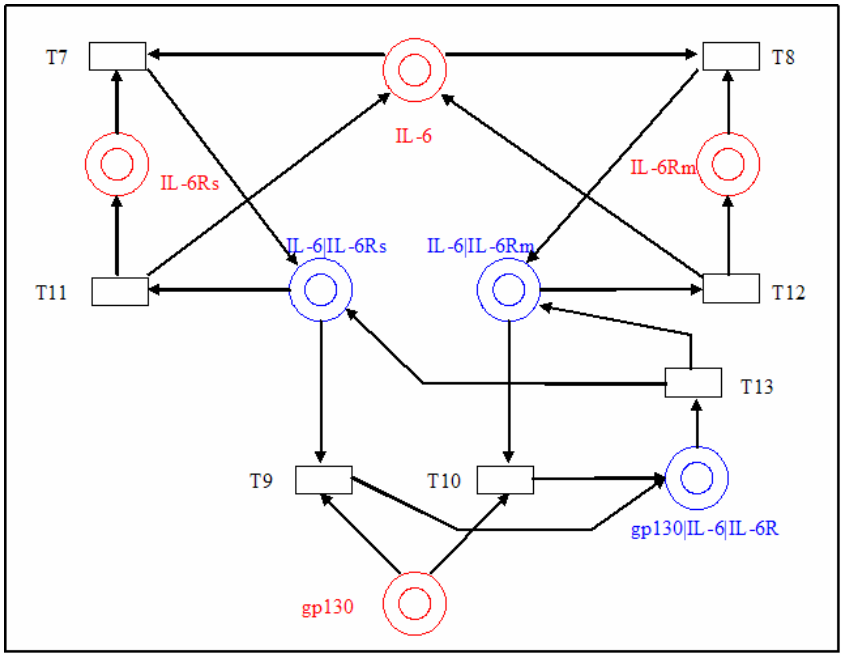


Figure 4 : *Molecular model.*

VI. Signalling pathway

In order to bring together these two parts of the model, we use the signalling pathway described above *(Figure 1).* Thus we have to consider kinetics of the different steps *(Figure 5)* [6]. When the complex gp130|IL-6|IL-6R is formed, this is internalized for one hour (Transition T15). In order to represent this delay, we use a discreet transition (T14). In the same way we use a second discreet transition (T16) to show the negative loop acting as a refractor period of eight hours. After this delay, we observe the synthesis of gp130, IL-6Rm as well as the cleavage of IL-6Rm to IL-6Rs (T17). This pathway also enables the activation of self-renewal of permissive cells (Pp). The places called Pv are virtual entities that haven't any biological meaning.

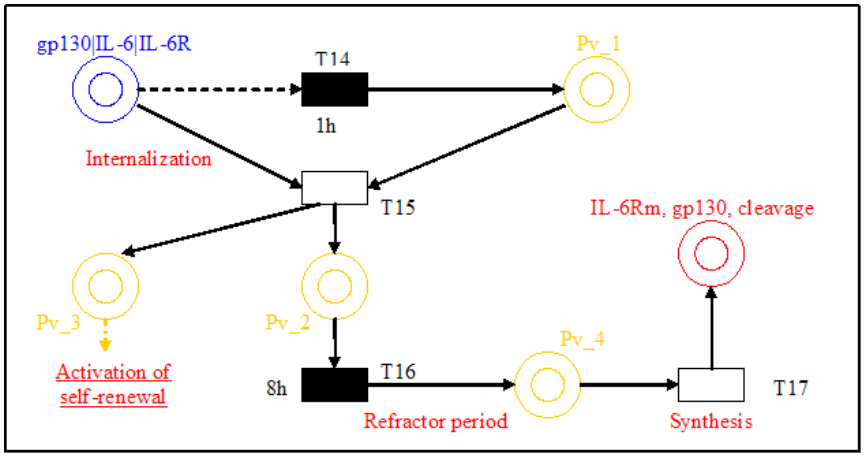


Figure 5: *Model of the signalling pathway*

Thus the bond between the molecular and cellular models is made. Concatenating these models we obtain a part of the global model *(Figure 6).* The transition T18 describes the synthesis of IL-6Rs by permissive cells (Pp). In red, we can see the retroaction positive loop. Indeed, Pp cells synthesize IL-6Rs enabling the formation of gp130|IL-6|IL-6R. When sufficiently complexes are formed the signalling pathway is activated allowing self-renewal as well as synthesis of IL-6Rm whose a part is truncated to IL-6Rs *(not shown in the figure).* This synthesis entitles again the formation of gp130|IL-6|IL-6, and so on.

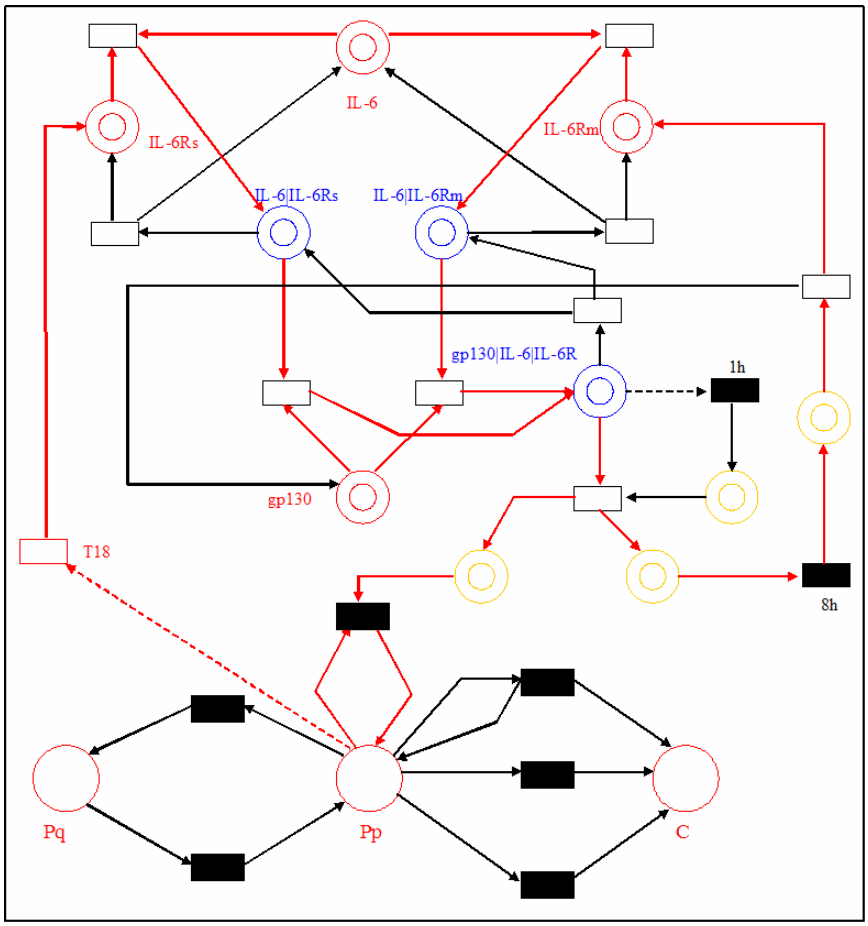


Figure 6: *A part of the global model.*

VII. Results and discussion

The presented model is not complete. Nevertheless it permits understanding the following procedure to model early hematopoiesis.

With this model we obtain the disappearance of HSCs as observed *in vitro*. Thanks to this model we would validate or refute the hypothesis of an epigenetic modification of HSCs thanks to the mechanisms of a positive retroaction of IL-6. An epigenetic modification involves that under a threshold HSCs differentiate and above it they renew. To test this hypothesis we need to model pulse of different molecules, such as IL-6Rs in order to try to go beyond this threshold. In spite of pulse we first observe the activation of self-renewal and then all cells go to differentiation. The biological experiments obtain the same loss of primitive cells.

Thus our future work is to vary parameters of the models, in order to try getting a set of parameters enabling the survival of a pool of HSCs. Following the results, we will incorporate other cytokines in the model, such as TGF-β *(Transforming Growth Factor)* which keep primitive cells in quiescent state.

References

[1]. "A ligand-receptor signaling threshold model of stem cell differentiation control: a biologically conserved mechanism applicable to Hematopoiesis". P.Zandstra, D.Lauffenburger and C.Eaves. *Blood*. 2000.

[2]. "Interleukin-6 and Soluble Interleukin-6 Receptor : Direct Stimulation of gp130 and Hematopoiesis". M.Peters, A.Müller and S.Rose-John. *Blood.* 1998.

[3]. "Biopathways representation and simulation on hybrid functional Petri net". H.Matsuno, Y.Tanaka, H.Aoshima, A.Doi, M.Matsui and S. Miyano. *In silico Biology*.2003

[4]. “Constructing biological pathway models with hybrid functional Petri nets”. A.Doi, S.Fujita, H.Matsuno, M.Nagasaki and S.Miyano. *In Silico Biology*. 2004.
[5]. “Asymmetric Cell Divisions Sustain Long-Term Hematopoiesis from Single-sorted Human Fetal Liver Cells”. Brummendorf et al. *The Rockefeller University Press*. 1998.
[6].”In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells”. S.Cheshier, S.Morrison, X.Liao and L.Weissman. *Medical Sciences*. 1999.
[7]. “Towards predictive models of stem cell fate”. S.Viswanathan and W.Zandstra. *Cytotechnology*. 2003.