Proceedings of
The Nice Spring School
on

Modelling Complex Biological Systems in the Context of Genomics

March 30th - April 3rd, 2009

Edited by

Patrick Amar, François Képès, Vic Norris, Gilles Bernot

"But technology will ultimately and usefully be better served by following the spirit of Eddington, by attempting to provide enough time and intellectual space for those who want to invest themselves in exploration of levels beyond the genome independently of any quick promises for still quicker solutions to extremely complex problems."

Strohman RC (1977) Nature Biotech 15:199

FOREWORD

What are the salient features of the new scientific context within which biological modelling and simulation will evolve from now on? The global project of high-throughput biology may be summarized as follows. After genome sequencing comes the annotation by 'classical' bioinformatics means. It then becomes important to interpret the annotations, to understand the interactions between biological functions, to predict the outcome of perturbations, while incorporating the results from post genomics studies (of course, sequencing and annotation do not stop when simulation comes into the picture). At that stage, a tight interplay between model, simulation and bench experimentation is crucial. Taking on this challenge therefore requires specialists from across the sciences to learn each other's language so as to collaborate effectively on defined projects.

Just such a multi-disciplinary group of scientists has been meeting regularly at Genopole, a leading centre for genomics in France. This, the *Epigenomics project*, is divided into six subgroups. The *GolgiTop* subgroup focuses on membrane deformations involved in the functionning of the Golgi. The *Hyperstructures* subgroup focuses on cell division, on the dynamics of the cytoskeleton, and on the dynamics of *hyperstructures* (which are extended multi-molecule assemblies that serve a particular function). The *Organisation* subgroup has adopted a systems biology approach with the application and development of new programming languages to describe biological systems which it has been applying to problems in the growth and differentiation of plants and in the structure and functioning of mitochondria. The *Observability* subgroup addresses the question of which models are coherent and how can they best be tested by applying a formal system, originally used for testing computer programs, to an epigenetic model for mucus production by *Pseudomonas aeruginosa*, the bacterium involved in cystic fibrosis. The *Bioputing* group works on new approaches proposed to understand biological computing using computing machine made of biomolecules or bacterial colonies. The *SMABio* subgroup focuses on how multi-agents systems (MAS) can be used to model biological systems.

The works of subgroups underpinned the conferences organised in Autrans in 2002, in Dieppe in 2003, in Evry in 2004, in Montpelliers in 2005, in Bordeaux in 2006, back to Evry in 2007 and in Lille in 2008. The conferences in Nice in 2009 which as reported here, brought together over a hundred participants, biologists, physical chemists, physicists, statisticians, mathematicians and computer scientists and gave leading specialists the opportunity to address an audience of doctoral and post-doctoral students as well as colleagues from other disciplines.

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

Patrick Amar, Gilles Bernot, Marie Beurton-Aimar, Jean-Louis Giavitto, Christophe Godin, Bruno Goffinet, Janine Guespin, François Képès, Jean-Pierre Mazat, Franck Molina, Victor Norris, Vincent Schächter, Bernard Vandenbunder.

ACKNOWLEDGEMENTS

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

Of course the organisation team would like to express gratitude to all the staff of the Club Belambra VVF *La Bergerie* for the very good conditions we have found during the conference.

Special thanks to the Epigenomics project for their assistance in preparing this book for publication. The cover photography shows The *Baie des Anges* Copyright *Ville de Nice*.

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

- Centre National de la Recherche Scientifique (CNRS):
  http://www.cnrs.fr
- Genopole® Evry:
  http://www.genopole.fr
- GDRE CNRS 513 Biologie Systémique:
  http://www.mpi-magdeburg.mpg.de/CNRS_MPG
- INRIA Sophia Antipolis - Méditerrannée:
  http://www-sop.inria.fr/
- Réseau national des systèmes complexes (RNSC):
  http://rnsc.csregistry.org/tiki-index.php
- GDR CNRS 3003 Bioinformatique Moléculaire:
  http://www.gdr-bim.u-psud.fr
- Université de Nice Sophia Antipolis:
  http://portail.unice.fr
- Laboratoire I3S, Sophia Antipolis:
  http://www.i3s.unice.fr/I3S
- Fondation Scientifique Fourmentin-Guilbert:
  http://www.fourmentinguilbert.org

THE EDITORS

INVITED SPEAKERS

PASCAL BARBRY	Université de Nice, Sophia Antipolis (F)
FRÉDÉRIQUE CLÉMENT	INRIA Paris-Roquencourt (F)
OLIVIER GANDRILLON	Université de Lyon 1 (F)
MARCELLINE KAUFMAN	ULB, Bruxelles (Belgium)
FRANÇOISE LIVOLANT	Laboratoire de Physique des Solides, Orsay (F)
CLAUDE LOBRY	INRIA Sophia Antipolis (F)
JEAN-PIERRE MAZAT	Université Bordeaux 2 (F)
JUDIT OVADI	Biological Research Center, Budapest (Hungary)
JACQUES PROST	ESPCI and Curie Institute, Paris (F)
UDO REICHL	Max Planck Institute, Magdeburg (D)
GERMAN RIVAS	Centro de Investigaciones Biológicas, Madrid (Sp)
PAUL RUET	Institut de Mathématiques de Luminy, Marseille (F)
CHRISTIAN SARDET	Université P. et M. Curie, Villefranche sur Mer (F)
FRANCISCA SÀNCHEZ JIMÉNEZ	University of Málaga (Sp)
RANDY THOMAS	Université d'Évry (F)
MARIE-JOËLLE VIROLLE	Institut de Génétique et Microbiologie, Orsay (F)

CONTENTS

PART I INVITED TALKS

From cell to tissue dynamics

Jacques Prost[1,2]

[1] Physicochimie Curie (CNRS-UMR168), Institut Curie, Section de Recherche, 26 rue d'Ulm 75248 Paris Cedex 05 France

[2] E.S.P.C.I, 10 rue Vauquelin, 75231 Paris Cedex 05, France

Abstract

I give a survey of the work which we have done over the last years on "active gels". In particular, I show how one can construct a set of equations describing gels in which the cross-links can be moved around by active elements constantly consuming energy. This situation corresponds to the cell cytoskeleton, which is thought to control most of cell dynamics. We illustrate the potential usefulness of the equations by discussing cell behavior such as motility, oscillations, wound healing and cytokinesis. Eventually, I discuss tissue growth, introducing the concept of homeostatic pressure and tumor nucleation.

1 Introduction

Our knowledge in Biology has improved significantly over the last fifty years, with impressive successes in molecular biology, genetics, developmental and cell biology. The wealth of information is such that it is hard to make use of all of them. Although it is clear that details matter in biological systems, it is also clear that one currently needs to develop a global picture taking into account the main features and recognizing what is universal. Cell biology provides a good example of this need: with exactly the same genome cells can differentiate in about three hundred different types in complex animals such as vertebrates [1]. Physicists would say that they can go to three hundred stable attractors depending on external conditions. Considering that cell phase space is controlled among other things by the expression of a few 10^4 genes, three hundred is a very small number. A possible explanation for this small number of cell types is that they are not only controlled by gene expression, but that they are also constrained by generic physical laws. We are far from being able to discuss this problem in its generality, but in the following I address a simpler problem which illustrates how physics could provide generic tools for raising these questions. Namely I investigate what can be learned from using symmetry arguments and conservation laws in describing cell morphology and dynamics. In view of the acknowledged specificity of biology such an endeavor may seem futile. I hope to convince the attendee that it is on the contrary helpful. Indeed I will present:

- one simple aspect of cell motility, namely the shape and speed of a lamellipodium, thin protrusion leading the cell motion on a substrate,

- cell oscillations which are observed when cells are suspended in a physiological serum,
- wound healing of xenopus eggs and the onset of cytokinesis,
- tissue growth in specific geometrical conditions.

For the first three examples I use the the same theoretical framework. Lessons and numbers learned from one example are useful to understand the next. For tissue growth the approach is very similar, but conservation laws differ.

In order to do so, one needs to construct the tools. It is nowadays textbook knowledge that the shape of cells is maintained by a network of cross-linked biofilaments: the cytoskeleton [2]. At this stage, all we need to know is that the network constitutes a physical gel which would be rather conventional in the absence of molecular motors. At short time scales, it behaves like a conventional solid, at long time scales like a liquid. There are in fact some added complexities which will be discussed in the conclusion. The essential novelty comes from molecular motors. They consume continuously ATP (Adenosine Triphosphate) and are able to exert stresses on the cross-links of the gel. The question is then how to describe such a gel, which we call ”active”. Using conservation laws and symmetry arguments only I will derive the relevant equations. Since they result from general considerations these equations can describe many different situations and are very similar to those derived in different contexts such as motions of bacterial colonies, fish shoals and bird flocks [3, 4, 5]. Active gels could also be made artificially, leading to original material properties [6]. Some of these expected original properties, are the spontaneous transition to a moving state of a thin slab and the rotation of disclinations.

The main difference between tissue growth description and active gel description comes from the fact that cell number is not conserved. Although this difference may seem to be innocuous, it has profound influence on growth properties. In particular it allows for the definition of the homeostatic pressure of a tissue in a given biochemical environment. This pressure is different from the hydrostatic pressure and corresponds to the mechanical force per unit area that a tissue can exert on a wall permeable to nutrients and growth factors, when duplication exactly balances apoptosis. Orders of magnitude of the homeostatic pressure may be obtained from the analysis of cell behavior presented before. I will further show that if two tissues are in competition

for finite space, the tissue which has the largest homeostatic pressure will always win with no compromise. If one adds that the tissues interface exhibit (active) surface tension, this leads naturally to the concept of nucleation of small tumors in three dimensions. This concept is consistent with experiments showing ”metastasis inefficiency”.

Acknowledgments

The work I present here results from very pleasant collaborations with M. Basan, J.F. Joanny, F.Julicher, K.Kruse, T. Risler, G. Salbreux, K. Sekimoto on the theory side, and M. Bornens, E. Paluch, P. Pullarkat, C. Sykes and X. Sastres on the experimental side.

The reference list is intentionally long to help the reader start its own bibliography.

References

[1] M. Kirshner, J. Gerhart, The Plausibility of Life: Resolving Darwin's dilemma, Yale Uniersity Press, 2007.

[2] *et al.*. B. Alberts, Molecular Biology of the Cell, Garland, 2002.

[3] R. Simha, S. Ramaswamy, Hydrodynamic fluctuations and instabilities in ordered suspensions of self-propelled particles, Phys.Rev.Lett. 89 (2002) 058101.

[4] J. Toner, Y. Tu, Long range order in a two-dimensional dynamical xy model: how birds fly together?, Phys.Rev.Lett. 75 (1995) 4326–4329.

[5] I. Tuval, L. Cisneros, C. Dombrowski, C. W. Wolgemuth, J. O. Kessler, R. E. Goldstein, Bacterial swimming and oxygen transport near contact lines., Proc Natl Acad Sci U S A 102 (7) (2005) 2277–2282.
URL `http://dx.doi.org/10.1073/pnas.0406724102`

[6] J. Prost, Conference solvay, in: Chemistry of molecular motors, 2007.

[7] K. Kruse, J. F. Joanny, F. Jï¿½licher, J. Prost, K. Sekimoto, Asters, vortices, and rotating spirals in active gels of polar filaments., Phys Rev Lett 92 (7) (2004) 078101.

[8] K. Kruse, J. Joanny, F.Jï¿½licher, J.Prost, K.Sekimoto, Generic theory of active polar gels: a paradigm for cytoskeletal dynamics, Eur.Phys.J.E 16 (2005) 5–16.

[9] F. Julicher, K. Kruse, J. Prost, J. F. Joanny, Active behavior of the cytoskeleton, Physics Reports 449 (2007) 3–28.

[10] T.Svitkina, G.G.Borisi, Correlative light and electron microscopy of the cytoskeleton of cultured cells, Methods Enzymol. 298 (1998) 570-592.

[11] P. DeGennes, J. J. Prost, The Physics of Liquid crystals, Oxford University Press, 1993.

[12] P. M. Bendix, G. H. Koenderink, D. Cuvelier, Z. Dogic, B. N. Koeleman, W. M. Brieher, C. M. Field, L. Mahadevan, D. A. Weitz, A quantitative analysis of contractility in active cytoskeletal protein networks., Biophys J 94 (8) (2008) 3126–3136.
URL http://dx.doi.org/10.1529/biophysj.107.117960

[13] Y. Hatwalne, S. Ramaswamy, M. Rao, R. Simha, Rheology of active-particle suspensions, Phys.Rev.Lett. 92 (2004) 118101.

[14] L. Giomi, M. C. Marchetti, T. Liverpool, Complex spontaneous flows and concentration banding in active polar films, phys.Rev.Lett (2008).

[15] R. Voituriez, J. Joanny, J. J. Prost, Spontaneous flow transtions in active polar gels, Europhys. Lett. 70 (2005) 404–410.

[16] R. Voituriez, J. F. Joanny, J. Prost, Generic phase diagram of active polar films., Phys Rev Lett 96 (2) (2006) 028102.

[17] A. . Basu, J. Joanny, F. Julicher, J. J. Prost, Thermal and non-thermal fluctuations in active polar gels, submitted to European Physical Journal E (2008).

[18] J. Prost, C. Barbetta, J.-F. Joanny, Dynamical control of the shape and size of stereocilia and microvilli., Biophys J 93 (4) (2007) 1124–1133.
URL http://dx.doi.org/10.1529/biophysj.106.098038

[19] P. Vallotton, G. Danuser, S. Bohnet, J.-J. Meister, A. B. Verkhovsky, Tracking retrograde flow in keratocytes: news from the front., Mol Biol Cell 16 (3) (2005) 1223–1231.
URL http://dx.doi.org/10.1091/mbc.E04-07-0615

[20] F. Gerbal, P. Chaikin, Y. Rabin, J. Prost, An elastic analysis of listeria monocytogenes propulsion., Biophys J 79 (5) (2000) 2259–2275.

[21] A. C. Callan-Jones, J.-F. Joanny, J. Prost, Viscous-fingering-like instability of cell fragments., Phys Rev Lett 100 (25) (2008) 258106.

[22] K. Kruse, J. F. Joanny, F. Jï¿½licher, J. Prost, Contractility and retrograde flow in lamellipodium motion., Phys Biol 3 (2) (2006) 130–137.
URL http://dx.doi.org/10.1088/1478-3975/3/2/005

[23] Y. Marcy, J. Prost, M.-F. Carlier, C. Sykes, Forces generated during actin-based propulsion: a direct measurement by micromanipulation., Proc Natl Acad Sci U S A 101 (16) (2004) 5992–5997.
URL http://dx.doi.org/10.1073/pnas.0307704101

[24] E. Paluch, M. Piel, J. Prost, M. Bornens, C. Sykes, Cortical actomyosin breakage triggers shape oscillations in cells and cell fragments., Biophys J 89 (1) (2005) 724–733.
URL http://dx.doi.org/10.1529/biophysj.105.060590

[25] G. Salbreux, J. F. Joanny, J. Prost, P. Pullarkat, Shape oscillations of non-adhering fibroblast cells., Phys Biol 4 (4) (2007) 268–284.
URL http://dx.doi.org/10.1088/1478-3975/4/4/004

[26] J. H. Zang, J. A. Spudich, Myosin ii localization during cytokinesis occurs by a mechanism that does not require its motor domain., Proc Natl Acad Sci U S A 95 (23) (1998) 13652–13657.

[27] M. Zhou, Y.-L. Wang, Distinct pathways for the early recruitment of myosin ii and actin to the cytokinetic furrow., Mol Biol Cell 19 (1) (2008) 318–326.
URL http://dx.doi.org/10.1091/mbc.E07-08-0783

[28] D. J. Fishkind, Y. L. Wang, Orientation and three-dimensional organization of actin filaments in dividing cultured cells., J Cell Biol 123 (4) (1993) 837–848.

[29] C. Mandato, W. Bement, Contraction and polymerization cooperate to assemble and close actomyosin rings around xenopus oocyte wounds, J.Cell Biol. 154 (2001) 785–797.

[30] C. A. Mandato, W. M. Bement, Actomyosin transports microtubules and microtubules control actomyosin recruitment during xenopus oocyte wound healing., Curr Biol 13 (13) (2003) 1096–1105.

[31] G. Salbreux, J. Prost, J.-F. J.F. Joanny, Hydrodynamics of cellular cortical flows and formation of contractile rings, physs.Rev.Lett (2009).

[32] Cameron, M. D., Schmidt, E. E., Kerkvliet, N., Nadkarni, K. V., Morris, V. L., Groom, A. C., Chambers, A. F., and MacDonald, I. C. (2000). “Temporal Progression of Metastasis in Lung: Cell Survival, Dormancy,

and Location Dependence of Metastatic Inefficiency 1." *Cancer Research* **60**, 2541–2546.

[33] Chambers, A. F., Groom, A. C., and MacDonald, I. C. (2002). "Dissemination and growth of cancer cells in metastatic sites." *Nat Rev Cancer* **2**, 563–572.

[34] Chambers, A. F. and Wilson, S. (1988). "Use of Neo R B16F1 murine melanoma cells to assess clonality of experimental metastases in the immune-deficient chick embryo." *Clinical and Experimental Metastasis* **6**, 171–182.

[35] Duguay, D., Foty, R. A., and Steinberg, M. S. (2003). "Cadherin-mediated cell adhesion and tissue segregation: qualitative and quantitative determinants." *Developmental Biology* **253**, 309–323.

[36] Engler, A. J., Richert, L., Wong, J. Y., Picart, C., and Discher, D. E. (2004) "Surface probe measurements of the elasticity of sectioned tissue, thin gels and polyelectrolyte multilayer films: Correlations between substrate stiffness and cell adhesion." *Surface Science* **570**, 142–154.

[37] Fidler, I. J. (2003). "The pathogenesis of cancer metastasis: the seed and soil hypothesis revisited." *Nat Rev Cancer* **3**, 453–458.

[38] Forgacs, G., Foty, R. A., Shafrir, Y., and Steinberg, M. S. (1998) "Viscoelastic Properties of Living Embryonic Tissues: a Quantitative Study." *Biophysical Journal* **74**, 2227–2234.

[39] Foty, R. A., Forgacs, G., Pfleger, C. M., and Steinberg, M. S. (1994). "Liquid properties of embryonic tissues: Measurement of interfacial tensions." *Physical Review Letters* **72**, 2298–2301.

[40] Frisch, U., Hasslache, B., and Pomeau, Y. (1986). "Lattice-Gas Automata for the Navier-Stokes Equation." *Physical Review Letters 56* **14**, 1505-1508.

[41] Hanahan, D. and Weinberg, R. A. (2000). "The hallmarks of cancer." *Cell* **100** (1), 57–70.

[42] Helmlinger, G., Netti, P. A., Lichtenbeld, H. C., Melder, R. J., and Jain, R. K. (1997). "Solid stress inhibits the growth of multicellular tumor spheroids." *Nature Biotechnology* **15**, 778–783.

[43] Kong, H. J., Polte, T. R., Alsberg, E., and Mooney, D. J. (2005). "FRET measurements of cell-traction forces and nano-scale clustering

of adhesion ligands varied by substrate stiffness." *Proceedings of the National Academy of Sciences* **102**, 4300–4305.

[44] Lecuit, T. and Lenne, P. F. (2007). "Cell surface mechanics and the control of cell shape, tissue patterns and morphogenesis." *Nat Rev Mol Cell Biol* **8**, 633–644.

[45] Luzzi, K. J., MacDonald, I. C., Schmidt, E. E., Kerkvliet, N., Morris, V. L., Chambers, A. F., and Groom, A. C. (1998). "Multistep Nature of Metastatic Inefficiency Dormancy of Solitary Cells after Successful Extravasation and Limited Survival of Early Micrometastases." *American Journal of Pathology* **153**, 865–873.

[46] Marcy, Y., Prost, J., Carlier, M. F., and Sykes, C. (2004). "Forces generated during actin-based propulsion: A direct measurement by micromanipulation." *Proc. Natl. Acad. Sci. USA* **101**, 5992–5997.

[47] Schötz, E. M., Burdine, R. D., Jülicher, F., Steinberg, M. S., Heisenberg, C. P., and Foty, R. A. (2008). "Quantitative differences in tissue surface tension influence zebrafish germ layer positioning." *HFSP Journal* **2**, 42–56.

[48] Shraiman, B. I. (2005). "Mechanical feedback as a possible regulator of tissue growth." *Proc. Natl. Acad. Sci. USA* **102**, 3318–3323.

[49] Talmadge, J. E. and Fidler, I. J. (1982). "Evidence for the clonal origin of spontaneous metastases." *Science* **217**, 361–363.

[50] Weinberg, R. A. (2007). "The biology of cancer." Garland Science, New York.

How to play with DNA inside and outside of the bacteriophage capsid

Françoise Livolant[1]

[1] Laboratoire de Physique des Solides, Bât 510, UMR 8502, Université Paris-Sud, 91405 Orsay Cedex, France

Abstract

Tailed bacteriophages are complex macromolecular machineries that deliver their genome into the host cytoplasm while their capsid and tail remain bound to the cell surface. Although several models have been proposed, DNA organization is still unknown in the phage head but dimensions of the capsids are adjusted to keep DNA at a concentration close to 500mg/ml (interhelix distances close to 27Å) independent of the species. DNA ejection from the capsid is triggered by specific interaction of a phage tail protein with a receptor inserted in the wall of the bacteria. DNA progresses in the tail and is injected in the cytoplasm. For some species (T5, λ and SPP1), the bacterial receptors have been isolated, allowing to reconstitute the ejection process *in vitro* and to investigate the underlying mechanisms. Our goal here is to understand the DNA ejection process, using the bacteriophage T5.

1 DNA organization inside the capsid

Once triggered, the DNA release from the bacteriophage T5 is not an all-or-none process but occurs in a stepwise fashion as revealed by fluorescence imaging [1] of individual bacteriophages deposited on a surface and also by light scattering on whole T5 populations [2]. During this process, the concentration of DNA goes down from more than 500mg/ml to zero. We followed the organization of DNA inside the capsids at different steps of the ejection process. Our goal was to understand what happens when the amount of DNA progressively decreases. Using cryoElectron microscopy (cryoEM), we observed the organization of DNA inside individual capsids and correlated these observations with the lengths and concentrations of encapsidated DNA. We show that upon decrease of its length and concentration, DNA always occupies the total volume of the capsid and reorganizes under confinement. The single DNA chain undergoes several phase transitions, from constrained hexagonal to hexagonal, cholesteric and isotropic phases [3].

2 DNA ejection under pressure

Both theoretical [1, 2, 3] and experimental [4, 5, 6, 7] studies have shown that full packaging of the genome inside the capsid requires forces of the order of 50-100 pN that would correspond to internal pressures of the order of 50-100 Atm. These high values would result from the confinement and bending of the long double-stranded DNA chain [8, 9] (typically tens of μm with a persistence length of 50 nm) inside the small volume of the capsid (50 to 80 nm in diameter). It has been hypothesised that this internal pressure is responsible for DNA release after interaction with the receptor protein, in the absence of any external source of energy.

The role of pressure in the process can be investigated by opposing an external osmotic pressure to DNA ejection. The external pressure can be tuned precisely and over a very large range using solutions of an osmolyte that cannot permeate the capsid such as polyethylene glycol (PEG). According to theoretical models, a decrease of the length of ejected DNA is expected with the increase of the external pressure [4, 5, 6]. Experiments with phages λ/LamB and SPP1/YueB systems showed the behaviour predicted from theory. For bacteriophage T5 in the presence of its receptor FhuA, DNA ejection is sensitive to the pressure but the process is far more complex. In the high pressure range, (7-16 atm), the ejected DNA length decreases with the increase of the pressure, as predicted. However, in the low and moderate pressure ranges (2-7Atm), the behaviour is puzzling: multiple discrete steps coexist in the population independently of the applied pressure. Moreover, a fraction of the phages eject significantly more DNA than expected (and sometimes their total content). Under pressure, T5 DNA ejection can be blocked at different stages and, surprisingly, the DNA remaining inside the phage may be shorter or longer than the one expected considering an equilibrium between the internal pressure exerted by the non-ejected DNA packing and the outside. We suspect kinetics effects or structural characteristics of the tail and connector involved in the ejection process to play a yet unexplained role in this process [13].

3 Toroidal DNA structure

The bacteriophage can also be used as a nanocontainer, permeable to water and ions. Individual DNA chains, 3000 to 55000bp long ($1.4 - 18\mu m$) have been collapsed inside the volume of the bacteriophage capsid, (80 nm in diameter), after partial ejection of its DNA. The tetravalent cation spermine was diffused through the capsid to condense DNA under conditions chosen to determine a hexagonal packing of DNA. Using cryoEM, we describe the toroidal structure and show how the frustration arising between chirality and hexagonal pack-

ing combined with the strong curvature imposed by the small volume of the container impose phasing of the helices and variations the DNA helical pitch [14].
Many questions remain unanswered yet and would deserve more experiments and modelisation.

References

[1] Mangenot, S. M. Hochrein, J. Rädler1 and L. Letellier (2005) Real-Time Imaging of DNA Ejection from Single Phage Particles. *Current Biol.*, **15**, 430-435.

[2] M. de Frutos, L. Letellier and E. Raspaud, DNA ejection from bacteriophage T5: analysis of the kinetics and energetics. *Biophys. J.* **88** (2004), pp. 1364-1370.

[3] Leforestier A. and Livolant F. (in preparation)

[4] Kindt, J., Tzlil, S., Ben-Shaul, A. & Gelbart, W. M. (2001). DNA packaging and ejection forces in bacteriophage. *Proc. Natl. Acad. Sci. USA*, **98**, 13671-13674.

[5] Tzlil, S., Kindt, J. T., Gelbart, W. M. & Ben-Shaul, A. (2003). Forces and Pressures in DNA Packaging and Release from Viral Capsids. *Biophys. J.* **84**, 1616-1627.

[6] Purohit, P. K., Inamdar, M. M., Grayson, P. D., Squires, T. M., Kondev, J. & Phillips, R. (2005). Forces during Bacteriophage DNA Packaging and Ejection. *Biophys. J.* **88**, 851-866.

[7] Smith, D. E., Tans, S. J., Smith, S. B., Grimes, S., Anderson, D. L. & Bustamante, C. (2001). The bacteriophage phi 29 portal motor can package DNA against a large internal force. *Nature* **413**, 748-752.

[8] Fuller, D. N., Rickgauer, J. P., Jardine, P. J., Grimes, S., Anderson, D. L. & Smith, D. E. (2007). Ionic effects on viral DNA packaging and portal motor function in bacteriophage phi 29. *Proc. Natl. Acad. Sci. USA* **104**, 11245-50.

[9] Fuller, D. N., Raymer, D. M., Kottadiel, V. I., Rao, V. B. & Smith, D. E. (2007). Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proc. Natl. Acad. Sci. USA* **104**, 16868.

[10] Fuller, D. N., Raymer, D. M., Rickgauer, J. P., Robertson, R. M., Catalano, C. E., Anderson, D. L., Grimes, S. & Smith, D. E. (2007). Measurements of Single DNA Molecule Packaging Dynamics in Bacteriophage λ Reveal High Forces, High Motor Processivity, and Capsid Transformations. *J. Mol. Biol.* **373**, 1113-1122.

[11] Odijk, T. (1998). Hexagonally Packed DNA within Bacteriophage T7 Stabilized by Curvature Stress. *Biophys. J.* **75**, 1223-1227.

[12] Odijk, T. (2004). Statics and dynamics of condensed DNA within phages and globules. *Phil. Trans. Math, Phys. Eng. Sci.* **362**, 1497-1517.

[13] Leforestier A., Brasiles, S., de Frutos, M., Raspaud, E., Letellier, L., Tavares, P., Livolant, F. (2008) *J. Mol. Biol.* **384**, 730-739.

[14] Leforestier A., Livolant F (submitted)

HARNESSING THE FORCE: Physicochemical modulation of contractile bacterial division complexes in the test tube

Francisco Monroy[1], Germán Rivas[2a], Pedro Tarazona[3], Marisela Vélez[2b], Jesús Mingorance[4] and Miguel Vicente[2c]

[1] Dep. Química Física, Facultad de Química, Univ. Complutense, Madrid
[2a] Centro de Investigaciones Biológicas, CSIC, Madrid
[2b] Instituto de Catálisis, CSIC, Madrid
[2c] Centro Nacional de Biotecnología, CSIC, Madrid
[3] Dep. Física Teórica, Facultad de Ciencias, Univ. Autónoma, Madrid
[4] Unidad de Investigación, Hospital La Paz, Madrid

Abstract

The components of the bacterial cell division machinery assemble to form a dynamic ring at mid-cell towards the end of the cell cycle. The ring is formed by at least ten division specific proteins, most of them integral membrane proteins. The first multi-protein complex formed is the bacterial proto-ring that initiates division. In *E. coli* the proto-ring is a complex of three proteins (FtsZ, FtsA, and ZipA) assembling on the cytoplasmic membrane, which is required for the incorporation of the remaining proteins at the mature ring [1]. The GTP-mediated assembly and disassembly of FtsZ (a bacterial ancestor of the eukaryotic tubulin) are thought to be essential for the formation of the septal ring. We will review our contributions to understand quantitatively the behavior of *E. coli* FtsZ (in different states, from soluble monomers, to single filaments and bundles) and how it is modulated by membrane components of the division machinery, namely the ZipA protein and the lipid composition of the inner bacterial membrane [2, 3, 4, 5, 6, 7, 8]. We use a multidisciplinary approach in which we combine experimental (physicochemical, biochemical, biophysical and structural) and theoretical tools to attain structural, dynamical and ensemble mechanical (single molecule versus collective behavior) information of the molecular events that control the reversible interactions leading to the formation of the complexes active in division.

References

[1] Vicente and Rico (2006) *Mol. Microbiol.* **61**: 5-8

[2] Rivas *et al.* (2000) *J. Biol. Chem.* **275**:11740-11749

[3] Rivas *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98**:3150-3155

[4] González *et al.* (2003) *J. Biol. Chem.* **278**:37664-37671

[5] González *et al.* (2005) *Proc. Natl. Acad. Sci. USA* **102**:1895-1900

[6] Mingorance *et al.* (2005) *J. Biol. Chem.* **280**:20909-20914

[7] Hölger *et al.* (2008) *Biophys J.* **94**(11):L81-3

[8] López-Montero *et al.* (2008) *Langmuir* **24**:4065-4076

Non coding RNAs infiltrate the cell: the micro RNA revolution

Pascal Barbry[1]

[1] CNRS and Université de Nice Sophia Antipolis, UMR 6097, Institut de Pharmacologie Moléculaire et Cellulaire, 660, route des lucioles, F-06560, Sophia Antipolis, France.

Abstract

In human, only ˜22,000 genes out of ˜69,000 correspond to protein-coding genes. Emphasis has been put recently on the large fraction (˜2/3) of non-protein-coding genes, and especially on those encoding micro RNAs (miRNAs). miRNAs have been discovered in 1993 by the groups of V. Ambros and G. Ruvkun as a class of small regulatory RNA controlling developmental timing in *C. elegans*. Since then, hundreds of distinct miRNAs have been discovered. They act at a post-transcriptional level and down-regulate mRNA expression. As many as 30% of human mRNAs may interact functionally with miRNAs. This implies that miRNAs potentially regulate many important cellular functions. Particular patterns of miRNAs expression are associated to specific stem cells, terminally differentiated cells, or cancer cells. This specifies a strong relationship between a given cell type and a small repertoire corresponding to the miRNAs showing the highest level of expression into a cell.

Recent data derived from high throughput sequencing suggest than less than 20 distinct miRNAs can probably cover more than 80% of the total miRNA available in a airway epithelial cell. The interactions between these highly expressed miRNAs and their targets therefore represent one minimal network of interactions specifying such a cell. Different levels of interaction have to be considered since: (1) several miRNAs can interact with the 3'-non-coding region of a same mRNA, but can also interact with many others mRNA targets; (2) one genomic locus can encode several miRNAs (e.g. cluster mir-17/mir-18a/mir-19a/mir-20a/mir-20b/mir-19b1/mir-92), which are likely to share similar profiles of expression. Intronic miRNAs can also act in synergy with other products from the same operational genetic unit. These different modes of regulation need to be correctly integrated in order to define the exact cellular impact of miRNAs.

I will illustrate how such relationships can play a role in the context of the respiratory epithelium. Airway epithelial cells (AECs) develop specific traits of differentiation (such as the development of cilia), in order to protect

efficiently the mucosa of the respiratory tract against external aggressions, and for providing specific responses against various stresses. AECs not only represent an excellent paradigm to study mechanisms of differentiation, but are also central to important health issues (asthma, cystic fibrosis, cancer,?). Our group has identified several miRNAs which expression is altered during differentiation, transformation and/or after infection of the airways. The use of several quantitative approaches derived from the microarray technologies (DNA microarray, high throughput sequencing) highlights how: (1) chromosomal rearrangements can affect the expression of some miRNAs; (2) production of miRNAs can be altered by genetic and epigenetic mechanisms; (3) the interplay existing between miRNAs and their mRNAs targets.

We anticipate from this project a definition of new specific markers of normal or pathological airway epithelial cells (basal, ciliated, goblet cells). The potential of some of these miRNAs in the context of wound healing, innate defense against virus infection and balance between proliferation and differentiation will be discussed.

Supports: CNRS, FP7 (Microenvimet n° 201279), InCA (PL0079), Infectiopole SUD, Région PACA, CG06, VLM.

References

[1] Le Brigand, K., Russell, R., Moreilhon, C., Rouillard, J. M., Jost, B., Amiot, F., Magnone, V., Bole-Feysot, C., Rostagno, P., Virolle, V., Defamie, V., Dessen, P., Williams, G., Lyons, P., Rios, G., Mari, B., Gulari, E., Kastner, P., Gidrol, X., Freeman, T. C., and Barbry, P. An open-access long oligonucleotide microarray resource for analysis of the human and mouse transcriptomes. *Nucleic Acids Res*, 34: e87, 2006.

[2] Le Brigand, K. and Barbry, P. Mediante: a web-based microarray data manager. *Bioinformatics*, 23: 1304-1306, 2007.

[3] riboulet R, Mari B, Lin YL, Chable-Bessia C, Bennasser Y, Lebrigand K, Cardinaud B, Maurin T, Barbry P, Baillat V, Reynes J, Corbeau P, Jeang KT, Benkirane M. Suppression of microRNA-silencing pathway by HIV-1 during virus replication. *Science*. 315:1579-1582, 2007.

[4] Saumet A, Vetter G, Bouttier M, Portales-Casamar E, Wasserman WW, Maurin T, Mari B, Barbry P, Vallar L, Friederich E, Arar K, Cassinat B, Chomienne C, Lecellier CH. Transcriptional repression of microRNA genes by PML-RARA increases expression of key cancer proteins in acute promyelocytic leukemia. *Blood*. 113:412-421, 2009.

[5] Ortholan C, Puissegur MP, Ilie M, Barbry P Mari B and Hofman P. MicroRNAs and lung cancer: new oncogenes or tumor suppressors, prognosis factors and potential therapeutic targets. *Current Medicinal Chemistry*. In press. 2009

Systems Biology Metabolic Modeling Assistant (SBMM): An ontology-based tool for the integration of metabolic data in kinetic modeling

Armando Reyes-Palomares, Raul Montañez, Alejando del Real-Chicharro, Othmane Chniber, Amine Kerzazi, Ismael Navas-Delgado, Miguel-Ángel Medina, José Aldana-Montes and Francisca Sánchez-Jiménez

Dpt. Molecular Biology and Biochemistry. F. Sciences. University of Málaga, Spain

Abstract

In systems biology it is becoming a routine task to build models of increasing complexity on a given biochemical network or pathway of interest. In this communication, we present SBMM Assistant [1], a tool built using an ontology-based mediator, and designed to facilitate metabolic modeling through the integration of data from repositories that contain valuable metabolic information. SBMM Assistant is an SBML-compatible [2] and user-friendly tool that gives the user the ability to capture, enrich, generate and visualize biological networks, to make basic queries about enzymatic kinetics and regulation, and it can be used as an assistant for kinetic modelling [3]. Furthermore, SBMM Assistant annotates this information using MIRIAM specifications [4] and facilitates friendly cross-talk among different resources and tools (for instance, COPASY: **`http://www.copasi.org`**). Assistant is freely available for academic use at **`http://www.sbmm.uma.es`**.

References

[1] Reyes-Palomares *et al.* (2009) Systems Biology Metabolic Modeling Assistant (SBMM): An ontology-based tool for the integration of metabolic data in kinetic modeling. *Bioinformatics* (Appl. notes), in press.

[2] Hucka,M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524-531.

[3] Rodríguez-Caso *et al.* (2006) Mathematical modeling of polyamine metabolism in mammals. *J Biol Chem.*, **281**:21799-812.

[4] Le Novère,N. *et al.* (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.*, **23**, 1509-1515.

Metabolic adjustments consequence of an energetic deficit and antibiotic production in *Streptomyces*

N. Seghezzi, A.Smirnov, C. Esnault, H. Chouayekh, S. Ghorbel and M.-J. Virolle.

Laboratoire de "Métabolisme Energétique des Streptomyces" (MES).
Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris XI, 91405 Orsay, FRANCE.

Abstract

There is an urgent need to found new drugs leads to face a major threat of this century, the development of multi-resistant pathogenic bacteria. The Actinomycetes genus that includes several hundred species is at the origin of the production of over two third of all known antibiotics. Each specie is known to produce 2 to 4 different antibiotics but the sequence of the genome of five species of this large genera (mainly *Streptomyces* species: *S. coelicolor, S. avermilis, S. scabies, S. griseus* and *S. ambofaciens*) revealed that each specie possesses 5 to 10 fold more biosynthetic pathways that could have been predicted from their biosynthetic abilities. This enormous genetic diversity is likely to result in the production of secondary metabolites with different chemical structures and bio-activities. A very important, but incompletely explored, and thus under-exploited reservoir of genetic and metabolic diversity resides in the hundred species of this genus.

It has long been known that, in *Streptomyces*, antibiotic production occurs in the period of slow or no growth (weak anabolism) and is triggered by a nutritional limitation in phosphate, condition that correlates with a weak energetic charge. Antibiotics are synthesised by huge enzymatic complexes that use precursors originating from primary metabolism. Sugars, nucleosides, amino acids, lipids etc. can be modified, assembled, combined, to generate the great chemical diversity of antibiotics (**`http://www.123bio.net/cours/antibio/index.html`**). However, many antibiotics fall in two major classes: the polyketides and the peptidyls. The polyketides are synthesized by polyketide synthetase (PKS) from precursors derived from acetyl CoA (butyryl CoA, malonyl CoA, methylmalonyl CoA etc.) that originate from the catabolism of sugars, fatty acids or amino acids. The peptidyls are synthesized by None Ribosomal Peptide Synthetase (NRPS) using modified or not amino acids. The genes responsible for the biosynthesis of a given antibiotic are grouped in cluster on the chromosome and their coordinated expression is under the control of activators (the "SARP" for "*Streptomyces* Antibiotic Regulatory Protein" of the OmpR family) or repressors linked to the biosynthetic pathways.

However, the cascade of regulation that goes from the sensing of a nutritional limitation in phosphate to the triggering of antibiotic production is still poorly understood. The expression of the putative biosynthetic pathways, detected *in silico*, is often weak and the resulting metabolites too poorly produced to be characterised. In consequence, a major challenge of the coming years is to find ways to enhance the expression of these "cryptic" biosynthetic pathways in order to exploit the outstanding genetic potential and thus metabolic richness of the several hundreds of Actinomycetes species present on earth. To do so, one needs to get a deeper understanding of what triggers the expression of the biosynthetic pathways and what is the nature of the regulatory changes that lead the cell to redirect its central metabolism, usually aimed at biomass construction, toward the production of secondary metabolites.

A novel understanding of the regulation of antibiotic biosynthesis was inferred from the analysis of a mutant of *Streptomyces lividans* (fig.1), a naturally very weak antibiotic producing strain, that became a very high producer of some coloured antibiotics, upon the interruption of a single gene, called *ppk* [1]. This gene was shown to encode an enzyme acting, in vitro, as a polyphosphate kinase (PPK), polymerising the ? phosphate of ATP into polyphosphate when the ATP/ADP ratio in the reaction was high and as a nucleoside diphosphate kinase (NDPK), regenerating ATP from ADP and polyphosphate, when this ratio was low [1]. In vivo, the expression of *ppk* was shown to be induced in condition of Pi limitation (condition correlating with a weak energetic charge) [2]. In this condition, the expression of *ppk* was shown to be positively controlled by the two components system PhoR/PhoP whereas in condition of phosphate sufficiency, the expression of *ppk* is thought to be negatively controlled by a repressor using ATP as co-repressor [3].

In order to assess the in vivo enzymatic function of Ppk, a comparative analysis of the intracellular content in ATP, ADP and polyphosphate of the wild type strain of *S. lividans* and of the *ppk* mutant was carried out, in condition of phosphate limitation. During active growth, the absence of this enzyme in the *ppk* mutant, was indeed shown to correlate with an higher intracellular concentration of ADP than in the wild type strain suggesting a default of ADP to ATP regeneration in that strain. During active growth, the internal concentration in ATP was the same in both strains but at stationary phase the concentration of ATP was, unexpectedly, much higher in the *ppk* mutant strain than in the wt strain. This high internal concentration of ATP in the *ppk* mutant is thought to result from a strong activation of central metabolism, that coupled to the respiratory chain, is the main ATP producing route within the cell.

The energetic deficit characteristic of the *ppk* mutant strain, that is lacking an important ATP regenerating enzyme, obviously provokes important changes

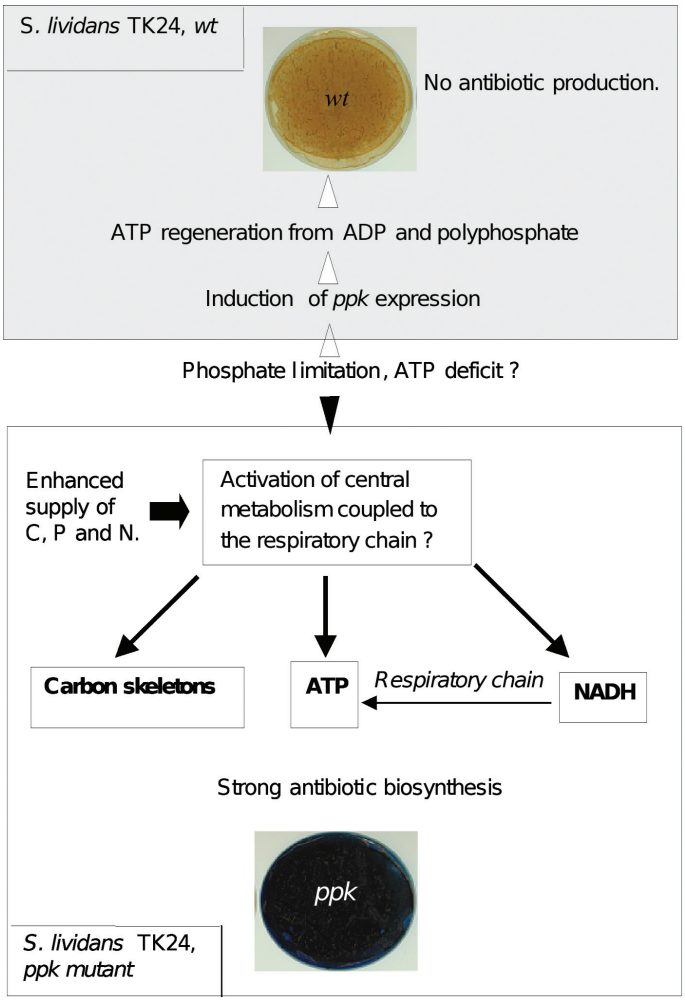


Figure 1: Novel model of the regulation of antibiotic biosynthesis in *Streptomyces lividans*

in the functioning of central metabolism. These changes still remain to be precisely described but some results of our proteomic studies suggested that a sequential activation of specific central metabolic pathways (glycolysis, Krebs cycle and β-oxydation of fatty acids) was taking place in the *ppk* mutant. This strong activation, likely aimed at maintaining the energetic balance of the *ppk* strain (ATP generation), concomitantly yields carbon skeletons that are constituting "building blocks" for antibiotics biosynthesis. As a matter of fact, in condition of active growth, the energetic and anabolic functions of the Krebs cycle are both required whereas when growth ceases (conditions

of antibiotics production) the anabolic function of the Krebs cycle is not required. However, in these conditions, the bacteria still has to produce ATP, in order to maintain its integrity and survive and some of our results suggest that the bacteria generates ATP *via* a strong activation of glycolysis that leads to the accumulation of acetylCoA stored as fatty acids in triacylglycerols. However, if the energetic deficit becomes very severe, as in the *ppk* mutant that is lacking a very important ATP regenerating enzyme, the degradation of these storage lipids would be triggered yielding glycerol, acetylCoA and numerous reduced co-factors (FADH2) whose regeneration by the respiratory chain would produce the necessary ATP. The glycerol will be metabolised via the Krebs cycle yielding carbon skeletons that, in this period of slow or no growth, will not be used for anabolism but will constitute precursors entering in the biosynthesis of antibiotics other than polyketides (peptidyl antibiotics for instance) whereas the generated acetylCoA will be used for the biosynthesis of polyketide antibiotics.

Our work demonstrated for the first time that an energetic deficit is the real trigger of antibiotic biosynthesis in *Streptomyces* and that this energetic deficit leads to specific adjustments of central metabolic pathways that results in the accumulation of precursors of antibiotic biosynthesis. The comparison of the metabolic fluxes determined in condition of energetic surplus or energetic deficit will lead to a better understanding of these specific adjustments and of the conditions required to promote good antibiotic production.

The MES group is funded by the French CNRS, the University Paris XI and the European program ACTINOGEN
(**`http://www.swan.ac.uk/research/ActinoGEN`**).

References

[1] Chouayekh H. and M.-J. Virolle. (2002) The polyphosphate kinase is involved in the phosphate control of actinorhodin production in *Streptomyces lividans*. *Mol. Microbiol.* **43**(4):919-30.

[2] Ghorbel S., Smirnov A., Chouayekh H., Sperandio B., Esnault C., Kormanec J.and M-J Virolle. (2006) Regulation of ppk Expression and In Vivo Function of Ppk in *Streptomyces lividans* TK24. *J. Bacteriol.* **188**(17):6269-76.

[3] Ghorbel S., Kormanec J., Artus A.and M-J Virolle (2006) Transcriptional studies and regulatory interactions between the *phoR-phoP* operon and the *phoU, mtpA,* and *ppk* genes of *Streptomyces lividans* TK24. *J. Bacteriol.* **188**(2):677-86.

[4] A.Smirnov, C. Esnault, H., A. Nowacka-Mazurek, C. Martel, E. Darbon-Rongère, H. Chouayekh, S. Ghorbel, N. Shikura, N. Seghezzi and M.-J. Virolle (2008) Toward a better understanding of why and how antibiotic biosynthesis is being triggered in *Streptomyces*? *Chemical Engineering Transactions.* **14**: 47-53. Proceedings of 1^{rst} International Conference on Industrial Biotechnology IBIC 2008.
(**http://www.aidic.it/IBIC2008**)

[5] Virolle MJ (2007) Antibiotic Production in Actinomycetes: Towards a Better Understanding. *BioforumEurope* Oct:37-39.
(**http://www.eMagazineBIOforum.com**)

Ordinary Differential Equations and Modelling of Microbial Ecosystems

Claude Lobry[1]

[1] INRIA-MERE and Université Nice Sophia-Antipolis

Abstract

A bioreactor used for the purpose of wastewater depollution contains thousands of different species with populations ranging from 10^9 to 10^{12} individuals. A possible model for the description of the dynamics of such an ecosystem is :

$$\begin{cases} \dfrac{ds}{dt} &= d(\varepsilon t)(S_{in}-s) - \sum_{i=1}^{N} \mu_i(s)x_i \\ \\ \dfrac{dx_i}{dt} &= (\mu_i(s) - d(\varepsilon t))x_i + \sigma\{\sum_{j=1}^{N} \alpha_{ij}\mu_j(s)x_j - \mu_i(s)x_i\} \quad i = 1, 2, ..., N \end{cases}$$

where :

- $x_i(t)$ is the concentration of the species i at instant t,
- $s(t)$ is the concentration of the nutriment at instant t,
- $\mu_i(s)$ is a function with "Monod" shape which represents the growth rate of the species i,
- α_{ij} is the mutation rate from species j to species i,
- d is the flux through the reactor.

If one integrates such a system, a typical picture obtained is shown on Fig. 1. We do not comment this picture in this short abstract; actually the main objective of the course is to explain such pictures. For that purpose we shall:

- Recall the well known theory of the chemostat (Spicer 1950)
- The competitive exclusion principle (Hardin 1960)
- Experiments of Hansel and Hubbell (1980)
- Evidences of coexistence of many species in bioreactors using "finger prints"
- Possible explanation through "Ratio Dependence" (Arditi Ginzburg 1990)
- Variable environments
- "Kill the winner hypothesis".

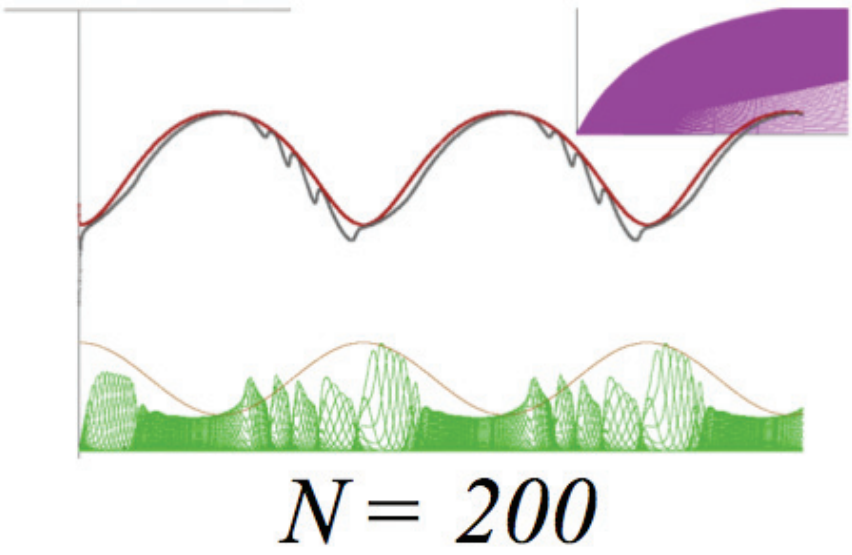


Figure 1: $N = 200 \quad \sigma = 0.000001 \quad \varepsilon = 0.001$

References

[1] J. Harmand, A. Rapaport, D. Dochain, C. Lobry; *Microbial ecology and bioprocess control : opportunities and challenges,* in "International Journal of Robust and Nonlinear Control", vol 18, n^o 18, 2008 p. 1637-1656.

[2] A. Rapaport, D. Dochain, J. Harmand *Practical coexistence in the chemostat with arbitrarily close growth functions,* in ARIMA, vol 9, 2008, p. 231-243

[3] A. Rapaport, D. Dochain, J. Harmand *Long run coexistence in the chemostat with multiple species* in Journal of Theoretical Biology, to appear 2008.

[4] C. Lobry, F. Mazenc, A. Rapaport. *Persistence of ecological models of competition for a single resouce,* C.R. Acad. Sci. Paris, Mathématiques, 1340 (2005) p. 199-204.

Circuits, differentiation and homeostasis in genetic regulatory networks

Paul Ruet[1]

[1] CNRS, Institut de Mathématiques de Luminy, Campus de Luminy, Case 907, 13288 Marseille Cedex 9, France. Email: ruet@iml.univ-mrs.fr.

Abstract

The purpose of my lecture is to present some recent mathematical works on the relationships between the structure of genetic regulatory networks (typically the presence of positive or negative circuits) and the emergence of significant dynamical properties such as multistability (involved in differentiation) and periodic oscillations (involved in homeostasis).

Introduction

It is acknowledged that the activity of a biological cell is to a large extent controlled by genetic regulation, an interactive process usually represented by graphs called genetic regulatory networks. In these graphs, vertices represent genes, and a directed edge from i to j represents a regulatory influence of gene i on gene j. Such an influence may be an activation (in which case the edge is said positive) or an inhibition (negative edge).

Since this interactive process is in general very complex, it is interesting to relate the dynamical properties of these networks to their (simpler) structure. In particular, in the early 1980's, the biologist R. Thomas, following earlier work by Delbrück, suggested that the existence of a positive circuit in the regulatory network (the sign of a circuit being defined as the product of the signs of its edges) is a necessary condition for multistability (i. e., the existence of several stable fixed points in the dynamics), and that a negative circuit is necessary for the presence of sustained oscillations.

The purpose of my lecture is to present some recent mathematical works inspired by these two rules, or somehow related to them.

References

[1] J. Aracena, J. Demongeot, and E. Goles, (2004) On limit cycles of monotone functions with symmetric connection graph. *Theoretical Computer Science*, **322(2)**:237-244.

[2] A. Crumière and P. Ruet, (2008) Spatial differentiation and positive circuits in a discrete framework. *Developments in Computational Models*, ENTCS **192**: 85-100.

[3] J.-L. Gouzé, (1998) Positive and negative circuits in dynamical systems. *Journal of Biological Systems*, **6**:11-15.

[4] D. Gale and H. Nikaido, (1965) The jacobian matrix and global univalence of mappings. *Mathematische Annalen*, **159**:81-93.

[5] E. Plahte, T. Mestl, and S. W. Omholt, (1995) Feedback loops, stability and multistationarity in dynamical systems. *Journal Biological Systems*, **3**:409-413.

[6] E. Remy, B. Mossé, C. Chaouiya, and D. Thieffry, (2003) A description of dynamical graphs associated to elementary regulatory circuits. *Bioinformatics*, **19(2)**:172-178.

[7] E. Remy, P. Ruet, and D. Thieffry, (2008) Graphic requirements for multistability and attractive cycles in a Boolean dynamical framework. Preprint IML-2005-8. *Advances in Applied Mathematics*, **41(3)**: 335-350.

[8] E. Remy and P. Ruet, (2007) On differentiation and homeostatic behaviours of boolean dynamical systems. In *Trans. on Comput. Systems Biology VIII*, Springer LNCS **4780**: 92-101.

[9] E. Remy, P. Ruet, and D. Thieffry, (2006) Positive or negative regulatory circuit inference from multilevel dynamics. In *Positive Systems: Theory and Appl.*, Springer LNCIS 341: 263-270.

[10] E. Remy and P. Ruet, (2008) From minimal signed circuits to the dynamics of Boolean regulatory networks. *Bioinformatics*, **24**: i220-i226.

[11] A. Richard and J.-P. Comet, (2007) Necessary conditions for multistationnarity in discrete dynamical systems. *Discrete Applied Mathematics*, **155(18)**: 2403-2413.

[12] A. Richard, (2007) On the link between oscillations and negative circuits in discrete genetic regulatory networks. In *Journées Ouvertes en Biologie, Informatique and Mathématiques*.

[13] F. Robert, (1986) *Discrete iterations: a metric study*, volume **6** of *Series in Comput. Mathematics*. Springer.

[14] M.-H. Shih and J.-L. Dong, (2005) A combinatorial analogue of the Jacobian problem in automata networks. *Advances in Applied Mathematics*, **34(1)**:30-46.

[15] H. Siebert, (2008) Deriving behavior of Boolean bioregulatory networks from subnetwork dynamics. Preprint Freie Universität Berlin.

[16] E. H. Snoussi, (1998) Necessary conditions for multistationarity and stable periodicity. *Journal of Biological Systems*, **6**:3-9.

[17] C. Soulé, (2003) Graphic requirements for multistationarity. *ComPlexUs*, **1**:123-133.

[18] C. Soulé, (2006) Mathematical approaches to gene regulation and differentiation. *Comptes Rendus de l'Académie des Sciences, Biologies*, **329(1)**:13-20.

[19] R. Thomas, (1981) On the relation between the logical structure of systems and their ability to generate multiple steady states and sustained oscillations. In *Series in Synergetics*, volume **9**, pages 180-193. Springer.

From the Birth of the Cell to Molecular Movies

Christian Sardet[1]

[1] BioMarCell group, Biodev UMR7009 CNRS / UPMC, Observatoire Océanologique de Villefranche sur Mer, Université Pierre et Marie Curie - Paris VI, France

Abstract

It took 200 years for the cell theory to emerge. Robert Hook, a british polymath considered one of the father of microscopy coined the term "cell" in 1665 to describe the small pores he observed in a sliver of cork. After contributions by a multitude of Europeans including Leewenhoek, Dutrochet, Raspail, Oken, Fontana, Purkinje, in 1847, Schleiden and Schwann declared that "All living things are composed of cells and cell products" and stressed the central role of the nucleus - the Cell theory was born [1, 2]. Discovery of chromosomes, organelles, macromolecules followed and the cell biology has now federated and united many fields from zoology and genetics, to molecular and developmental biology.

The first moving images of cell a dividing sea urchin egg - appeared in 1909 [3]. Films of cells, embryos and tissues have multiplied in the last 20 years since the introduction of video and digital imaging techniques coupled with the development of fluorescent or luminescent probes [4, 5]. Some of these documents are beginning to reach TV, feature movies and video sites such as YouTube and impact the general public [6, 7] . Video journals [8], sites and even festivals [9] devoted to movies of cells and macromolecular structure are now available for students of the cell.

Most interestingly, computer animations can generate molecular movies which propose possible representations of living "nanomachines" in action [10]. No doubt advances in 3D structures of molecules [11], high resolution cell imaging and computer animation will lead to modelling the cell and it's components.

The time is ripe for collectively building a "Virtual Cell".

References

[1] "The Birth of the Cell", Henry Harris, Yale University Press, New Haven (1999)

[2] ”The Discovery of the Cell” video document by D Deliekiein the DVD ”Exploring the Living Cell” by C Sardet, V Kleiner, CNRS Image production (2006)
http://videotheque.cnrs.fr/index.php?urlaction=doc&id_doc=1763&rang=6

[3] See BioClip

[4] ”Imaging Living Cells (Springer lab manual)”, R Rizzuto, C Fasolato, Springer (2008)

[5] ”Live Cell Imaging”, R.D. Goldman, David L. Spector, CSHL (2004)

[6] See sample of ”Voyage inside the Cell” by C Sardet with Digital Studio (1995) on **http://fr.youtube.com/watch?v=GW0lqf4Fqpg**
The full DVD version is distributed by Sinauer:
http://www.sinauer.com/voyage/video.php

[7] See ”Inner Life of Cells” by R Lue and A Viel with EXVIVO, Harvard BioVision. **http://multimedia.mcb.harvard.edu**

[8] Journal of Vizualized Experiments: **http://www.jove.com**

[9] Multimedia documents and ”Cinema of the Cell” festival:
http://www.bioclips.com

[10] Molecular movies: a portal for cell and molecular animation:
http://www.molecularmovies.com

[11] Molecule of the month by D. Goodsell:
http://www.rcsb.org/pdb/home/home.do

Experiment-based mathematical modelling of energy metabolism in diseases caused by unfolded/misfolded proteins

Judit Ovádi[1]

[1] Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, H-1113 Budapest, Hungary

Abstract

Research in the near past has revealed that a major part of neurodegeneration is a multistep process during which unfolded/misfolded proteins initiate a cascade of aberrant protein-protein interactions resulting in inclusions with toxic aggregates that culminate in neuronal dysfunction. These complex pathological processes are potent targets for development of early diagnosis and of drugs to improve therapies of these diseases. The hallmark proteins of these conformational diseases such as Parkinson's, Alzheimer's or Huntington's diseases, are α-synuclein, tau or mutant huntington, respectively, do not have well-defined 3D structures and require protein partners to express their functions. Recently we isolated and identified a new unstructured protein denoted Tubulin Polymerization Promoting Protein, TPPP/p25 which was found to be enriched in human pathological inclusions characteristic for synucleinopathies (fig1).

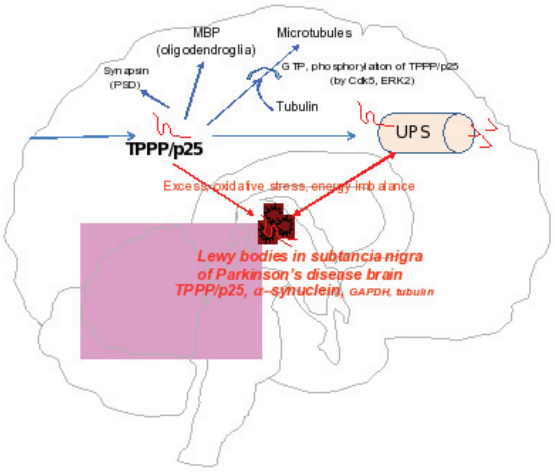


Figure 1: Tubulin Polymerization Promoting Protein, TPPP/p25

This unfolded protein has become the target of many structural, functional and pathological studies due to its potential involvement in different CNS

diseases. TPPP/p25 was discovered as the first member of a new protein family which primarily targets microtubule structures displaying dominant MAP-like function by affecting the dynamics and stability of microtubule networks.

The formation of diverse protein aggregates frequently induces mitochondrial dysfunction and failure of synaptic, transport and other crucial physiological processes, which contributes to the development of pathological symptoms. Genomic and proteomic data provided useful information for understanding the structural and functional perturbations in the cases of diverse neurodegenerative diseases frequently coupled with mitochondrial dysfunction. However, our knowledge whether and how the presence of the unfolded/misfolded proteins affects the energy metabolism in the pathogenesis of the conformational diseases is scarce.

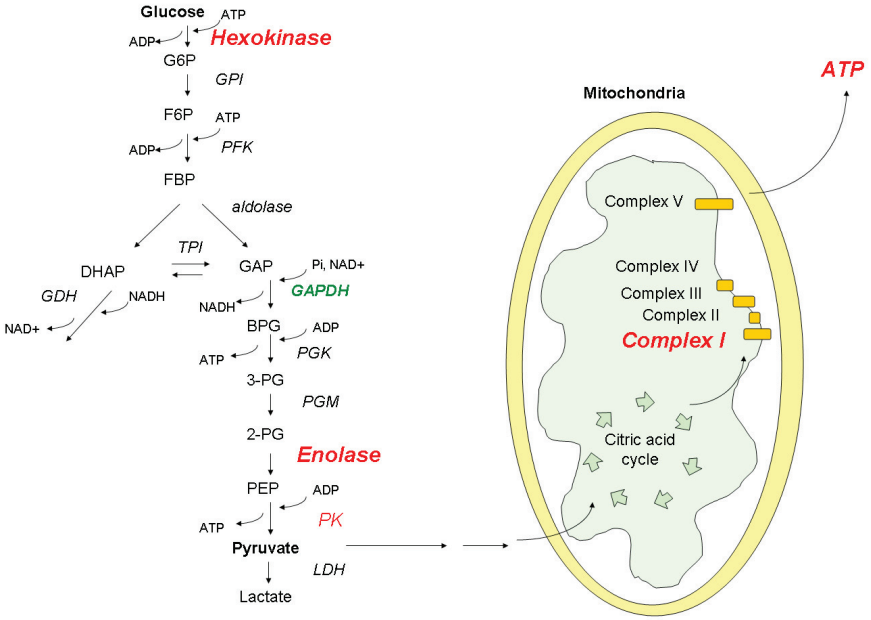


Figure 2: metabolization of glucose via glycolysis.

This gappy area has been an object of our recent research applying of biosimulation based upon well-established rate equations of glycolytic enzymes and the mitochondrial complexes and experimentally determined kinetic parameters. These studies were performed, on one hand, on blood cells of human models suffering from a rare glycolytic enzymopathy, caused by the mutation of triosephosphate isomerase, and coupled with neurological disorder; on the other hand, on brain tissue of mouse model for Huntington's disease caused

by the mutation-derived insertion of an unfolded polyglutamine track in the huntingtin protein. In addition, a specific human cell model, SK-N-MC cell expressing TPPP/p25 protein stably was established which rendered it possible to characterize the effect of the unfolded protein on the mitochondrial membrane polarization related to the energy metabolism at single cell level, as well as on the glucose metabolism at system level.

Glucose is the major or the only energy source of the brain tissue and the red blood cells. It is metabolized primarily via glycolysis (fig 2). To evaluate the effect of the expression of the unfolded/misfolded proteins on the energy metabolism, the activities of the glycolytic enzymes were measured in the controls and the ”diseased” cells, and, on one hand, compared with that of the isolated enzymes, on the other hand, used to compute the glycolytic fluxes using the well-established rate equations of the individual glycolytic enzymes.

These data gave information on the local changes, on the influence of intracellular heterologous interactions and for their effects on the glycolytic fluxes. Then the glycolytic fluxes were also measured in the control and ”diseased” cells and compared with the computed ones. These studies rendered it possible to evaluate the mechanisms responsible for the altered glycolytic fluxes such as enzyme activation/inhibition or metabolite channelling of key metabolites. Various mechanisms were identified in the three systems studied which were responsible for the altered metabolic fluxes, i.e., association of the mutant misfolded isomerase to the red cell membrane, or the microcompartmentation of glyceraldehyde-3-phosphate intermediate by the simultaneous association of the glyceraldehyde-3-phosphate dehydrogenase and aldolase to the mutant huntingtin protein resulting in enzyme inhibition. Therefore, the local changes were found to be compensated at system level either by activation of other glycolytic enzymes or the direct transfer of the intermediate of the aldolase-dehydogenase-catalysed consecutive reactions. No metabolic impairments were observed in any cases studied rather enhanced ATP production was detected when either mutant isomerase or mutant huntingtin or unfolded TPPP/p25 proteins were accommodated in the ”diseased” cells. In addition, immunofluorescence images using reporter molecule in TPPP/p25-transfected cells showed that the mitochondrial membrane was highly polarized which concerns with the kinetic data obtained with brain tissue of the transgenic mice: there was no impairment in activities of the mitochondrial complexes.

Therefore, our results suggest that unfolded/misfolded proteins cause enhancement of the intracellular ATP level which associated primarily with increased glycolytic flux leading to altered metabolite pattern with pathological consequences manifesting themselves in distinct clinical symptoms.

References

[1] Orosz, F., Wágner, G., Liliom, K., Kovács, J. Baróti, K., Horányi, M., Farkas, T., Hollán, S. and Ovádi, J. Distinct associations of normal and mutant triosephosphate isomerases to red cell membrane and brain microtubules. Molecular bases of enzyme deficiency *Proc. Natl. Acad. Sci. USA* (2000) **97**, 1026-1031

[2] Ovádi, J., Orosz, F. Hollán, Zs: Functional aspects of cellular microcompartmentation in the development of neurodeneration. Mutation induced aberrant protein-protein associations *Mol. Cell. Biochem.* (2003) **256**, 83-93

[3] Orosz, F., Kovács, G. G., Lehotzky, A., Oláh, J., Vincze, O., Ovádi, J. TPPP/p25: from unfolded protein to misfolding disease: prediction and experiments. *Biol. Cell.* (2004) **96**, 701-11.

[4] Oláh, J., Orosz, F., Puskás, L.G., Hackler, L., Horányi, M., Polgár, L., Hollán, S., Ovádi, J. Triosephosphate isomerase deficiency: consequences of an inherited mutation at mRNA, protein and metabolic levels. *Biochem J.* (2005) **392** (3), 675-683

[5] Ovádi, J., Orosz, F.: Energy metabolism in conformational diseases (2007) In: *Biosimulation in drug development*. (Eds.: Bertau M., Mosekilde, E., Westerhoff, H.) Wiley, Wilhelm 233-257

[6] Oláh J, Klivényi P, Gardián G, Vécsei L, Orosz F, Kovacs GG, Westerhoff HV, Ovádi J. Increased glucose metabolism and ATP level in brain tissue of Huntington's disease transgenic mice. *FEBS J.* (2008) **275**, 4740-4755.

[7] Orosz, F., Lehotzky, A., Oláh, J., Ovádi, J. TPPP/p25: A New Unstructured Protein Hallmarking Synucleinopathies (eds J. Ovádi and F. Orosz) *Protein Folding and Misfolding: Neurodegenerative Diseases*, Springer 2009

Virus Dynamics in Mammalian Cell Culture

Udo Reichl[1,2]

[1] Max Planck Institute for Dynamics of Complex Technical Systems, Bioprocess Engineering, Sandtorstrasse 1, 39106 Magdeburg, Germany
[2] Otto-von-Guericke-University Magdeburg, Chair of Bioprocess Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany

Abstract

In the next decades mammalian cell culture-derived vaccines will play an increasing role in prevention and control of seasonal disease but also of pandemic outbreaks. To guarantee an affordable supply of safe and highly immunogenic vaccines significant efforts in design and optimization of process technologies in vaccine manufacturing are required. Particularly, virus yields of large-scale bioreactor harvests are of interest, which mainly depend on host cell line, cell concentration at time of infection, virus seed and cultivation conditions. For process monitoring, comprehensive analytical methods for characterization of cell growth and virus replication are required. In addition, quantitative analyses by mathematical modelling approaches are of crucial importance for interpretation of experimental data, establishment of control strategies, scale-up studies and improvement of overall productivity.

Main goal of our group is the development of integrated concepts to optimize viral-based production processes [1]. As an example we investigate influenza virus replication in various adherent and suspension cell lines in stirred tank and wave bioreactors. At present, we focus on optimization of cell growth and virus yield, design of high cell density cultivations, and detailed characterization of cellular metabolism and intracellular virus replication kinetics. On-line data from bioreactors are monitored and extra and intracellular metabolites quantified with numerous newly established assays. This includes off-line analysis of glucose, lactate, ammonia and amino acids as well as phosphorylated sugars, organic acids, ATP and NAD(P)H, all relevant in glycolysis and TCA cycle. Influenza virus yield is determined by hemagglutination and infectivity assays. From flow cytometry measurements additional information concerning cell cycle status, progress of infection and virus-induced cell death by apoptosis are gained.

Mathematical modelling plays a crucial role with regard to a quantitative understanding of cell growth and virus replication for this process [2]. Based on extensive sets of experimental data, models of different complexity are being developed and validated. These include basic dynamic models to describe

cell growth in bioreactors and models for metabolic flux analysis. Furthermore, virus replication dynamics is described at various levels of complexity.

Focus of this presentation will be on influenza A virus replication in adherent Madin Darby canine kidney (MDCK) cells covering:

- basic virus dynamics in bioreactors (unstructured, segregated models) [3],
- single cell infection models (structured models) [4], and
- virus spreading in populations of cells (structured, segregated models) [5, 7].

Use of such mathematical models not only allows the simulation of the overall dynamics of virus replication in bioreactors but also supports the analysis of limiting factors for increasing process yields. Furthermore, models facilitate interpretation of an ever increasing amount of experimental data available at all process levels. In particularly, they encourage systems biology studies regarding virus replication at a cellular level to obtain a more detailed knowledge of the basic laws and mechanisms that control virus replication in mammalian cells. This is a prerequisite not only for the optimization of virus-related production processes but also essential for a better understanding of viral diseases and the identification of molecular targets for antiviral therapies.

References

[1] Genzel, Y., Reichl, U. (2007): Vaccine production - state of the art and future needs in upstream processing, In: *Animal Cell Biotechnology - Methods and Protocols*, Pörtner, R., Humana Press Inc., Totowa, New Jersey, 457-473.

[2] Reichl, U., Sidorenko, Y. (2006): Virus-host cell interaction, In: *Bioinformatics: From Genomes to Therapies*, Lengauer, T., Vol. II, Chapter 23, 861-898, Wiley-VCH.

[3] Möhler, L., Flockerzi, D., Sann, H., Reichl, U. (2005): A Mathematical Model of Influenza A Virus Production in Large-Scale Microcarrier Culture, *Biotechnology and Bioengineering*, **90**(1), 46-58.

[4] Sidorenko, Y. and Reichl, U. (2004): Structured model of influenza virus replication in MDCK cells, *Biotechnology and Bioengineering*, **88**(1), 1-14.

[5] Müller, T., Schulze-Horsel, J., Sidorenko, Y, Reichl, U., Kienle, A. (2008): Population balance modelling of influenza virus replication in MDCK cells during vaccine production, *18th European Symposium on Computer Aided Process Engineering - ESCAPE 18, in: Computer-Aided Chemical Engineering 25*, edts.: Bertrand Braunschweig (IFP, Frankreich) and Xavier Joulia (LGC- ENSIACET - IPT, Frankreich), 33-138.

[6] Sidorenko, Y., Schulze-Horsel, J., Voigt, A., Reichl, U., Kienle, A. (2008): Monte Carlo simulations of influenza virus replication in vaccine production processes, *Chemical Engineering Sciences*, **63**(1), 157-169.

[7] Schulze-Horsel, J., Schulze, M., Agalaridis, G., Genzel, Y., Reichl, U. (2008): Infection dynamics and virus-induced apoptosis in cell culture-based influenza vaccine production - flow cytometry and mathematical modeling, *Vaccine*, submitted.

Oscillations, variability and frequency tuning in the p53-Mdm2 network

M. Kaufman[1], W. Abou-Jaoudé[1] and D.A. Ouattara[1]

[1] Université Libre de Bruxelles (U.L.B.), Faculté des Sciences,
Unit of Theoretical and Computational Biology

Abstract

We investigate the dynamical properties of a simple four-variable model describing the interactions between the tumour suppressor protein p53, its main negative regulator Mdm2 and DNA damage, a model inspired by the work of Ciliberto et al. [1]. Its core consists of an antagonist circuit between p53 and nuclear Mdm2 embedded in a 3-element negative circuit involving p53, cytoplasmic and nuclear Mdm2. Rather than choosing a unique mode of description, we develop an integrated approach combining a multilevel logical method with a differential approach and stochastic simulations. We show that the essential dynamical properties of our network are described by a small number of bifurcation scenarios that can be interpreted in terms of the balance between the positive and negative loops of the core of the network. These bifurcation scenarios depend on two parameters linked to post-translational modifications of p53, the DNA-binding affinity and transcriptional activity of p53. Since these parameters are known to be cell- and stress-type specific, we propose that different types of cells or stresses could be characterized by different bifurcation schemes and lead to different responses upon irradiation. Our results also account for important features of the kinetics of the p53 response to damage that, to our knowledge, have not been addressed in other modeling approaches. In particular, we provide an interpretation of the tuning of the oscillation frequency that has been observed experimentally depending on the irradiation dose [2], and predict that the rate of damage repair should play an important role for this behaviour [3].

References

[1] Ciliberto, A., Novak, B., and Tyson, J. J., "Steady states and oscillations in the p53/Mdm2 network". Cell Cycle 4 (3), pp. 488-493, 2005.

[2] Geva-Zatorsky, N., Rosenfeld, N., Itzkovitz, S., Milo, R., Sigal, A., Dekel, E., et al., "Oscillations and variability in the p53 system". Mol. Syst. Biol., 2, 2006.

[3] Abou-Jaoudé, W., Ouattara, D.A. and Kaufman, M., ”From structure to dynamics: Frequency tuning in the p53-Mdm2 network. I. Logical approach”. J. Theor. Biol. (in press).

A system's biology approach to understand stochasticity in gene expression

Olivier Gandrillon[1]

[1] Centre de Génétique Moléculaire et Cellulaire, CNRS UMR 5534, Université de Lyon 1, France

Abstract

There is an increasing body of evidence showing the importance of stochasticity in gene expression in various biological phenomena [1, 12, 5, 3] ; for recent review see [7, 11, 9], including differentiation [13, 6, 8, 10, 2].

We are aiming at understanding the molecular causes of this stochasticity in gene expression in higher eukaryotic cells through a system's biology approach, combining modeling and experimental evidences. Our modeling focuses on the promoter level, and integrates the complex interplay of the dynamics of transcription factors and their combinatorial mutual influences. In parallel we have started acquiring expression level in single cells in real time, using a model of primary chicken erythrocytic progenitors [4]. Although the completion of the virtuous circle (modeling >experiments >modeling) still lies ahead of us, preliminary evidences indicate that stochasticity is a complex phenomenon that can only be fully understood by a systemic approach.

References

[1] Benzer, S. (1953) Induced synthesis of enzymes in bacteria analyzed at the cellular level. *Biochim Biophys Acta*, **11**, 383-395.¡BR¿

[2] Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E. and Huang, S. (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, **453**, 544-547.¡BR¿

[3] Elowitz, M.B., Levine, A.J., Siggia, E.D. and Swain, P.S. (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1183-1186.¡BR¿

[4] Gandrillon, O., Schmidt, U., Beug, H. and Samarut, J. (1999) TGF-beta cooperates with TGF-alpha to induce the self-renewal of normal erythrocytic progenitors: evidence for an autocrine mechanism. *Embo J*, **18**, 2764-2781.¡BR¿

[5] Golding, I., Paulsson, J., Zawilski, S.M. and Cox, E.C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell*, **123**, 1025-1036.¡BR¿

[6] Hume, D.A. (2000) Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. *Blood*, **96**, 2323-2328.¡BR¿

[7] Kaern, M., Elston, T.C., Blake, W.J. and Collins, J.J. (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet*, **6**, 451-464.¡BR¿

[8] Kupiec, J.J. (1997) A Darwinian theory for the origin of cellular differentiation. *Mol Gen Genet*, **255**, 201-208.¡BR¿

[9] Maheshri, N. and O'Shea, E.K. (2007) Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu Rev Biophys Biomol Struct*, **36**, 413-434.¡BR¿

[10] Paldi, A. (2003) Stochastic gene expression during cell differentiation: order from disorder? *Cell Mol Life Sci*, **60**, 1775-1778.¡BR¿

[11] Raser, J.M. and O'Shea, E.K. (2005) Noise in gene expression: origins, consequences, and control. *Science*, **309**, 2010-2013.¡BR¿

[12] Ross, I.L., Browne, C.M. and Hume, D.A. (1994) Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol Cell Biol*, **72**, 177-185.¡BR¿

[13] Wernet, M.F., Mazzoni, E.O., Celik, A., Duncan, D.M., Duncan, I. and Desplan, C. (2006) Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature*, **440**, 174-180.¡BR¿

Multiscale modelling of structured cell populations: application to ovulation control

Frédérique Clément[1]

[1] INRIA Paris-Rocquencourt, F-78753 Le Chesnay, France

Abstract

We will present an instance of multiscale modelling of coupled structured populations and its application to the selection process of ovarian follicles for ovulation. The follicle selection process is viewed as an FSH (follicle stimulating hormone)-dependent controlled process. FSH acts on the molecular scale through a signal transduction pathway. Integration of FSH signaling on the cellular scale rules the dynamics of the transition rates between different cellular states, hence cell commitment towards either proliferation, differentiation or apoptosis. Further scaling determines the follicular fate based on its cell content and contribution to ovarian endocrine status. Summing-up each follicular contribution finally defines the ovarian feedback pressure on FSH release, closing the loop. Our approach thus merges a detailed multi-scaled description of follicular development with population dynamics issues, due to competition for FSH resource.

1 Introduction

In vertebrates, the gonadotrope axis is made up of the hypothalamus, belonging to the central nervous system, the pituitary gland and the gonads. These organs communicate with one another within entangled endocrine loops. We will focus here on the gonadic level and present an instance of multiscale modelling for coupled structured cell populations and its application to the selection process of ovarian follicles for ovulation.

1.1 Physiological background: follicular development

The ovulation process is the endpoint of follicular development, the process of growth and functional maturation undergone by ovarian follicles, from the time they leave the pool of primordial follicles until ovulatory stage. Its biological meaning is to release one (in mono-ovulating species) or several (in poly-ovulating species) fertilisable oocyte(s) enabled to subsequent embryo development. Actually, very few follicles reach an ovulatory size; most of them undergo a degeneration process, known as atresia. The species-specific ovulation rate (number of ovulatory follicles) results from an FSH-dependent follicle selection process.

1.2 Biological and medical challenges

The development of ovarian follicles is a crucial, limiting step for the success of reproduction in mammals. Yet, the process of follicle selection, the regulation of the species-specific ovulation rate and the meaning of the tremendous wasting of follicles through atresia are still incompletely understood. Resolving these basic scientific questions correspond to both clinical and zootechnical challenges. A better understanding of follicular development is required to improve the control of anovulatory infertility in women and to control ovulation rate and ovarian cycle chronology in domestic species.
Beyond the frame of reproduction physiology, follicular development is a unique instance of rapid and controlled development in adult organisms, and follicular cells (particularly the granulosa cells surrounding the oocyte) constitute an interesting model in cell kinetics studies. Also, the FSH receptor belongs to the seven transmembrane spanning receptor family, which represents the most frequent target (over 50%) amongst the therapeutic agents currently available.

2 Middle-out modelling approach

Our working hypothesis is that understanding how FSH controls the outcome of follicular development (ovulation or not) amounts to understand how FSH acts on its target cells within ovarian follicles (the granulosa cells) and controls their commitment towards either proliferation, differentiation or apoptosis. The changes in the cellular composition of the granulosa ultimately determine the follicular fate. The ovulatory trajectories correspond to a FSH-controlled transition from a proliferative state to a differentiated one, while the atretic trajectories correspond to either the proliferation-apoptosis sequence or the proliferation-differentiation-apoptosis sequence.
Accordingly, the dynamics of the granulosa cell population was first expressed as a function of follicular age, by a system of ordinary differential equations associated to a compartmental model, where the compartments correspond to each cellular state (Proliferative-Differentiated-Apoptotic) [3]. The transition rates between compartments act as control variables mediating the interactions of the follicle with its hormonal environment, while the model outputs consist in the cell numbers in each compartment. Such a modelling framework raised the notion of follicular proliferative resources, that can be investigated on the experimental ground by cell kinetics experiments (measurement of the growth fraction [4]). This notion leads to question whether there exists an optimal strategy of exploiting these resources to reach ovulation [5]. Both the experimental and modelling studies have highlighted the crucial role of the dynamics of the cellular transition rates in the outcome of follicular development.

Hence, to further understand the control of the granulosa dynamics, we needed to investigate the action of gonadotrophins on the cellular transition rates. Since the response to FSH depends on the cell maturity level (granulosa cells from small follicles tend to proliferate, while those from big follicles tend to differentiate), we had to change for a less simple formalism than the ODE-compartmental one, that considered homogenous cell populations. The first step in designing another formalism was to characterise the biochemical basis of FSH action on the intracellular level, so that we shifted to a top-down modelling phase.

3 Top-down modelling approach

FSH is a pituitary glycoprotein hormone that specifically acts on granulosa cells of the ovary in female. FSH exerts its biological functions via a 7 transmembrane spanning receptor (7TMR), namely the FSH receptor (FSH-R), which preferentially couples to the Gα_s subunit. So far, the FSH-induced downstream signalling pathways were thought to be activated solely by the cAMP (cyclic Adenosine Mono-Phosphate)-PKA (Protein Kinase A) pathway [6]. We have thus focused up to now on this pathway to characterise the sensitivity of granulosa cells to FSH, hence their maturity.

The control of cAMP levels in granulosa cells involves both fast biochemical processes, occurring on a time scale of a few minutes, such as binding and desensitisation, and slower physiological processes lasting hours or even a few days, which mainly result in changing the efficiency of the enhancement of cAMP synthesis by stimulated FSH receptors via adenylyl cyclase activation. The design of our model followed from the interactions between these contrasting biochemical and physiological dynamics and aimed at accounting for both the FSH-dependent and auto-amplified settling and efficiency of the cAMP cascade through follicular development. The use of quasi-steady state approximations allowed us to reduce the biochemical system to a single differential equation [7].

Recently, the paradigm for 7TMR signalling has been renewed [8]. At least three protein families have been involved in signal transduction: not only Gα_s, but also GRKs (G protein-coupled Receptor Kinases) and β-arrestins. We are currently investigating the relative contributions of these different pathways within FSH signalling network, from a systems biology viewpoint [9]. The qualitative and quantitative properties of this network will be analysed and summarised to enrich the intracellular level of the model.

4 Bottom-up modelling approach

The bottom-up embedding of FSH signal transduction dynamics into the granulosa cell population dynamics results in multiscale structured conservation laws in each follicle. Coupling of the follicular densities through a common control term further allows to consider the upper, ovarian (tissular) scale [10]. For each follicle, the cell population dynamics is ruled by a conservation law with variable coefficients, which describes the changes in age and maturity of the granulosa cell density. A coupling control term, representing FSH signal, intervenes both in the velocity and loss terms of the conservation law. Two acting controls are distinguished: a global control resulting from the ovarian feedback and corresponding to FSH plasmatic levels, and a local control, specific to each follicle, accounting for the modulation in FSH bioavailability due to follicular vascularisation. Besides, cells are characterised by their position within or outside the cell cycle and their sensitivity to FSH. This leads to distinguish 3 cellular phases within the granulosa cell population. Phases 1 and 2 correspond to the proliferation phases (describing respectively the G1 phase and S to M phases of the cell cycle), and phase 3 corresponds to the differentiation phase, after cells have exited the cell cycle.
The cell population within an ovarian follicle, f, is thus represented by the density function $\phi_f^i(a,\gamma,t)$ whose dynamics is described by the following equation:

$$\begin{array}{ccc} \dfrac{\partial \phi_f^i}{\partial t} & + & \dfrac{\partial g_f(u_f)\phi_f^i}{\partial a} + \dfrac{\partial h_f(\gamma,u_f)\phi_f^i}{\partial \gamma} = -\lambda(\gamma,U)\phi_f^i \\ \text{on } \Omega_i, & & i \in \{1,2,3\}\ ,\ f \in \{1,\ldots,N\} \end{array}$$

The structuring variables, a and γ, correspond respectively to the age and maturity of the cells. $\phi_f^i(t;a,\gamma)$ represents the density of cells of age a and maturity γ in phase i at time t within the follicle f. Ω_1, Ω_2 and Ω_3 correspond to the different phases of the cell cycle within the age-maturity plane.The transfer from one phase to another is governed by appropriate boundary conditions. The aging rate, $g_f(u_f)$, and the maturation rate, $h_f(\gamma,u_f)$, are controlled by the level of FSH available on the follicular scale, noted by u_f, whereas the lost term, $\lambda(\gamma,U)$, is controlled by the plasmatic level of this hormone. The FSH levels are themselves defined in closed-loop and as a function of some moments of the density, leading to integro-differential terms. The stopping time coincident with ovulation triggering by the hypothalamus is scheduled by the maturity first-order moment reaching a threshold.

Numerical simulations of conservation laws can be performed in the framework of finite volume methods [11]. In our case, the numerical scheme is

based on high-resolution, wave-propagation algorithms developed for multi-dimensional hyperbolic systems [12]. Running simulations allows to predict the outcome of follicular development (ovulation and ovulation rate, or anovulation) for a given parameter combination. It can also be useful in testing control strategies. For instance, preliminary trials of exogenous FSH administration to compensate the drop in FSH suggest that it is possible to tune the ovulation rate finely [10].
The multi-level property of the model allows to propose an integrative scenario for the control of ovulation rate, where multiple ovulations ensue from the combined increased follicular sensitivity to FSH, leading to lower whole follicle cell numbers at ovulation time decreased sensitivity of the pituitary gland to ovarian negative feedback and/or decreased sensitivity of the hypothalamus to ovarian positive feedback. Whatever the ovulation rate, the ”losing” atretic follicles are those whose cells are confined within the maturity zone of vulnerability whereas cells of ”winning” ovulatory follicles can escape from this zone before the drop in FSH be too severe. An alternative follicular fate is to escape from atresia without becoming able of ovulating, if the cellular mass is too low.

Current works consist in exploring systematically the different situations engendered by the model from given combinations of parameters and solving the associated control problems (control of ovulation rate and chronology). Since the original PDE conservation law is not directly tractable, solving of such problems first goes through a step of model reduction. For instance, we have used the quite standard method of characteristics, in the framework of reachability theory, to look for the initial conditions which are either compatible with ovulation or atresia [13]. An alternative model reduction strategy is based on the specific asymptotic properties of the model. It amounts to handling a dynamical system of coupled nonlinear ordinary differential equations ruling the change in the mass and maturity of each follicle [14]. The dynamical system can be considered as a game where each follicle plays against the others for its survival and analysed in the framework of bifurcation theory.

References

[1] Greenwald G., Roy S., (1994) Follicular development and its control, in: Knobil E., Neill J.D. (eds.), The Physiology of Reproduction, Raven Press, New York, 629–724.

[2] Richards J.S., (1980) Maturation of ovarian follicles: actions and interactions of pituitary and ovarian hormones on follicular cell differentiation. *Physiol. Rev.* **15**: 51–89.

[3] Clément F., Gruet M.A., Monget P., Terqui M., Jolivet E., Monniaux D., (1997) Growth kinetics of the granulosa cell population in ovarian follicles: an approach by mathematical modelling. *Cell Prolif.* **30**: 255–270.

[4] Pisselet C., Clément F., Monniaux D., (2000) Fraction of proliferating cells in granulosa during terminal development in high and low prolific sheep breeds, *Reprod. Nutr. Dev.* **40**: 295-304.

[5] Clément F., (1998) Optimal control of the cell dynamics in the granulosa of ovulatory follicles,) *Math. Biosci.* **152**: 123-142.

[6] Richards J.S., Hedin L., (1988) Molecular aspects of hormone action in ovarian follicular development, ovulation and luteinization. *Annu. Rev. Physiol.* **50**: 441-463.

[7] Clément F., Monniaux D., Stark J., Hardy K., Thalabard J.C, Franks S., Claude D., (2001) Mathematical model of FSH-induced cAMP production in ovarian follicles. *Am. J. Physiol. (Endocrinol. Metab.)* **281**: E35-E53.

[8] Reiter R., Lefkowitz R.J., (2006) GRKs and β-arrestins: roles in receptor silencing, trafficking and signaling. *Trends Endocrinol. Metab.* **17**: 159-165.

[9] Heitzler D., Durand G., Dupuy L., Gauthier C., Piketty V., Crépieux P., Rizk A. , Soliman S., Fages F., Clément F., Reiter E., Modelling of FSHR-induced signalling network. In *International Conference on the Gonadotropins and their Receptors (July, 6th-7th)*, 2008.

[10] Echenim N., Clément F., Sorine M, (2007) Multi-scale modeling of follicular ovulation as a reachability problem. *Multiscale Model. Simul.* **6**: 895-912.

[11] LeVeque R.J., (2002) Finite Volume Methods for Hyperbolic Problems, *Cambridge University Press*,

[12] LeVeque R.J., (1997) Wave propagation algorithms for multidimensional hyperbolic systems. *J. Comput. Phys.* **131**: 327-353.

[13] Echenim, N. Clément F., Sorine M., (2007) Multi-scale modeling of follicular ovulation as a reachability problem. *Multiscale Model. Simul.* **6**: 895-912.

[14] Michel P., (2008) A singular asymptotic behavior of a transport equation. *C. R. Math. Acad. Sci. Paris*, **346** :155-159.

The VPH Toolkit for collaborative multi-scale modelling of multi-organ systems for the Physiome

S. Randall Thomas[1]

[1] CNRS FRE 3190, Evry, F-91000, France

Abstract

European Physiome activity is currently funded under the 7th Framework Program VPH call (Virtual Physiological Human), which supports more than a dozen targeted projects and also one Network of Excellence (NoE), whose mission is to coordinate these efforts, explore training possibilities, disseminate information about VPH resources and projects, and furnish a VPH ToolKit. I briefly summarize this activity and then focus on progress in the Renal Physiome, which is linked to the VPH NoE through the SAPHIR Exemplar Project treating blood pressure regulation in a Guyton-inspired modular modeling environment. The Renal Physiome includes: (1) the Quantitative Kidney DataBase (QKDB) of experimental measurements and anatomical details, (2) links to other disciplines (incl. high-throughput data) through common ontologies (e.g., FMA, GO...), (3) a library for models of channels and transporters expressed along the nephron, and (4) a repository of legacy models of kidney physiology at various scales (tubule transport, tubuloglomerular feedback, medullary models addressing the problem of urine concentration, etc.), with internet access through a KidneyGrid Portal (in collaboration with Univ. Melbourne).

1 Introduction

The IUPS Physiome project [4], and its European manifestation, the Virtual Physiological Human (in the European 7th Framework Program), is a complex systems approach to human physiology: the aim is to put in place a toolbox that will enable a cooperative environment for mathematical modeling of human physiology over a wide range of time and space scales. Associated with the model development are several tools to facilitate sharing and interconnection of models developed under various software environments and by different research groups.

The core model environment is a modular multi-resolution approach to the problem of accomodating in a single mathematical model the global effects (at the level of integrated physiological parameters such as blood pressure or salt balance) of local perturbations (such as the modification of the kinetics of a coupled ion transporter in a particular segment of the nephron, due, for

example, to a drug effect or to a mutation of the gene coding for the transporter protein). A basic requirement is to bridge the molecular-to-organism scales while keeping computation time manageable on a personal computer, since such a system is called for, for example, in order to provide in silico exploration of clinical treatment scenarios. In partial fulfillment of this need, we have implemented several modular versions of the classic Guyton [6] models, centered on blood pressure regulation, and one of the offshoots, the model of Ikeda et al. [8] targeting acid-base regulation. These are implanted in a multi-formalism simulation package (M2SL, developed at LTSI, Inserm U642, Rennes)[5] that allows replacement of the individual low-resolution basic modules of any given organ system by detailed mechanistic models representing the mechanisms involved in a target process [10]. I will present recent progress on two aspects of this project.

Towards a library of detailed local models for the Renal Physiome

There exists a rich literature of mathematical models of various aspects of renal function at all scales from the kinetics of membrane channels and coupled transporters, inclusion of these in tubule models of segmental reabsorption/ secretion along the nephron, to models of tubuloglomerular feedback regulation and renal hemodynamics, and medullary models of nephro-vascular solute and water recycling involved in the urine concentrating mechanism. However, there has been scant effort at integrating these detailed 'local' models into global descriptions of renal function sensitive to hormonal and neural controls and relevant to questions of the role of the kidneys in whole organism processes such as blood pressure regulation. This is precisely the goal of our core modelling environment.

Via the dependence of salt excretion on arterial pressure (PA), known as the 'pressure-natriuresis relation' or the 'renal function curve' [7], the kidney is responsible for setting the long-term level of PA. As a consequence, the kidney is necessarily involved in the genesis of hypertension, as reflected in the fact that genetic polymorphisms linked to hypertension invariably involve the kidney. In particular, several hypertension-related defects involve gain-of-function mutations in membrane proteins responsible for salt reabsorption in the late portions of the nephron. Major anti-hypertensive drugs, such as the thiazides, act by inhibiting these transporters. However, the effectiveness of such drugs depends on several other steps in the salt reabsorption process, including especially the Na, K-ATPase, which in turn depends on connections to the cytoskeleton.

In order to explore these scenarios in silico, we present a model of distal tubule transport based on kinetics of the implicated membrane channels and transporters [1, 2, 3, 9, 11, 12]. This involves the constitution of a library

of available kinetic descriptions for membrane channels and transporters (and their various isoforms) in the different cell types along the nephron. We are particularly interested in the role of polymorphisms of alpha-adducin (a cytoskeleton protein associated with the Na,K-ATPase in distal tubule cells) in the effectiveness of thiazide diuretics. The use of this library of kinetic models as a set of building blocks for integrated tubular models of secretion and reabsorption, and inclusion as part of the modular core modelling environment, is facilitated by coding the descriptions in an XML markup language, which serves as the pivot for translation into an arbitrary variety of numerical solution environments.

Brief Biography S. Randall Thomas received his M.A. in Biology in 1973 (Swarthmore College), a Ph.D. in Physiology in 1977 (Medical College of Virginia, Richmond), and the HDR from Univ. Paris 5 in 1990. After postdocs at the French CEA at Saclay (1979) and in the Dept. of Physiology at Univ. Texas Med. Center, Houston (1980-81), he became Chargé de Recherche with the CNRS, France, in 1982, where he is now a Director of Research. He currently works at IBISC CNRS FRE 3190 in Evry. His research centers on mathematical modeling of integrated transport systems in renal and epithelial physiology and on the development of related databases and physiome infrastructure. He is on the editorial board of the journals Nephron Physiology, Philosophical Transactions of the Royal Society, and Synthetic & Systems BIology. He is Director of the joint CNRS-INSERM Stic-Santé GdR (Groupement de Recherche), and is a founding member of the Institute of Theoretical Medicine (Lyon).

References

[1] Andersen, O. S., J. E. Silveira, et al. (1985). "Intrinsic characteristics of the proton pump in the luminal membrane of a tight urinary epithelium. The relation between transport rate and delta mu H." *J Gen Physiol* **86(2)**: 215-34.

[2] Chang, H. and T. Fujita (1999). "A kinetic model of the thiazide-sensitive Na-Cl cotransporter." *Am J Physiol* **276(6 Pt 2)**: F952-9.

[3] 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.

[4] Crampin, E. J., M. Halstead, P. Hunter, P. Nielsen, D. Noble, N. Smith and M. Tawhai (2004). "Computational physiology and the Physiome Project." *Exp Physiol* **89(1)**: 1-26.

[5] Defontaine, A., A. I. Hernández, and G. Carrault, "Multi-formalism Modelling of Cardiac Tissue," *Lecture Notes in Computer Science*, pp. 394-403, 2005.

[6] Guyton, A. C., T. G. Coleman and H. J. Granger (1972). "Circulation: overall regulation." *Annu Rev Physiol* **34**: 13-46.

[7] Guyton, A. C. (1987). "Renal function curve–a key to understanding the pathogenesis of hypertension." *Hypertension* **10(1)**: 1-6.

[8] Ikeda, N., F. Marumo, M. Shirataka and T. Sato (1979). "A model of overall regulation of body fluids." *Annals of Biomedical Engineering* **7**: 135-166.

[9] Monroy, A., C. Plata, et al. (2000). "Characterization of the thiazide-sensitive Na(+)-Cl(-) cotransporter: a new model for ions and diuretics interaction." *Am J Physiol Renal Physiol* **279(1)**: F161-9.

[10] Thomas, S. R., P. Baconnier, J. Fontecave, J.-P. Françoise, F. Guillaud, P. Hannaert, A. Hernandez, V. L. Rolle, P. Maziere, F. Tahi and R. J. White (2008). "SAPHIR: a physiome core model of of body fluid homeostasis and blood pressure regulation." *Philos Transact A Math Phys Eng Sci* **(in press)**.

[11] Weinstein, A. M. (2005). "A mathematical model of rat distal convoluted tubule. I. Cotransporter function in early DCT." *Am J Physiol Renal Physiol* **289(4)**: F699-720.

[12] Weinstein, A. M. (1995). "A kinetically defined Na+/H+ antiporter within a mathematical model of the rat proximal tubule." *J Gen Physiol* **105(5)**: 617-41.

PART II ARTICLES

Hyperstructures 2008-2009

Vic Norris, Patrick Amar, Marie Aimar, Pascal Ballet, Anne-Francoise Batto, Georgia Barlovatz, Gilles Bernot, Guillaume Beslon, Armelle Cabin, Sylvie Chevalier, Anthony Delaune, Jean-Marc Delosme, Eric Fanchon, Hongjun Gao, Nicolas Glade, Yohann Grondin, Danielle Hernandez-Verdun, Laurent Janniere, Francois Képès, Catherine Lange, Guillaume Legent, Corinne Loutelier-Bourhis, Franck Molina, Nicole Orange, Derek Raine, Camille Ripoll, Michel Thellier, Alain Thierry, Philippe Tracqui, Abdallah Zemirline

1 Introduction

Many functions in both prokaryotes and eukaryotes are performed by large structures in which molecules, macromolecules and ions are physically associated [23]. In the case of *Escherichia coli, Bacillus subtilis, Caulobacter crescentus* and other model bacteria, examples of such hyperstructures include: the array of chemotaxis-specific receptors (Tar, Tsr, Trg, Tap, and Aer) [4]; dynamic, coupled transcription-translation and transcription-translation-insertion (transertion) hyperstructures comprising active RNA polymerases and ribosomes along with the nascent mRNAs and nascent proteins and indeed the highly expressed genes themselves [3]; the cytoskeletal filaments MreB, CreS and FtsZ; filaments of elongation factor EF-Tu [20]; metabolons of glycolytic enzymes [33]; foci of ATP synthase and succinate dehydrogenase in the membrane [15]; foci of the enzymes E1 of the phosphoenolpyruvate:sugar phosphotransferase system; clusters of secretion enzymes such as SecA; nucleofilaments of recombination enzymes such as RecA; the cell division machinery (comprising a lipid domain and proteins such FtsZ, FtsA, FtsI, FtsK and AmiC); the DNA replication factory (comprising enzymes such as PolC, DnaB, DnaG and DnaE as well as enzymes responsible for the synthesis of precursors such as ribonucleoside diphosphate reductase); cellulosomes and polycellulosomes [10]. Ambiquitous enzymes can occupy two different positions in the cell [21] and some hyperstructures depend on such enzymes and are functioning-dependent structures that assemble only when functioning and that disassemble when no longer functioning [31]. Other hyperstructures are equilibrium or quasi-equilibrium structures that remain even in the absence of a flow of energy or nutrients.

It has been proposed that hyperstructures constitute a level of organisation intermediate between macromolecules and the bacterial cell itself [23]. Communication between hyperstructures would then take the form of changes in: DNA supercoiling (which may be modulated by the transcription in one hyperstructure to affect the transcription needed for another hyperstructure), ion

condensation on charged filaments (which by being concentrated in stabilising one hyperstructure may be in short supply for another hyperstructure), signalling molecules (which may be emitted by hyperstructures), water structures (which may lead to the formation of one hyperstructure at the expense of another), and distribution of membrane domains (whereby a hyperstructure with particular lipid preferences may stabilise another with the same preferences). At this intermediate level of organization, hyperstructures would control the phenotype and, in particular, the bifurcations that occur, as during the cell cycle, so that events take place in the right place, at the right time and in the right order.

Exploration of the hyperstructure concept may also prove useful for understanding eukaryotic cells. Interactions between hyperstructures in eukaryotes have been invoked to explain the structure and functioning of mitochondria and to explain the existence of mitochondrial DNA [32]. Members of the hyperstructures group work on podosomes, which are important in motility and mechanotransduction [9] and on the nucleolus, which has a multitude of roles in addition to the assembly of ribosomes [29]. Here we review briefly new discoveries about hyperstructures and describe some of the work being done by members of the hyperstructures group.

2 New hyperstructures

Evidence for new hyperstructures or for new aspects of known hyperstructures since our reviews in 2007 includes:

1. RNaseE is the main component of the RNA degradosome of *Escherichia coli*, which plays an essential role in RNA processing and decay. The degradosome also contains RNA helicase B, polynucleotide phosphorylase and, intriguingly in view of possible interactions between hyperstructures, the glycolytic enzyme, enolase. It has now been found that the degradosome forms helical filaments just under the membrane that may regulate access to substrates to prevent uncontrolled degradation [30]. The significance of the two hybrid finding for interaction between RNaseE and MinD (part of the Min hyperstructure for regulating cell division) remains to be determined.
2. Carboxysomes in bacteria such as *Synechococcus, Synechocystis* and *Halothiobacillus* enhance autotrophic carbon fixation via the Calvin cycle and are widely distributed among chemoautotrophs and cyanobacteria [7]. These hyperstructures, which the authors term 'microcompartments', are about 80-150 nm in cross section and are bounded by a 3-4 nm thick protein shell. Their mass is about 300 MDa and they are composed of several thousand polypeptides of 10-15 different types. There

is no evidence they contain lipids, RNA or DNA. The interior of the carboxysome contains the sequential metabolic enzymes carbonic anhydrase and ribulose bis-phosphate carboxylase monooxygenase. Carbonic anhydrase converts HCO3 to CO2 within the carboxysome, then ribulose bis-phosphate carboxylase monooxygenase converts CO2 and ribulose bisphosphate to 3-phosphoglycerate. The shell of the carboxysome may prevent diffusion of CO2 (and exclude O2) and so keep it close to the ribulose bis-phosphate carboxylase monooxygenase.

3. Pdu microcompartments or hyperstructures are responsible in bacteria such as *Salmonella enterica* for B12-dependent 1,2-propanediol utilization (pdu) [7]. They are 100-150 nm in cross section with a 3-4 nm protein shell. Their mass is about 600 MDa and they are composed of about 18,000 individual polypeptides of about 14-18 different types. They are believed to sequester an intermediate of 1,2-propanediol degradation (propionaldehyde) to prevent toxicity and diffusive loss. 1,2-propanediol degradation is also implicated in the pathogenesis of *Salmonella* and *Listeria*. In *S. enterica*, the genes involved in 1,2-propanediol degradation form a contiguous, twenty-three-gene cluster, one of the largest clusters of functionally related genes in this organism. It might be expected that coupled transcription, translation and assembly would mean that genes, nascent mRNA and nascent protein would at some stage be physically coupled to the Pdu hyperstructure yet, like carboxysomes, the Pdu hyperstructures are not believed to contain RNA, DNA or lipids. At least, not at present.

4. Eut microcompartments or hyperstructures are responsible in S. enterica for B12-dependent ethanolamine utilization (Eut) [7]. Ethanolamine utilization begins with conversion of ethanolamine to acetaldehyde by the B12-dependent enzyme, ethanolamine ammonia lyase, after which acetaldehyde is degraded to acetate and ethanol by a pathway analogous to 1,2-propanediol degradation. Overall, the degradation of 1,2-propanediol and ethanolamine share many features with the main difference being that 1,2-propanediol is a C3 compound and ethanolamine is a C2 compound. It has been suggested that the Eut hyperstructure concentrates substrates and enzymes to either increase metabolic efficiency or regulate metabolite levels.

5. Pyr hyperstructures are involved in *E. coli* in the phosphorylation of uridine monophosphate [35]. PyrH and MetK formed discrete protein foci within the cell as well of which 1 or 2 moved around the cytoplasm in a circles. PyrH also regulates the transcription of carbamoyl phosphatase and is thought to play a key role in chromosome partitioning and cell division.

6. The 'stressosome' in *B. subtilis*, a 1.8-megadalton hyperstructure of many different proteins that integrates a variety of signals to effect a single outcome [19]. These include the Rsb proteins which interact physically with one of the enzymes responsible for DNA replication, DnaE (Laurent Janniere, unpublished)
7. Podosomes are transient intracellular organelles with a biological role that is the subject of much active research and debate. Podosomes display many features of hyperstructures, notably as potential functioning-dependent structures. Indeed, podosomes are actin-rich membrane structures, with size ranging from 0.5 to few ?m, that form close contact with the surrounding substrate. Podosomes are more dynamic and instable than focal adhesions, dissolving and then reforming in new locations. They typically appear as a ring of adhesive molecules centred around a dense F-actin core [12]. It is still uncertain whether the first event in the formation of podosomes is actin nucleation or adhesion to the extracellular matrix (which leads to the activity of scaffolding proteins made up of combinations of protein/lipid and protein/protein interaction motifs involved in podosome formation and turnover). Furthermore, circular superstructures of podosomes, called rosettes, are formed spontaneously or after the stimulation with growth factors. The formation of podosomes and sealing zones as primary adhesive structures of resorbing osteoclasts upon contact with bone is well documented, but podosomes can be formed on a variety of other substrates. Interestingly, they appear as functioning-dependent mechanosensors [8]. Thus, they are sensitive to changes in substrate rigidity [9] and influenced by cellular contractility. Podosomes represent a powerful paradigm to study the mechanosensory machinery that integrates cytoskeletal elements with adhesion and signalling.

3 Modelling hyperstructures

3.1 HSIM

The stochastic automaton, HSIM, has been used to show that PTS and glycolytic metabolons can increase production of pyruvate eightfold at low concentrations of phosphoenolpyruvate [1]. A fourfold increase in the numbers of enzyme EI led to a 40% increase in pyruvate production, similar to that observed in vivo in the presence of glucose. Surprisingly, little improvement resulted from the assembly of metabolons into a hyperstructure. However, HSIM allowed us to see that such assembly is a powerful way of generating gradients of metabolites and signalling molecules. During the last year, we have been using HSIM to investigate how pausing during the replication of

the chromosome might lead to differential gene expression [25]. Although the situation analysed was somewhat different from the one we envisage, it is encouraging that there is a report of replication-associated changes in the expression of the dps, pyrI and gapA genes [2].

3.2 BioDyne

BioDyn is a multiagent software made to simulate self-assembly systems at cellular level. An agent is a cube located in a 3D grid. An agent can move according to the three axes (x,y and z) with a certain probability per simulation step. To allow self-assembly, two kinds of interactions are modelled. A short-distance adhesion interaction allows the agents to aggregate and form large rigid structures like actin filaments. Adhesion is given a probability which allows structures to be reconfigurable. A medium-distance attraction/repulsion interaction is designed to reproduce the action of 'soft' interactions which lead, for example, to the formation of the mitochondrial network. In order to improve the number of biological mechanisms that BioDyn can simulate, a software engine of reactions between agents has been made. It can reproduce phenomena like agent proliferation, agent destruction, agent creation or agent modification. The combination of the interactions and the reaction engine makes possible, in an abstract and qualitative manner, the reproduction of different biological hyperstructures. This system has different drawbacks. For instance, it is impossible to rotation an agent (only translations are allowed) and the multiscale is limited to one order of magnitude. BioDyn is available as a java applet at the url http://netbiodyn.tuxfamily.org. Example and tutorials are also available online.

3.3 Interaction networks

Interaction networks can prove useful in modelling hyperstructures. Indeed, a hyperstructure can be modelled by means of an interaction network $R = (N, I, f)$ which is defined as follows:

- N is a set of nodes representing the molecules, macromolecules and ions which, by associating physically, constitute the hyperstructure;
- I is the set of arcs and hyperarcs representing, respectively, the simple interactions (involving two or more elements of N, i.e. two molecules, macromolecules or ions) and the multiple interactions (involving more than two elements of N);
- f is an interaction function defined on the set I and whose the values could represent properties of the interactions (for example their intensities).

Such a representation would allow the simulation (e.g. via a Multi-Agent System) of processes involving hyperstructures such as the formation of a hyperstructure, the interactions between its constituents, or the interaction function related to it, etc.

Let us consider the example of the compaction hyperstructure (for references see [24]. In *E. coli*, the MukB protein is localized to discrete structures, with reports suggesting that it forms either foci at the [1/4] and [3/4] positions during the cell cycle or larger oblongs in the nucleoid. In vitro, MukB is associated with MukE and MukF in a large complex. MukB is a member of the SMC superfamily (like RecN), while MukF is a non-SMC protein or kleisin. It is generally believed that the MukB, MukE, and MukF proteins form a "condensin" that compacts DNA, probably in association with DNA gyrase, and that this condensin assists in the separation of sister chromosomes. With such a role in chromosome topology, it is not surprising that the MukB and SeqA foci are related. Indeed, the latter are perturbed in both size and distribution in the mukB null mutant. An interaction network of the type $R = (N, I, f)$ could then be used to obtain an interaction network $R' = (N', I', f')$ modelling a compaction hyperstructure; this would have a node set N = { condensing(MukB, MukE, MukF); chromosomes; DNA gyrase; DNA}, and a interaction set I which express the interactions respectively between condensing(MukB, MukE, MukF) and DNA gyrase and DNA, and between condensing(MukB, MukE, MukF) and chromosomes. The interaction network R' would have a node set I' = {sister chromosomes, complex(Mukb, SeqA foci), compacted DNA}.

3.4 Functioning-dependent structures

The concept of functioning-dependent structure (FDS) describes an assembly of objects that forms and maintains itself as a result of its action in accomplishing a task and that dissociates when its task is over. It is therefore a scale-free concept. To explore quantitatively this concept at the level of macromolecules, we have studied the behavior in steady state of a simple model of the functioning-dependent association of two sequential enzymes of the Michaelis-Menten type (Thellier et al., 2006). We are now studying the properties of an FDS in transient states using the MAPLE 9.5 software to solve the set of ordinary non-linear differential equations that constitute the model. Our initial findings are that a two-enzyme FDS can have properties that are very different from those of a system of free enzymes. The most salient of these properties is the generation of bursts of the final product. We suggest that, in principle, almost any metabolic pathway in which the enzymes can associate in a functioning-dependent manner can generate signals.

Some of the hyperstructures mentioned above fall into the class of functioning-dependent structures. This is the case for the pdu hyperstructures which disappear when the substrate, 1,2-propanediol, is removed and which cannot be purified in the absence of 1,2-propanediol. Remarkably, pdu hyperstructures do not form in vivo in the absence of 1,2-propanediol even under conditions where the genes are induced (Tom Bobik, personal communication). This may also be the case of the RNA degradosome (see above) where the association between RNase E, Hfq (an RNA binding protein) and sRNA (small, non-coding regulatory RNA) may also depend on the binding of the substrate. The sRNA act by base pairing with mRNA targets to inhibit translation and promote mRNA degradation. In *E. coli*, Hfq is required for the function of sRNA and RNase E has a role in the degradation of mRNA targets. Note that sRNAs are not present in normally growing cells; they are induced by a 'stress' or 'shock', accumulating to relatively high concentration for a transient period. Since researchers in the field have failed to detect a direct protein-protein interaction between Hfq and RNase E, one model for the interaction is that the sRNA is acting as a 'bridge' between Hfq and RNase E. The evidence suggests that sRNA induction leads to remodeling in which the canonical RNA degradsome is 'transformed' into a complex containing RNase E, Hfq and sRNA (for references see [6]).

3.5 Globally Constrained Shape Deformation

Jean-Marc Delosme has been working on the deformation of shapes under constraints on both perimeter and area which occur during the dynamics of hyperstructures such as the Golgi. Modelling these deformations is a challenging task due to the highly non-trivial interaction between the need for flexible local rules for manipulating the boundary and the global constraints. In collaboration with Freddy Bruckstein and Ishay Goldin, he has developed several methods to address this problem that entail "random walks" in the space of shapes obeying time-varying constraints on their perimeter and area [13]. Design of perimeter and area preserving deformations are an interesting and useful special case of this problem. The resulting deformation models are employed in annealing processes that evolve original shapes toward shapes that are optimal in terms of boundary bending-energy or other functionals. Furthermore, such models may find applications in the analysis of sequences of real images of deforming objects obeying global constraints as building blocks for registration and tracking algorithms.

3.6 Hyperstructure interactions

A think-tank in the Epigenomics Project (a 'Nirvana') brought together physical chemists (Jerry Manning, Max Berkowitz and Camille Ripoll), a physicist specialising in water structures (Alfons Geiger), and biologists interested in hyperstructures (Frank Mayer, Francois Kepes and Vic Norris). The first objective was to relate the phenomenon of counterion condensation (whereby a fraction of the counterions will run along a linear, charged polymer provided that the number of charges per unit length exceeds a critical value) to water structures and to 2-D structures such as membrane domains [28, 18]. The second objective was to treat counterion condensation as a solution looking for a problem in biology and then to come up with that problem. There is an urgent need in systems biology for solutions to the problem of how cells negotiate the enormity of phenotype space and, in particular, of how cells generate the coherent and reproducible phenotypes on which natural selection acts. The think-tank therefore had the objective of exploring a possible contribution in terms of ion condensation. A significant advance was made when the regulation of the cell cycle of bacteria was identified as the area in which suitable biological problems might be found - and the nature of the key step in the initiation of chromosome replication was identified as the most promising fundamental problem to explore. A hypothesis was then formulated in which Initiation is controlled by an initiation hyperstructure in which:

1. The key event of the separation of the strands at the origin of replication is mediated in part by the DnaA protein. This separation is favoured by the decondensation of ions from the origin region which leads to the strands in the origin region repelling one another. Such movements of ions involve linear filaments, membrane domains and water structures.
2. Sequestration of newly replicated origins of replication is controlled by a sequestration hyperstructure. The formation of this hyperstructure is favoured by the recondensation of ions occurring as the GATC sequences in the origin DNA go from being fully methylated to being hemi-methylated so increasing the charge parameter. This condensation also stabilises polymers of the SeqA protein which binds to these sequences.
3. Sequestration ends when ions decondense from this DNA and from the SeqA polymers in the sequestration hyperstructure to condense elsewhere on other hyperstructures.
4. The fundamental nature of the cell cycle has its origins in the coupling between ion condensation and the growth of the cell in terms of the production of the linear polymers and membranes that constitute certain hyperstructures. When the mass of polymers in the form of, for example,

ribosomal hyperstructures, reaches a critical threshold, condensation on these polymers at the expense of condensation on origin DNA leads to the decondensation step in 1/ that triggers initiation.

Encouragingly, it was realised that this mechanism might also underlie a related phenomenon in *E. coli*, that of growth rate control in which there is regulation of the proportion of the bacterial mass in the form of the transcriptional and translational apparatus. The group also realised that ion condensation and water structures might play important roles in the separation of the chromosomes and in membrane domain formation at the start of cell division.

4 Experimental advances

4.1 DNA replication and glycolysis

The replication of DNA and the central metabolism of carbon have recently been shown to be related in *Bacillus subtilis* where the chain elongation step of replication can be modulated by the flux of carbon through the bottom part of the glycolytic pathway [14]. Our expectation is that hyperstructure dynamics would underpin this coupling. How might DNA replication hyperstructures and glycolytic hyperstructures communicate? The alarmone ppGpp is over-produced when translating ribosomes meet uncharged tRNAs or when carbon sources are depleted. Over-production of ppGpp is strongly implicated in the coupling between replication and translation when there is a dramatic shortage of amino acids [34]. Is ppGpp produced at different basal level in steady state cells grown in different media also involved in this coupling? Our preliminary results indicate that it is not involved.

4.2 Membrane cartography

A collaboration between microbiologists and chemists in Rouen is intended to provide information on the lipid preferences of abundant membrane proteins as a way to study those hyperstructures that affect the dynamics and composition of the *E. coli* membrane. Initial work has focussed on a thorough mass spectometry-based analysis of the lipids making up this membrane [26, 27]. This is the basis for future experiments designed to reveal the changes in lipid composition that accompany overproduction of particular membrane proteins.

5 Discussion

Extending the concept of functioning-dependent structures to hyperstructures raises some interesting questions. The synthesis of macromolecules (DNA,

RNA, proteins, glycans, polyphosphates) is an important part of what makes a cell a cell. There is now ample evidence that such synthesis occurs in hyperstructures such as the lac hyperstructure where 30 or so RNA polymerases work together to generate the nascent mRNAs that are translated by 300 or so ribosomes [16]. Perhaps the most important one, again in *E. coli*, would be the putative ribosomal hyperstructure bringing together rRNA genes, rprotein genes and their products in a giant assembly plant (like the eukaryotic nucleolus) [5] although this is controversial [17]. The question we should ask is 'what do our FDS studies tell us about the dynamics of hyperstructures'? But first, what do our studies tell us about the signalling properties inherent in the synthesis of macromolecules by a single RNA polymerase or ribosome?

At the level of the synthesis of an individual protein, there is a set of individual reactions that add amino acids to the growing peptide. An individual reaction is between the amino acid and the nascent peptide as catalysed by the codon and the tRNA. The separate reactions that must be linked to make a protein require the channelling that is provided by the ribosome. So we might suppose the synthesis of an individual protein is an extreme example of an FDS. In terms of signalling, we can ask about what happens when protein synthesis is interrupted. Are peptide fragments released as signals (for references see [11])? Are ribosomes released (perhaps post-translationally modified)? And what about GTP/GDP and alarmones such as ppGpp and Ap4A? Note that at certain times in bacteria, many proteins are made only to be degraded. We might ask similar questions about RNA synthesis ...

So how is the functioning-dependent lac hyperstructure different from the FDS for synthesizing a single protein? In terms of complete assembly and disassembly of the FDH, specific lipids normally bound to nascent proteins in the FDH might be released by disassembly (with accompanying changes in viscosity of the membrane). Assembly and disassembly might affect water structures, DNA supercoiling and ion condensation. So topological signals are probably greater for an FDH than for an FDS. More usefully for our modelling, when a ribosome meets an uncharged tRNA, what happens to the following ribosomes (and what would be the equivalent for RNA polymerases?) - maybe they are all released? In other words, is there an amplification of effects in an FDH?

Finally, the insight that the hyperstructure hypothesis offers into bacterial physiology via hyperstructures extends to pathogenesis. The degradosome plays a key role in the virulence of bacteria such as Yersinia pestis [36]. This virulence is extremely sensitive to levels of calcium and to temperature, factors that are important in ion condensation. As mentioned about, it may be significant that the degradosome exists in the form of the linear filaments on

which, conceivably, ions might condense to regulate virulence. In which case, targetting such hyperstructures via hybrid metabolites or hybolites may prove of value [22].

References

[1] Amar, P., G. Legent, M. Thellier, C. Ripoll, G. Bernot, T. Nystrom, M. H. Saier, Jr. & V. Norris, (2008) A stochastic automaton shows how enzyme assemblies may contribute to metabolic efficiency. BMC systems biology 2: 27.

[2] Bechtloff, D., B. Grunenfelder, T. Akerlund & K. Nordstrom, (1999) Analysis of protein synthesis rates after initiation of chromosome replication in Escherichia coli. Journal of bacteriology 181: 6292-6299.

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

[4] Cabin-Flaman, A., C. Ripoll, M. H. Saier, Jr. & V. Norris, (2005) Hypothesis: chemotaxis in Escherichia coli results from hyperstructure dynamics. Journal of molecular microbiology and biotechnology 10: 1-14.

[5] Cabrera, J. E. & D. J. Jin, (2003) The distribution of RNA polymerase in Escherichia coli is dynamic and sensitive to environmental cues. Molecular microbiology 50: 1493-1505.

[6] Carpousis, A. J., (2007) The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E. Annual review of microbiology 61: 71-87.

[7] Cheng, S., Y. Liu, C. S. Crowley, T. O. Yeates & T. A. Bobik, (2008) Bacterial microcompartments: their properties and paradoxes. Bioessays 30: 1084-1095.

[8] Collin, O., S. Na, F. Chowdhury, M. Hong, M. E. Shin, F. Wang & N. Wang, (2008) Self-organized podosomes are dynamic mechanosensors. Curr Biol 18: 1288-1294.

[9] Collin, O., P. Tracqui, A. Stephanou, Y. Usson, J. Clement-Lacroix & E. Planus, (2006) Spatiotemporal dynamics of actin-rich adhesion microdomains: influence of substrate flexibility. Journal of cell science 119: 1914-1925.

[10] Doi, R. H. & A. Kosugi, (2004) Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nature Reviews Microbiology 4: 541-551.

[11] Gerdes, K., S. K. Christensen & A. Lobner-Olesen, (2005) Prokaryotic toxin-antitoxin stress response loci. Nature reviews 3: 371-382.

[12] Gimona, M., R. Buccione, S. A. Courtneidge & S. Linder, (2008) Assembly and biological role of podosomes and invadopodia. Current opinion in cell biology 20: 235-241.

[13] Goldin, I., J.-M. Delosme & A. Bruckstein, (2008) Vesicles and amoebae: on globally constrained shape deformation. In., pp.

[14] Janniere, L., D. Canceill, C. Suski, S. Kanga, B. Dalmais, R. Lestini, A. F. Monnier, J. Chapuis, A. Bolotin, M. Titok, E. L. Chatelier & S. D. Ehrlich, (2007) Genetic evidence for a link between glycolysis and DNA replication. PLoS ONE 2: e447.

[15] Johnson, A. S., S. van Horck & P. J. Lewis, (2004) Dynamic localization of membrane proteins in Bacillus subtilis. Microbiology 150: 2815-2824.

[16] Kennell, D. & H. Riezman, (1977) Transcription and translation frequencies of the Escherichia coli lac operon. Journal of molecular biology 114: 1-21.

[17] Lewis, P. J., S. D. Thaker & J. Errington, (2000) Compartmentalization of transcription and translation in Bacillus subtilis. EMBO Journal 19: 710-718.

[18] Manning, G. S., (2007) Counterion condensation on charged spheres, cylinders, and planes. J Phys Chem B 111: 8554-8559.

[19] Marles-Wright, J., T. Grant, O. Delumeau, G. van Duinen, S. J. Firbank, P. J. Lewis, J. W. Murray, J. A. Newman, M. B. Quin, P. R. Race, A. Rohou, W. Tichelaar, M. van Heel & R. J. Lewis, (2008) Molecular architecture of the "stressosome," a signal integration and transduction hub. Science 322: 92-96.

[20] Mayer, F., (2006) Cytoskeletal elements in bacteria Mycoplasma pneumoniae, Thermoanaerobacterium sp., and Escherichia coli as revealed by electron microscopy. Journal of molecular microbiology and biotechnology 11: 228-243.

[21] Nemat-Gorgani, M. & J. E. Wilson, (1980) Ambiquitous behavior–a biological phenomenon of general significance? Curr Top Cell Regul 16: 45-54.

[22] Norris, V., P. Amar, M. Thellier, G. Legent, A. Cabin, J.-C. Plaquevent, M. H. Saier jnr, G. Bernot, G. Misevic & C. Ripoll, (2007a) Hybolites: novel therapeutical tools based on stochastic automata. In: Modelling complex biological systems in the context of genomics. P. Amar, F. Képès, V. Norris & G. Bernot (eds). Evry, France, pp. 83-88.

[23] Norris, V., T. D. Blaauwen, R. H. Doi, R. M. Harshey, L. Janniere, A. Jimenez-Sanchez, D. J. Jin, P. A. Levin, E. Mileykovskaya, A. Minsky, G. Misevic, C. Ripoll, M. Saier Jnr., K. Skarstad & M. Thellier, (2007b) Toward a Hyperstructure Taxonomy. Annual review of microbiology 61: 309-329.

[24] Norris, V., T. den Blaauwen, A. Cabin-Flaman, R. H. Doi, R. Harshey, L. Janniere, A. Jimenez-Sanchez, D. J. Jin, P. A. Levin, E. Mileykovskaya, A. Minsky, M. Saier, Jr. & K. Skarstad, (2007c) Functional taxonomy of bacterial hyperstructures. Microbiology and Molecular Biology Reviews 71: 230-253.

[25] Norris, V., L. Janniere & P. Amar, (2007d) Hypothesis: Variations in the rate of DNA replication determine the phenotype of daughter cells. In: Modelling complex biological systems in the context of genomics. P. Amar, F. Képès, V. Norris & G. Bernot (eds). Evry, France: EDP Sciences, pp. 71-81.

[26] Oursel, D., C. Loutelier-Bourhis, N. Orange, S. Chevalier, V. Norris & C. M. Lange, (2007a) Identification and relative quantification of fatty acids in Escherichia coli membranes by gas chromatography/mass spectrometry. Rapid Commun Mass Spectrom 21: 3229-3233.

[27] Oursel, D., C. Loutelier-Bourhis, N. Orange, S. Chevalier, V. Norris & C. M. Lange, (2007b) Lipid composition of membranes of Escherichia coli by liquid chromatography/tandem mass spectrometry using negative electrospray ionization. Rapid Commun Mass Spectrom 21: 1721-1728.

[28] Ripoll, C., V. Norris & M. Thellier, (2004) Ion condensation and signal transduction. BioEssays 26: 549-557.

[29] Sirri, V., S. Urcuqui-Inchima, P. Roussel & D. Hernandez-Verdun, (2008) Nucleolus: the fascinating nuclear body. Histochemistry and cell biology 129: 13-31.

[30] Taghbalout, A. & L. Rothfield, (2007) RNaseE and the other constituents of the RNA degradosome are components of the bacterial cytoskeleton. Proceedings of the National Academy of Sciences of the United States of America 104: 1667-1672.

[31] Thellier, M., G. Legent, P. Amar, V. Norris & C. Ripoll, (2006) Steady-state kinetic behaviour of functioning-dependent structures. The FEBS journal 273: 4287-4299.

[32] Trinei, M., J.-P. Vannier, M. Beurton-Aimar & V. Norris, (2004) A hyperstructure approach to mitochondria. Molecular microbiology 53: 41-53.

[33] Velot, C., M. B. Mixon, M. Teige & P. A. Srere, (1997) Model of a quinary structure between Krebs TCA cycle enzymes: a model for the metabolon. Biochemistry 36: 14271-14276.

[34] Wang, J. D., G. M. Sanders & A. D. Grossman, (2007) Nutritional Control of Elongation of DNA Replication by (p)ppGpp. Cell 128: 865-875.

[35] Watt, R. M., J. Wang, M. Leong, H. F. Kung, K. S. Cheah, D. Liu, A. Danchin & J. D. Huang, (2007) Visualizing the proteome of Escherichia coli: an efficient and versatile method for labeling chromosomal coding DNA sequences (CDSs) with fluorescent protein genes. Nucleic acids research 35: e37.

[36] Yang, J., C. Jain & K. Schesser, (2008) RNase E regulates the Yersinia type 3 secretion system. Journal of bacteriology 190: 3774-3778.

Trail Systems as fault tolerant wires and their use in bio-processors

Nicolas Glade[1], Hedi M. Ben Amor[1], Olivier Bastien[2]

[1] TIMC-IMAG Laboratory, UMR CNRS 5525, University Joseph Fourier, Faculty of Medicine of Grenoble, 38700 La Tronche
[2] INRA - PCV, CEA Grenoble, 17 Av. des Martyrs, 38054 Grenoble

Abstract

Motivated by the idea that one day, probably far in the future, the computers and robots will be architectureless, made of collections of numerous 'intelligent' subsystems or nanomachines able to self-organize each other into computational morphologies with perhaps more computational power than classical electronic-based computers, many studies are burgeoning in different fields (chemistry, biology, condensed matter, quantum physics, ...). Several systems inspired from Nature have indeed been proposed yet for designing unconventional computer architectures using processing modes of various nature and at different scales.

The heterogeneous set of natural or artificially designed systems called *trail systems*, commonly associated to self-driven particles (agents[1]) with tropistic activity (through a communication based on traces let in the environment), is a soft matter with self-organizing properties sufficiently robust and fine for designing biocomputing structures. In this context, individual trails systems could be viewed as single wires and logical gates in a self-organized bio-processor, in the same manner axons are connecting the neural nodes in a neuro-processor. Their efficiency as wires depends on their specific properties which are often related to their scale. The robustness of their self-organization at the microscopic scale level occurring in a noisy environment, can be studied by a model based on effective computing systems (*i.e.* Turing machines) programmed to behave first as deterministic and perfect trail systems, then as stochastic-working trailing agents subject to randomness.

Keywords: Trail systems, Self-organization, Ant-based model, Robustness, Biocomputing, Bio-wires

1 Introduction

It is obvious for most scientists that natural systems possess indeed that potentiality to compute things. Nevertheless biocomputing approaches are often

[1] In order to differentiate them from any multi-agent system, the agents of a trail system will be called 'TS-agents' in the article

very difficult since the researchers block in implementing effective computational calculi with their system of interest. In other terms, it is easy to imagine the possible computational power of a biological system but one can not calculate easily '1 plus 1' by using molecules, living cells or other natural systems, unless applying to those system a strong control at the expense of the real potential of these systems (self-organization capabilities notably). Moreover, in comparison to electronic based processors that have a deterministic behavior and whose structure is well defined, natural systems are on the contrary very subject to noise, show stochastic fluctuations to dramatic changes in their behavior, and are poorly structured. In addition, their structuring is non-permanent compared to silicon based hardwares. These two point were largely analyzed by M. Conrad – died in 2002 –, one of the most prolific fathers of the notion of 'molecular computing' [1, 2].

In this article we bring close together a class of self-organizing natural systems, the trail systems, and effective programmable systems such as Turing machines whose behaviors are well known.

Natural systems are in general noisy and tend to the extend of their entropy. However some of them self-organize either by static mechanisms (*e.g.* liquid crystals) or by dynamic dissipative processes as it is the case in almost all biological systems, e.g. the organization of cells from tissues to organisms, population dynamics, or trail systems. By a permanent consumption of energy (or matter, such as reactive compounds), the latter self-organize over space and time and maintain their order to low entropy levels (compared to the entropy corresponding to a total disordered distribution of their constituents) [3, 4]. Self-organization of the so-called dynamical systems, or collective systems, always occurs at a macroscopic level compared to the microscopic level at which their components act. It is due to the combination of the numerous individual actions of the microscopic constituents, or agents (*e.g.* animals, cells, molecules), and the communication between them that synchronizes those individual actions. In that, they get closer to effective programmable processors than other natural systems like disordered solutions of molecules.

Processors are arrangements of wires and elementary signal integrators such as logical gates in electronic processors or neural nodes in neuro-processors. Specialized processors are *programmed structurally*. This means that one architecture computes a given input signal according to only one program. This is the case for the neuro-processors once their organization is fixed, or for some electronic chips dedicated to one computational task. Moreover, some processors, like those we use in our computers, can be also *programmed dynamically*. Their architecture is designed for a general use. Applying on the architecture a set of instructions as the computing process goes along (in other terms a program) corresponds to its refinement (and a dynamic adaptation).

Among bioprocessors, neural networks are not the only natural or bio-inspired processors that are able to learn or change a configuration and to self-adapt to a given context so as to process the information differently.

Biological regulatory networks are very often subject to changes of external (environmental) or internal (intracellular) conditions, so they have to adapt so as the cells or organisms can survive. Biological feedbacks and feed forwards act as activators or inhibitors of biological pathways (*e.g.* sequential cascades of reactions), respectively in the past or in the future of the biochemical process trajectory. A learning bio-processor based on microtubules and motors self-organization has also been proposed by Pfaffmann and Conrad [5]. In this model, microtubules are assumed to be able to transport the signal in an electric form (*nb.* it is important indeed to highlight the importance of the difference of nature between the constituents of the processor and that of the signals that travel through its wires and gates for being treated). Starting from a solution of microtubules and linker molecules (such as molecular motors), a reticulated morphology forms and constitutes the architecture of a certain processor. Used for processing a certain input, it returns an output. The latter is compared to a reference (calculated differently) and adjust progressively the architecture of the microtubule-based processor by changing the level of reticulation. If the processor computes well, the architecture freezes; if not, the architecture is warmed and changes. Such microtubule-based processors are plausible given the subsequent results that concern the control of their self-organization with linkers or motors [6, 7] and by admitting the concept of information transfer and processing inside – or at the surface of – microtubules, largely defended by several authors such as Hameroff [8, 9, 10], Tuszinsky [11, 12] (who also made recently calculations of ionic waves propagation along actin fibers), or other authors more recently [13, 14].

The plasticity of a neuro-processor compared to an electronic one is another important point to keep in mind. If the latter can be dynamically programmed (after being designed and structured by engineers over several years of successive ameliorations), the structure of the former can be reconfigured both by physical changes in the set of connections and by adaptations of its weights and thresholds (which in real neural networks found an equivalence in the increase or decrease of the number of available receptors to neurotransmiters or in the increase or decrease of the number, shape and volume of sensing areas, i.e. dendritic spines). That way, by a learning process that compares for a given set of entries the result with a reference (professor), neural networks can reconfigure into different neural-processors. Since all neurons can potentially connect to all the others in the network, including themselves (in particular in formal networks), and given their high reconfigurability, their structural combinatory is largely higher than that of electronic programmable

processors. Plasticity is a advantageous feature of formal neural networks, but this also occurs within the brains when we learn a better or different manner to proceed for doing something. After an accident that destroyed a part of the brain or after some surgical resections, other parts can also reconfigure to reestablish the function lost. On the other hand, one must accept that such a processor is not optimal, in particular when the number of neurons implied is very important.

Neural networks are not the only natural or bio-inspired processors that are able to learn or change a configuration and to self-adapt to a given context so as to process the information differently. Biological regulatory networks are very often subject to changes of external (environmental) or internal (intracellular) conditions, so they have to adapt so as the cells or organisms can survive. Biological feedbacks and feed forwards act as activators or inhibitors of biological pathways (*e.g.* sequential cascades of reactions), respectively in the past or in the future of the biochemical process trajectory. A learning bioprocessor based on microtubules and motors self-organization has also been proposed by Pfaffmann and Conrad [5]. In this model, microtubules are assumed to be able to transport the signal in an electric form. Starting from a solution of microtubules and linker molecules (such as molecular motors), a reticulated morphology forms and constitutes the architecture of a certain processor. Used for processing a certain input, it returns an output. The latter is compared to a reference (calculated differently) and adjust progressively the architecture of the microtubule-based processor by changing the level of reticulation. If the processor computes well, the architecture freezes; if not, the architecture is warmed and changes.

We think that self-organized trail systems can be considered for structuring bioprocessor fine architectures and can work as neural networks do. Although they are less robust than electronic circuits or even neurones, they are however able to reconfigure more rapidly. Moreover, they sometimes make some errors. Errors confer a great advantage to collective systems : when some individuals behave differently to the mass, one would say they make errors, others would say they explore other potentialities. The well known example is the determination of the best pathway among several paths that join together two points (*e.g.* the nest and a food source) by a population of ants [15, 16]. By collective dynamics based on the release and the sensing of pheromone trails by individual ants, the ant colony determines progressively the best pathway, but some of the ants – that we will call 'explorers' – fail in following the principal track and use another pathway. If the better one is blocked by an experimentalist (or by an environmental accident such as a rockfall or due to flooding) the secondary options that has already been found by some ants are rapidly used. In the case of ants, the exploratory trajectory is mostly random but can be enhanced by other external factors (smells, geometry constrains of the landscape ...). Then,

moving to other pathways is something easy for the colony since the explorers have found yet these new pathways. The rest of the colony just has to follow their tracks (*i.e.* pheromone trails). By this way, the colony can also discover other sources of food (resources in general) or new territories for settlement.

Very few – but successful – attempts were done that showed that NP complex problems such as network routing problems are perfectly and more efficiently resolved when based on trail processes (as shown with the use of virtual ants by Bonabeau et al [15]). The different systems that belong to this heterogeneous set are well studied independently now. Nevertheless they have surprisingly never really been viewed more theoretically as a unique family of systems sharing the same properties and studied as it is, as a generic model described by a unique set of parameters for each TS-agent and its trail. Neither have they really been studied for their computational properties. In the best case, they are considered in a simplified manner as multi-agent systems [17], but this does not take unfair advantage of all their characteristics.

Trail systems possess however very interesting characteristics in terms of computational efficiency and control. Their ability to produce preferential tracks and to follow – or be influenced by – them results in the emergence of temporal and/or spatial order and forms. That way, the whole system behaves naturally as a self-organizing dynamical network (a sort of circuitry where cables are defined dynamically) that can – when perturbed by external factors acting globally on the whole system or locally in the form of perturbing nodes (*e.g.* food sources, poison, walls) – re-adapt to take into account the new geometry and features of the system in its environment. Moreover the structurability of such systems is intrinsically high due to their structural aspect and to their self-organizing properties, and it can be enhanced by a direct control of the dynamics and the orientation of the TS-agents. The majority of biocomputing researches based on multi-agents use them directly as processors so that their concerted actions in a given context (inputs) directly gives a morphological result. This biocomputing approach is very interesting for solving optimization and organization based problems. This is for example what is done in collective robotics [17] or internet routing problems [15]. Another manner to proceed is indirect : this time, the agents are used to construct a – very reconfigurable – processor architecture that is used afterwards for computing classical calculi (and whatever problem a classical electronic processor can compute). As for a neural network, such a trail system based network would then be able to learn a configuration that, stimulated by some signal of different nature (*e.g.* light, electricity ...), would process it such as it corresponds to the expected computation. This would be facilitated and enhanced by using TS-agents of fibrillar nature, i.e. the orientation of cytoskeletal supramolecular assemblies under the action of electric or magnetic fields [18, 19, 20, 21].

The properties of trail systems allow us to better identify them to structurally programmable processors compared to classical reaction-diffusion systems (*e.g.* see [22] or [23] for BZ-based computing which needs either a strong control of the BZ reaction by local light inhibiting stimuli or by using a fixed geometry for the gel in which the reaction is realized), particularly in the case of fibrillar trail systems. We propose here to browse through those properties and computational interests that depend on the nature and the scale of those systems, and on the influence of their environment. In this article we also present the first step of a research agenda, i.e. a model based on an analogy between individual trail systems and wires or connections of an electronic processor or a neural processor. The model is an automaton very similar on its form to a Turing machine where the TS-agent is programmed to release traces of its trajectory in the environment, and for using them, but subject to intrinsic error or stochasticity coming from the environment. In that very simplified model of trail system, we identify clearly all sources of stochasticity that can affect the behavior of TS-agents, thus allowing us to study the robustness of their spatial self-organizing ability due to their microscopic –individual – properties. We also discuss of what ensures the robustness of self-organization at the level of a population. In particular, we are interested in the manner information is shared between TS-agents, and discuss on how such transfer entropy can be measured in our model or how it could be done is real systems. Finally, we describe how a learning process similar to that used for configuring formal neural networks will allow to obtain well structurally programmed trail-system based processors. It is interesting to note that an evolutionary process unifies the couple engineer-processor in a feedback loop : the engineer (that is also able to use feed forward decisions), creates a processing chip and tests it until it realizes the expected task or calculus. Evolution played the role of the engineer in natural systems (this point has been largely highlighted in several papers [24, 25]). Actually, the properties of the natural trail systems (as those cited above) have been selected – in a Darwinian sense – so as the robustness of the self-organization depending function is ensured in respect to the features of the natural environment.

2 Characteristics of the trail systems

Trail systems are constituted of numerous active elements in interaction in their environment. These elements are composed of two distinct parts: an active agent (TS-agent) and a trail.

The TS-agent is an object perceived as a unique physical entity. It has its own rules of behavior: internal rules that describe the changes intrinsic to this entity (independently to external actions), and rules that allow it interacting with other TS-agents, with its surrounding environment or with external fields.

Compared to the TS-agents, the trails are not precisely defined and are not really limited physically until a molecular description. They are made of local variations of the environment. Those variations have a maximal intensity at their sources, the TS-agents, and spread out or are progressively degraded from the instant they are generated. Their description is very related to level of abstraction used for describing the matter that constitutes the environment: a discrete manner or a continuous one. Viewed from a macroscopic point of view, one can mark a trail out with a certain accuracy, but the precision of their frontier is limited by the sensitivity of the experimentalist to the variations observed. Moreover when the scene is magnified, the observer can distinguish the different components of the trail, showing that the trail is not in the form of a unique coherent entity. Nevertheless the latter can drive the self-organization of the agents at the macroscopic level.

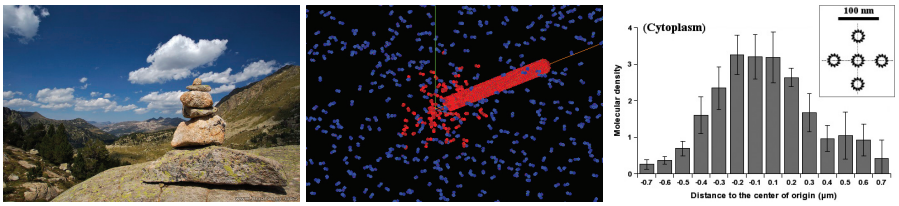


Figure 1: **Traces let by natural trail systems.** (**Left**) A cairn in the pyrénées in France (used by permission of P.-H. Muller, http://www.boreally.org/). One cairn constitutes an isolated element of the human trail signal clearly visible by humans over the landscapes. (**Center**) Model of disassembling microtubule (see [26]). Here, the microtubule shrinks at the unrealistic rate of 2000 $\mu m.min^{-1}$ so as we are able to observe the formation of a tubulin-GDP trail (red molecules; blue ones are tubulin-GTP). At realistic rates (¡20 $\mu m.min^{-1}$) no anisotropic trail forms and the heterogeneity observed is very weak as shown on the diagram to the right. (**Right**) Measurement from a simulation of the amount of tubulin-GDP released by disassembling microtubules in solution. Data correspond to the density profile of tubulin-GDP around the tips of disassembling microtubules, measured from the center of an array of 5 motionless microtubules (see the inset showing a transverse cross section of the x axis and of 5 MTs), each of them respectively separated by 30 nm (one microtubular diameter). All microtubules disassemble simultaneously at 20 $\mu m.min^{-1}$ (1.85 $ms.heterodimer^{-1}$) which is a quite fast disassembling rate. The macroscopic diffusion rate of individual tubulin dimers corresponds to that measured in the cytoplasm (5.9 10^{-12} $m^2.s^{-1}$)[27]. The quantity of tubulin-GDP molecules liberated is very low and needs to be integrated in time for obtaining average profiles. The graphic has been reconstructed by integration of the density maps of 6 independent simulations, during 1.8 ms (*i.e.* the average time separating the liberation of 2 tubulin-GDP molecules by a disassembling microtubule), between the simulation times 9.2 ms and 11 ms, along the 3 axis (a total of 6642 profiles)

The trails can be of various nature (chemical traces, physical tracks, hydrodynamic trajectories...) and have different effects (causing activating or inhibitory behaviors). Animals like social insects [28, 16, 29], marine snails [30], birds or flying drones [31], or human pedestrians [32, 33] or mountaineers (fig. 1 left) take advantage of this process for self-ordering and optimizing tasks, particularly path finding, hunting or foraging, thus saving energy. Other, such as harbor seals (for fishing activities)[34], 'know' how to use the trails produced by other systems for profiting from the same properties. All are macroscopic very efficient trail systems. Their efficiency is important in the sense that this mode of communication has a real influence on the synchronizing activity or the self-organizing behavior of the TS-agents that compose the system. Microscopic ones exist such as chemotactic cells [35] or bacteria such as actin comet systems (*e.g.* actin comets produced by TS-agents such as the *Listeria* or *Shigella* bacteria, or *Arp2/3*-coated latex beads [36, 37]), but also molecular ones such as the self-assembled biological fibers (*e.g.* microtubules or actin filaments) (fig. 1 right) or artificial DNA-designed programmable nanotubes [38] and carbon nanotubes [39]. In simulated systems such as the Conway's game of life or the Langton's ants [40], they are often called 'puffers', i.e. self-maintained gliders that produce persistent trails of numerical nature (*e.g.* a trajectory composed of cellular automata cells filled by active states).

For example a chemical trail, such as those made of pheromones that drive ant colony dynamics, is composed of numerous individual molecules that one could distinguish individually by a accurate observation at the 'nanoscopic' level. The same observed without or only at low amplification would be viewed as a macroscopic object, correctly defined along a certain area emerging from the source, and becoming progressively fuzzy at its undefined border. The quality of the frontiers observed depends on what the observer can perceive, as to say on its sensitivity to the signal that constitutes the trail. If the molecules that compose the trail are stained by a fluorescent marker, it will depend on the sensitivity of the experimentalist's camera to the light emitted. However evident it may be, that remark is the same for the TS-agent which is sensible to the trail signal only beyond a certain threshold.

Actually, numerous factors condition the efficiency of the trail-driven interactions between TS-agents, and thereby that of the resulting self-organizing processes. Considered aside, the system has a certain efficiency in processing information and self-ordering that can be inferred from the parameters enumerated below. What is the most important is nevertheless that the information generated and processed by such TS-agent – trail elements must be discriminated from the other interactions and processes (*e.g.* mechanical interactions, noise and thermal agitation, convection or flows, other concurrent reactions ...) that occur in the same time in their direct neighborhood. The importance of

a hierarchy in information processing systems and in particular in natural systems has been largely described by Conrad [2] : all interactions are processing events but all occur at different time and space scales, some of them being negligible in comparison to others that exist at the time and space scale level we are interested in. In our case, we are interested in trail system based processing and not in the possible different processing modes that could exist at a molecular level between some of the molecules or atoms that compose the system (as it is used in the hypothesis of information processing inside microtubules or actin [8, 9, 10, 11, 12, 13, 14]). An enumeration of the common parameters sufficient to describe whatever trail system and necessary for trail-based processing is given below.

***TS-agents*:**

- **Panel of actions (rules) available.** All TS-agents are random walkers since they generally can't 'see' directly the trail they follow (except perhaps in the particular case of animal pathways that can be directly viewed by the TS-agent before it encounters physically the trail). Once a trail is encountered, the internal state of the TS-agent is modified. Its internal rules take this new state into account so as the individual behavior of the TS-agent is now modified. In particular those changes induce an 'active searching' of the trail pathway by the TS-agent. Such 'intelligent' actively searching TS-agent are considered usually as active walkers but that notion is very imprecise: we can perfectly understand this meaning in the case of animals for example. It is more fuzzy in the case of cells or bacteria, and worth for pure molecular systems like biological fibers. The moving biological fibers encounter on their trajectory variations of composition and concentration in the medium, formed by the accumulation of numerous fuzzy, weak and extended trails (in this case, one may better say 'molecular clouds' instead of 'trails'). This affects the reactivity at their ends. When growing in favorable regions of the solution their growth is statistically enhanced and, on the contrary, when encountering unfavorable chemical compositions they can start to shrink or pause. They however can't reorient for following chemical trails as ants do. Nevertheless, a progressive process of selection – by a succession of growth, shrinkage and nucleation of new fibers – can lead to the selection of preferential orientation of the fibers. Despite a clear difference of efficiency, this process can be also assimilated to an active searching when considered from an upper level of observation.

- **Sensitivity to the trail.** The most important point is the efficiency of the TS-agent – trail recognition (or the attention of the TS-agent to the signals let by others in the environment). As evoked before, the TS-

agent – an animal, a cell, a supramolecular self-assembly – reacts to the signal contained in the trail structure only beyond a certain threshold. All TS-agents possess a sensing zone or a preferential region for physical interactions with the environment or for chemical ligand dependent reactions that can be assimilated to sensors (*e.g.* sense organs for animals, chemical receptors on a cell, reacting ends of biological fibers). The threshold is a combination of both the surface of recognition (reactive surface of the agent), the density of the sensing elements present on this surface (*e.g.* individual molecular receptors) and their individual affinity or sensitivity. Actually, it is the size of the surface of recognition on the TS-agent compared to the size of the trails encountered that is important. For better efficiency, and at least a better directional symmetry breaking, the trail must be thiner or of the same order of this region (*i.e.* of the TS-agent if the region is more or less confounded with the entire individual). For example, let us consider two cases: (i) the ant and (ii) the microtubular systems. In (i) the trail is thin and long, and the TS-agents are approximately as large as the width of the trail (about 1-5 mm). Such an anisotropy leads to a strong local symmetry breaking along the trajectory of the TS-agents and although the ants walk randomly, they are permanently biased in a preferential direction, often crossing the trail they follow and aiming to come back to it. This behavior is not unreminiscent to what occurs in human pathways and 'cairns' where those traces are very visible by the mountaineers, often over large distances, that way drawing kind of discrete tracks (fig. 1 left). On the contrary, in (ii) the trail produced by a microtubule is very extended (over microns) much more larger than the size of the 'sensor' of the fiber (30 nanometers) (see [26] and fig. 1 center & right). In that condition, this actions of the agent can not be biased in a preferential direction since the agent is included in the trail and can not cross it. The only possible effect of microtubule trails (and consequently of actin filament trails also) occurs at the level of huge populations of fibers at the millimeter scale, when macroscopic variations (compared to the size of the agents) of tubulin composition appear and cause massive synchronized reactions of the fibers, that are transmitted by diffusion. This is observed in solutions of microtubules in highly reactive conditions, as spatio-temporal variations (propagating waves) of the concentration of microtubules [41].

- **Moving rate of the TS-agent.** The motion of an agent has two origins: a motion that it controls due to its own dynamics, and a motion induced by external factors such as external fields, flows and diffusive motion. Their displacement relative to the environment will affect the manner the TS-

agent will sense the fluctuations of the medium. It depends actually on the ratio between the rate (efficiency) of recognition of the fluctuations by the TS-agent, and its moving rate. Moreover, the effect of the external factors on the TS-agents will strongly condition the success of specific trail-based processes, particularly when they are of the same order of the characteristic motion of the TS-agents. If animals are very weakly sensible to normal external factors (ex: wind) that's different for cells subject to flow or convective motions, and worth for biological fibers subject to strong mechanical effects and thermal agitation.

- **Internal memory.** When TS-agents evolve in their environment they are stimulated by the trail signals and other external factors. This modifies their internal state at least for a certain duration after stimulation. Animals have their neurons activated for an active searching of trails; particular pathways of the metabolic-genetic network of the chemotactic cells or bacteria are activated; the biological fibers store a part of the history of their reactive events in their composition and structure (*e.g.* There's a certain inertia in their dynamics, particularly for microtubules, stored in the form of chemical and conformational states at their reacting ends). In addition, some of them possess a sort of stack that memorizes a part of their trajectory. There are of course neuron-based memories, but also molecular assembly-based memories. Biological fibers and actin comets store chemical information in their fibrillar structure and restore it to the medium when disassembling in the form of the trail.

***trails*:**

- **Moving rate of the source.** The source is the TS-agent. Its characteristic moving rate condition partially the shape of the trail.

- **Persistence, diffusion and degradation.** The moving rate of the source has to be compared to the factors that tend to degrade and spread out the trail, such as molecular diffusion. Ants move quick compared to the diffusion of pheromones, thus leading to persistent trails. On the contrary, the trails formed by small molecules such as tubulin or actin diffuse very quickly compared to the growing or the shrinking rate of the microtubules or the actin filaments. In this case, the trails are not elongated as comets behind their source, but symmetric, diffusing all around their source.

- *Intensity (at the source).* The examples of trail systems cited above produce trails of various shapes but also of various intensity. As social insects release high concentrated trails of pheromones, biological fibers

on the contrary release individually very low amounts of protein bricks during disassembly. In addition to the fact their trails spread rapidly and far away from the source, they can not be considered as good 'molecular ants'. However, populations of N grouped and aligned fibers as microtubule bundles and most of all actin comet systems can release subunit amounts N times more important. Viewed from the TS-agents, this feature corresponds to its level of persuasion or to its efficiency of communication to other TS-agents.

It is then evidence that the relative scale and intensity of the agent-trail related processes compared to that of the interfering processes is critical as well as the relative size of the agents compared to that of the trail signals. The table below (Table. 1) gives different examples of trail systems observed in nature or of artificial (hardware or simulated) ones.

Nature	**O(*N*)**	**O(*Scale*)** (in O(m))			**O(*Rt*)** m/s	**Shape** ., o, -, –, —		**Efficiency** (– to ++)		
		Agent	Trail	System		Agent	Trail	Sig.	Ns.	Mn.
Pedestrians	1 to 4	0	1 to 2	1 to 2	0	.	—	++	- -	++
Drones	0 to 2	-1 to 1	0 to 1	1	1 to 2	.	-	++	-	++
Birds	0 to 2	-1	0	0 to 2	1	.	-	++	-	++
Fishes	1 to 4	-1 to 0	0 to 1	0 to 2	0 to 1	.	–	++	-	++
Snails	0 to 1	-2	-1 to 0	-1	-3	.	—	++	-	++
Ants	2 to 6	-3	-1	-1 to 3	-3	.	—	++	-	++
BZ wave	1 to 2	-4 to -3	-3	-4	-4	.	-	++	+	+
Lymphocytes	1 to 6	-5	-4 to 0	-3 to 0	-5	.	o to -	+	+	-
Dictyostelium	2 to 6	-6	-4 to -1	-2	-6	.	o to -	+	+	-
Bacteria	3 to 9	-6	-5 to -1	-3	-6	.	o to -	+	+	-
Actin comets	1 to 2	-6	-5	-5	- 7	. to –	–	+	+	+
Microtubules	3 to 9	-8	-6	-3	- 7	—	o	- -	++	- -
Actin F	3 to 9	-9	-6	-2	- 7	—	o	- -	++	- -
Carbon NT	> 6	-9	(?)	-6	(?)	–	o	- -	++	- -

Table 1: The table indicates the nature and the number of elements in a typical system. Semi-quantitative values (orders of magnitude expressed as exponents of ten) are given for their scales (agents, trails and the whole collective system) and the characteristic moving rates of the elements. Their shapes (agent and trail) can be assimilated as points '.', round symmetric spread areas 'o' , or fibers weakly to strongly orientable ('-' to '—')). An arbitrary notation between '–' and '++' characterizes the intensity of the trail signals. It is compared with the relative importance of the surrounding processes that interfere and perturb their behavior, thus giving an idea of the concurrence of the trail generation and trail degradation/spreading dynamics.

The table shows that very small and highly dynamic energy dependent elements behave potentially as trail systems: from the smaller to the bigger ones, actin filaments, microtubules, and actin comet systems. Due to the fact they are very small and numerous in a tiny volume of solution, and because they are fiber-shaped and consequently sensible to the action of weak external orienting fields like magnetic fields [18, 20, 21], they appear as excellent candidates for

realizing programmable chemical processors based on this principle. Unfortunately, the table also shows that their scale is very close or the same to the microscopic level where intense molecular diffusion and transports of matter rapidly delete all useful information.

Actin comet systems such as bacteria-actin comets or those formed by latex coated beads are probably the most efficient trail systems at the microscopic scale because they produce highly concentrated, thin (order of 1 - 3 μm) and long (10 - 20 μm) trails, because the size of their agent (about 2 μm) is comparable to the width of the trails, and because on the contrary to single biological fibers, they are able to reorient rapidly in the preferential direction of the trails followed by reorganizing their fibrillar network. Movies of actin-beads or actin-bacteria comets realized by several teams perfectly show this behavior. Regrettably, the interactions between actin comet agents and trails still have not been studied. It would however be of great interest.

3 Modeling a single TS-agent

As mentioned in the introduction, in its deterministic form, our model uses the formalism of a Turing machine. Such an automaton is composed of a tape that contains informations and a head that reads the informations of the tape at its current position, that applies internal rules (or transitions) such as moving or changing its current state, and that can write new informations on the tape, all of this conditioned by the set of instructions defined by the rules. Each Turing machine can be described as a sextuplet ($Q, \Sigma, \Gamma, E, q_0, F, \#,$), where Q is the finite set of control states $\{q_0, .., q_n\}$, Σ is the entry alphabet used for writing the values on the tape (it does not include the blank character $\#$ that separate to values or instructions), Γ is the alphabet of the tape and includes Σ and $\#$, q_0 ($q_0 \in Q$) is the initial internal state of the Turing machine, F is the set of final states that can reach the Turing machine, and E is a finite set of transitions (or rules) written in the form of a quintuplet of symbols $\{c, r, n, w, m\}$. The transition parameters are the following: the current state of the Turing machine c, the current read state r, the new state of the Turing machine n, the new state to write on the tape w, and a move m. Such rules can also be noted $c, r \longrightarrow n, w, m$ which means that if the Turing machine has the internal state c and reads r, its internal states changes to n, it replaces the symbol r by the symbol w in the tape and moves as indicated by m. All these symbols can be expressed into a unary, binary (or extended binary) or more compressed and symbolic (*e.g.* hexadecimal ...) codes. If it is of no importance for a Turing machine, we will see that the question of the encoding and the compression is really determining in the case of autonomous agents subject to stochasticity. For realizing any operation on values, the information tape has to contain at

least these numerical or symbolic values, but can also contain the rules (sets of instructions encoded within the same tape) necessary for realizing this operation. In the latter case, the Turing machine is called universal. All values or instructions are delimited between specific separators (or blanks) also encoded in a certain format. In our case, we are not interested in universal Turing machines since we want our agent to be very distinct to the environment.

Here, in our model, we will not use a complicated formalism. We will limit the notation of the values in the tape (Σ) to the unary format where decimal number (d) 1 is the unary (u) 1, d2 = u11, d3 = u111 ..., so as the set of decimal values $\{0, 1, 2, 3...\}$ corresponds to the unary values $\{0, 1, 11, 111...\}$. We differentiate the decimal number d0 (value u0) to the spacers used between these values by using the character star as a blank symbol ($\# = \{*\}$). The tapes and the sets of transitions are given separately (we are not in the case of universal Turing machines). The moves are elements of $\{S, L, R, P, T\}$ for respectively Stay, go Left and go Right, Pause, and the stopping signal T (as Terminated), and are of length 0 (for stay, pause or stop) to 1 step relative to their current position. In the model, we make a distinction between Stay and Pause. Stay implies that the state of the agent has changed during this step so as the agent is not stopped. This move instruction can exist in both deterministic or stochastic systems. The move instruction Pause does not concern deterministic systems. Real systems can maintain themselves in a stationary state during a certain time (depending on their available amount of energy and on their energy consumption rate) when they are blocked. For example, we can immobilize an ant during a certain – short – time and it continues to live. Once freed, the ant can continue to move. Although it does not include the notion of energy, the instruction Pause mimics this possibility. In our model, only an stochastic change in the machine state, in the environment or of decision (transition) can unjam the agent. Another analogy with a living agent is when the agent faces to an impossibility to decide what to do because of an unknown situation. A certain time pass until the agent has an idea of how to do. This corresponds in our model to a stochastic change of the current used instruction.

The agents will always start at the extreme left of the tape and end their computation at the position of the last rightest symbol of the result with a stopping signal. Their internal states are represented by the symbols of the latin lowercase alphabet $\{a, b, c, ..., z\}$. All agents begin with the initial state a.

Below is shown, as an example, a simple Turing machine (highlighted in gray) that computes N plus one (here $3 + 1$).

Transitions:
(1) $\mathbf{0}, \mathbf{a} \longrightarrow \mathbf{1}, \mathbf{a}, \mathbf{T}$;
(2) $\mathbf{1}, \mathbf{a} \longrightarrow \mathbf{1}, \mathbf{b}, \mathbf{R}$;
(3) $*, \mathbf{a} \longrightarrow *, \mathbf{a}, \mathbf{R}$;

(4) $\mathbf{1}, \mathbf{b} \longrightarrow \mathbf{1}, \mathbf{b}, \mathbf{R}$;
(5) $*, \mathbf{b} \longrightarrow \mathbf{1}, \mathbf{b}, \mathbf{T}$;

Tape before computation: * ****111***** $3 + 1$
Tape after computation : *****111** **1** ** $= 4$

In our case, we want the agent to behave as a TS-agent would do. Our agent realizes a more interesting simple task, i.e. transporting values from one part of the tape to another. This mimics the manner ants transport food or materials from one part of their environment to another (for example, [28] show the aggregation dynamics of dead ants transported by living ants that release preferentially their loads when encountering an existing accumulation of ant corpses). The most simple TS-agent that simulates a deterministic – and very simplified – ant in a 1D environment can be written as described below:

Blanks and zeros are not equivalent. We aim to represent an environment that contains (1) or not (0) values that correspond to amounts of matter (*e.g.* food or dead ants), but we also want to include the notion of explored or unexplored environment. When explored by TS-agents the environment is modified and shows specific paths (trails) that other TS-agents or the same TS-agents that created these paths can use preferentially. In the following, the environment is first considered as unexplored and is filled by blank characters $*$ except in two areas that correspond to a source (*e.g.* the food) and a destination (*e.g.* the nest) that are heterogeneities in the environment susceptible to be sensed and modified by an agent. Both source and destination areas can initially contain quantities of matter equal or more than 0. The TS-agent is conceived to transport all the N symbols 1 from the right heap (source), if its initial value is more than 0, by copying them to the left heap (the destination which contains 0 (e.g empty nest) to M symbols 1 at the beginning). Firstly, the TS-agent is looking for a food source by traveling through its nest and then through the unexplored area that separate the destination and the source areas. Then, when it finds some food, it comes back to the nest and let behind it a 'trail' of 0. The character 0 represents here (in this very simplified model) both amount values equal to 0 and the presence of pheromones. Then, it adds the carried value to the heap and goes again to the food source, this time by 'following' the traces of pheromones (succession of characters 0) between the two heaps. For doing that it needs the use of specific rules. At the end, the source disappears completely. This correspond to a calculus equivalent to $N + M$. Here $N = 8$ and $M = 1$. Of course much more simple models of Turing machines could do the same (*i.e.* transporting values from one heap to the other), but we aimed to be able to do a certain analogy with a trail system (with the notion of trail released by a TS-agent).

As described in table 2, the model contains 25 programmed transitions and 2 others – number 24 and 25 – (highlighted in gray) that where automatically added by the program so as to complete the table of relations between the TS-agent states and the environment values. The completion of the table is very important as is authorizes simulating stochastic systems. In such systems, . The two added instructions force the TS-agent to pause. As said before, only an accident let the TS-agent move again. Roughly, transitions 0-2 and 4-5 are used for reaching the left heap and crossing it, transitions 10-11 and 15-17 are used for realizing the first travel of 1 symbol 1 from the right to the left, transitions 3, 6-7 and 12-14 are used for transporting the N-1 symbols 1 from the right to the left, and finally the last transitions (8-9, 18-23 and 26) are used for reaching the left heap for the last time and finally stopping. The transitions that are characteristic to a trail system are those that imply the release and the use of a pheromone trail as a guiding rail, i.e. transitions 3, 6-10, 12, 17-18 and 21.

	0	**1**	*****
a	$(0)\ a,0 \longrightarrow b,0,R$	$(1)\ a,1 \longrightarrow b,1,R$	$(2)\ a,* \longrightarrow a,*,R$
b	$(3)\ b,0 \longrightarrow c,0,R$	$(4)\ b,1 \longrightarrow b,1,R$	$(5)\ b,* \longrightarrow d,*,R$
c	$(6)\ c,0 \longrightarrow c,0,R$	$(7)\ c,1 \longrightarrow p,0,L$	$(8)\ c,* \longrightarrow x,*,L$
d	$(9)\ d,0 \longrightarrow x,*,L$	$(10)\ d,1 \longrightarrow q,0,L$	$(11)\ d,* \longrightarrow d,*,R$
p	$(12)\ p,0 \longrightarrow p,0,L$	$(13)\ p,1 \longrightarrow p,1,L$	$(14)\ p,* \longrightarrow b,1,R$
q	$(15)\ q,0 \longrightarrow b,1,R$	$(16)\ q,1 \longrightarrow p,1,L$	$(17)\ q,* \longrightarrow q,0,L$
x	$(18)\ x,0 \longrightarrow y,*,L$	$(19)\ x,1 \longrightarrow x,1,T$	$(20)\ x,* \longrightarrow x,*,L$
y	$(21)\ y,0 \longrightarrow y,*,L$	$(22)\ y,1 \longrightarrow y,1,T$	$(23)\ y,* \longrightarrow z,*,R$
z	$(24)\ z,0 \longrightarrow z,0,P$	$(25)\ z,1 \longrightarrow z,1,P$	$(26)\ z,* \longrightarrow z,0,T$

Table 2: **Transitions.** The table gives the possible actions of the TS-agent modelized here, depending on its current state (first column) and on the symbols read in the environment (first row). To a given TS-agent state and a symbol read each transition associates a new state, a symbol newly written at the current place of the TS-agent and a move.

3.1 Deterministic simulations

The principal steps of a deterministic simulation are given below in table 3. Is this simulation, the TS-agent transports 5 symbols 1 (at the right of the environment) to its nest (close to the left end of the environment) already containing a symbol 1.

Such a TS-agent behaves, in a simplified manner, as a deterministic foraging ant in a 1D environment (see fig. 2). In this example, if the right heap is absent, the deterministic TS-agent will never stop (and is invalid). The TS-

agent would explore permanently its environment to the right until infinity. One could imagine by using different rules to place the food at the left and the nest to the right. In this case, if the nest was absent, it would be as if a unique dead ant or a very limited source of food was present in a box with a living deterministic one: the carrier ant would never release its load (until it dies in its turn). The only way for such a deterministic TS-agent to release the carried material is to include stochasticity in its perception of the environment, in its rules, or in the environment itself. Of course, more efficient rules could also be used that would ensure the deterministic TS-agent to stop such a set rules that would count the number of steps during when nothing has happened but our model is interesting in the sense that it is very simple and because its robustness is only due to the fact that there is no noise or errors and everything needs to be well planned before, until the presence of both a nest and a food source. Below is shown a complete simulation.

Environment	**Transitions used**
* **1*****11111***	starts from there
*******1 * **11111***	2,2,2,1
*******1*** 1 1111***	5,11,11,11
*******1*** 0 1111***	10
*******1 0 0001111***	17,17,17
******* 0 100001111***	16
******1 1 00001111***	14
******11 0 0001111***	4
******11000 0 1111***	3,6,6
******11000 0 0111***	7
******1 1 00000111***	12,12,12,12
*****11 1 00000111***	13,13,14,4
...	round trips
11111 1 00000000*	...4,4
**111111 00000000 * **	...4,4,3,6,6,...,6
111111 0000000 0 *	8
11111 1 ********	18,21,21,...,21,22

Table 3: **Example of deterministic simulation.** In the left column of the table, we show the principal steps of the evolution of the environment due to the action of the TS-agent. The position of the TS-agent is highlighted in gray. It becomes a TS-agent when it starts to release a trail of 0 (rules 10 and 17) on the return to its nest once it finds food (right heap of symbols 1), and when it uses it to come back from the nest (left heap of symbols 1) to the food. In the right column of the table, we give the respective transitions used for obtaining the environments at each step.

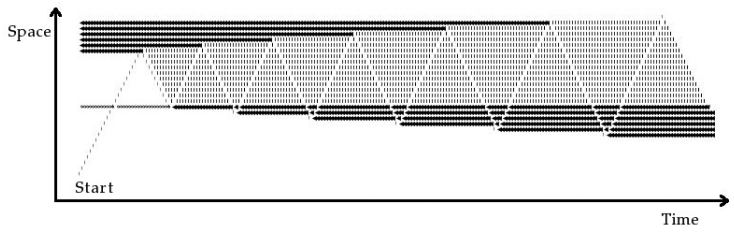


Figure 2: **Trajectory of a 1D-ant.** The TS-agent starts as indicated at the down left corner and moves vertically. The two heaps are shown in black: a big one (the food source which value = 6) at the top left corner and the smaller one (the nest which value = 0) below. They are separated initially by a blank space. Its trajectory during time corresponds to the zigzag that finishes to halt on the top of the nest. When moving, after it finds 'food', it releases a trail that appears in gray between the two heaps. At the end, the trail is removed by the TS-agent and the environment contains only one heap which value = 6, i.e. the sum of the two heaps.

Of course, this deterministic automaton realize its task perfectly and will serve as a reference. Let us consider now the introduction of stochasticity in both the environment and the behavior of the TS-agent.

3.2 Imperfect TS-agents in a changing environment

In real life, stochasticity is everywhere but living systems spend a lot of energy to counter entropy and maintain coherent self-organized forms and behaviors. Every living agent exists in a given environment. Both can be sources of errors that can affect the behavior of agents. A living agent makes 4 different things in its life: the agent can sense its environment, the agent can modify the environment, it moves through the environment, and finally it possesses its internal life, as to say it can think, make decisions, dream, change its internal states, etc, all of this independently to the environment. Mistakes of the agents can appear at these 4 levels. Problems of attention or sensing are related to the sensitivity of agents to their environment, e.g. our hearing or the sensitivity of a chemotactic cell (density and quality of the membrane receptors). Difficulties in the persuasion, talking or writing (communication) are due to a limited intensity of the signals emitted by agents towards other agents through the environment, compared with the ambient noise already present in the environment. Here, we distinguish between the errors of moving and the errors of decision. An agent can choose to move somewhere, i.e. to the left, and finally does something else like moving backwards or pausing. Stochastic changes of decision or, said differently, of use of behavioral rules always occur in our life when we feel uncertain about something to do before any choice of action (communicating or moving). Smaller living systems such as cells show most of the time coherent behaviors but sometimes, due to stochastic small molecular changes

(*e.g.* a small excess or lack at one moment of a transcription factor in a cell compared to another very similar cell will cause a small difference of behavior between them during a certain time. In this example, the brain that makes decisions is the cell internal machinery), they don't do what is expected. We are familiar to coherent behaviors of populations of cells, but when one look at all individual behaviors of each cells in the population, one find them not so coherent between each others.

In the environment, many events can also occur independently to living agents but all come down to transport of matter and transformation of matter. Transport of matter includes diffusion and flows but not the active transport of matter by living agents. Transformation of matter concerns chemical reactions, mutations, and disintegrations. All these phenomena are critical for the trails to be maintained. If weak trails are subject to a strong diffusion, the signal is rapidly diluted and the environment to become uniform again. On the contrary very intense trails produced by TS-agents in a noiseless environment will be maintained over a long time.

In our model, the representation of both the environment and the level of functioning of TS-agents are microscopic. At this level, one can consider that errors, noise and transport phenomena are rare events. Moreover, we consider distance matrices between the symbols of a given set (*i.e.* the set of rules, the set of moves, the set of TS-agent states, or the set of read symbols). Each distance value gives the relative proximity between two symbols (*e.g.* '99' is close to '100' but far from '3' as well as moving symbol 'L', 'R' and 'P' are very close together, but far from the halting signal 'T'). Then every event, whatever its type, behaves as a microscopic diffusion event in our model since the errors or mutations are just stochastic jumps from one given value (*e.g.* a TS-agent state, a given move or an environment symbol) to another of the same set. These jumps are only authorized as imposed and limited by the distance matrices. In addition, as in diffusion phenomena anisotropy can also be applied to the diffusion phenomena in our sets of symbols so as to force the system to evolve in one direction. As an example, one can consider a set of integer symbols comprised between 0 and 100 that represent the local number of molecules in a small volume. Due to chemical events called disintegration or mutation, these molecules will be degraded into others that are not perceived by our TS-agent; in other terms, for us, they disappear. Mutations events are then described as diffusion events of 1 step only from higher values to lower values of symbols.

Below is given the sequence of procedures followed during any simulation:

- **Environmental events**
  - Calculate the *mutations* (Diffusion at a given location within the set of environment symbols)
  - Calculate the *axial diffusion* along the 1D environment
  - Calculate the *lateral diffusion* between neighboring environments (in case of several parallel arrays of 1D environments)
- **Individual action and events of the TS-agent**
  - *Perception*: the TS-agent reads the environment. A possible error occurs during reading: the TS-agent reads something else, close to the good value as defined by the matrix of environmental symbols.
  - *Decision*: a rule is chosen given the environmental symbol read and the current state of the TS-agent. An error can occur that force the TS-agent to follow another transition close to the good one as indicated by the distance matrix of transitions. The state of the TS-agent takes a new value given by the transition that has been chosen.
  - *Communication*: the TS-agent writes into the environment the symbol given by the transition that has been chosen before. A possible error occurs during writing: the TS-agent writes something else, close to the good value as defined by the matrix of environmental symbols.
  - *Moving*: the TS-agent moves as described in the chosen rule. A possible error occurs during moving: the TS-agent moves in another direction or pauses or stops depending on the distance of the move states indicated by the matrix of move symbols.

At each time step, for one TS-agent, only one event of each type of event can occur. However, several events of mutation, and diffusion can occur in the environment during the same time because it is composed of several 'independent' volumes. Then, at each time step, a Poisson distribution law is used to determine the number of diffusion events of each type (axial diffusion, lateral diffusion and mutations).

Because errors can cause the TS-agent move out of the limits of the first defined environment, the 1D environment can extend in the two directions ore we can decide that its geometry is toric. When the environment extends, we generate a blank space and we fill it with the default symbols that characterize

the basic level of the environment (full of molecules or empty, or undefined...). During the simulation time many diffusion events and mutations occurred in the visible environment but many other should also have occurred in the undefined regions that we create for the TS-agent to continue moving. Because of this, when we extend the environment in a direction, we add a certain number (typically 5 or 10) of spatial elements, we determine the number of each type of diffusion events that should have occurred during all the simulation and we generate them.

As illustrated by figure 3, where we applied a mutation rate $\lambda_{Poisson} = 0.005$ (it is a quite small rate) on the environment with a strong anisotropy towards the smaller values ($0 \rightarrow 1 \rightarrow *$), some mutations occur and the trajectory of the TS-agent is punctuated by accidents. At the end, the result is not so different to the reference. Here, the resulting nest heap has the value 5 when the reference was 6. The execution of the simulations several thousand or million times will give an average result and its standard deviation. With these parameters, among 100000 simulations 90.2% had at least one error and we obtained an average result of 5.43 ± 1.76. Other comparisons could be made on other values such as the average Hamming distance between the final environments and the reference (here 0.092 ± 0.128), the computational time (number of steps) (here 220 ± 129 compared to the 225 steps of the reference), ... This constitutes a vector result that allows us measuring the efficiency of the considered trail system. This is a manner to quantify its robustness depending on its parameters.

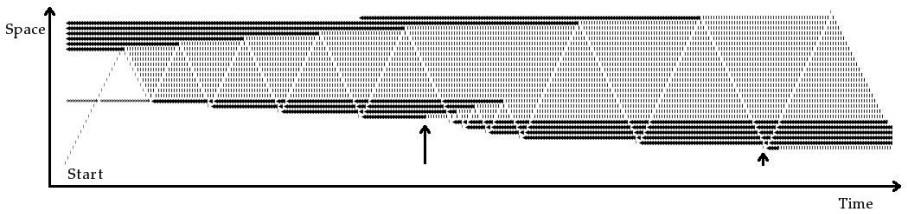


Figure 3: **Trajectory of a 1D-ant in a changing environment.** The simulations is identical to that described in figure 2 except that a mutation rate $\lambda_{Poisson} = 0.003$ is applied on the environment with a strong anisotropy towards the smaller values ($0 \rightarrow 1 \rightarrow *$). Only one mutation occurred as indicated by the arrow, and forced the TS-agent to follow a different trajectory during a certain time.

4 Increasing scales and degeneracy for robustness

For isolated TS-agents, in addition to the role of distance matrices that avoid dramatic changes when a perturbation occurs, two features ensure the robustness of TS-agent behaviors compared to a deterministic reference: the scale-depending accuracy of the agent in the perception of its environment, and the degeneracy of the values, internal states and rules in the system.

First, we mentioned that in the present model, TS-agents have a sight distance over the environment of only one element. Of course, we could decide that our TS-agents could look at several elementary regions of the environment at the same time and that the environment effectively perceived is an average value or a maximum, or values superiors to a certain threshold. In our model, one can consider the case of a TS-agent looking at 3 contiguous elements of the environment centered on the central one. If its actions need at least one value different to '0' or '*' among the set of values read, the agent will be very robust in an environment where the trail signals degrade rapidly (conversions of '1' into '0' and of '0' into '*'). The same phenomenon would have dramatic consequences on a TS-agent looking at only one element of space. The change of only one '1' into a '0' in that environment would change completely its computational trajectory. As evident they are, these remarks take sense when we think at biological systems such as cells that often perceive different regions of space at the same time due to a certain distribution of receptors on their surface. They are macroscopic agents that sense microscopic signals (individual molecules) with microscopic receptors. Microtubules on the contrary are microscopic agents that can only perceive very small regions which size are comparable to the dimensions of their sensing ends. In the first case (robust agent), the agent can ignore the fluctuation that exist at a scale level largely small than the agent itself; i.e. even if some of its receptors are not activated because of the local absence of molecules, a cell moving in a environment concentrated in activating molecules possess numerous other receptors that will perceive the signal. The cell feels globally concentrations instead of individual molecules and due to the integration of the whole concentration signal, the cell will adopt a certain behavior. On the contrary, when the size of agents is very limited compared to the size of the elements that compose the trail signal and to the trail itself, the agent will be very sensible to microscopic fluctuation.

Multivaluation on the symbols or values that describe the environment (*e.g.* $*$, 0, 1, 2, ... 10^6 ...) can be used for representing environments in a more macroscopic (and compact) form where elements of space can contain this time more than one 'molecule'. The two environments compared in table 4 contain the same values but shown at two different zoom levels: a microscopical one at which the components are individualized (the environment contains

0 or 1 molecule in each element of space), whereas the elements of space of the more macroscopic one can contain up to 4 molecules. It can also be useful to represent environments that can contains several equivalent kinds of molecules that can be used by agents. An example is described in table 5. Of course, in

Microscopic view	...	0000	0011	1111	1011	1000	0000	...
Macroscopic view	...	0	2	4	3	1	0	...

Table 4: **Scaling of the perception of environments.** Both show the same but at two different zoom levels. Each macroscopic element of space corresponds to 4 microscopic elements.

the latter, the TS-agent has to know what to do when values up to 4 are encountered. This implies a degeneracy of their rules. For example, rule 7 of the model that corresponds for the agent to the action of picking a '1' and replacing it by a '0' – that means at the same time presence of pheromone and absence of food –, could be rewritten for a 6-valued environment (*, 0, 1, 2, 3, 4) as follows:

Rule 7.1: $c, 1 \longrightarrow p, 0, L$
Rule 7.2: $c, 2 \longrightarrow p, 1, L$
Rule 7.3: $c, 3 \longrightarrow p, 2, L$
Rule 7.4: $c, 4 \longrightarrow p, 3, L$

Finally, another manner to introduce robustness is to multiply the number of redundant internal states (*i.e.* that are related to the same function). Such an increase of similar rules prevents the system from errors of decision or of fluctuations of the internal state.

5 Populations of TS-agents

Until now, we focused on the microscopic behavior of individual TS-agents in their environment. Of course TS-agents become interesting when they participate to the collective behavior of a colony. TS-agents communicate between each others via signals released to and get from their environment. In our model, this functionality is present. Diffusion can occur between neighboring 1D-environments. In real systems indeed, exchange of information helps maintaining a certain coherence of the whole system. This point is very well described in the recent article of Lizier et al [42]. The question is to quantify the information transfer as a dependence of the future states of the receiving agent on the past states that were emitted by the source. As said in that article, very logically, the predictive information transfer is the "average information contained in the source about the next state of the destination that was not

Environment	λ	Result	Computational time (steps)	Hamming distance	% modified environments
A or B	0.0	6.0	225	0.0	0
A	0.003	5.79±1.11	220±42	0.035±0.078	57
B	0.003	6.009±0.93	222±35	0.024±0.067	57
A	0.006	5.17±2.04	229±209	0.13±0.16	96
B	0.006	6.03±1.88	221±132	0.10±0.22	96
A	0.009	4.0±2.63	0.27±0.25	321±495	99.9
B	0.009	5.82±2.96	259±338	0.27±0.57	99.9

Table 5: **Effect of redundancy in the environment.** We defined two very similar environments A and B. Environment A is as previously defined in the article. Environment B contains 2 symbols ('1' and '—') that are equivalent to symbol '1' and is initiated with symbols '—' instead of the '1' in environment A. The distance matrix for mutations is defined such as a mutation of '—' gives '1' that can give in its turn '0' due to another mutation. The rules of the corresponding agent are adapted so as the actions are equivalent when symbols '1' and '—' are encountered. Result data are given as a vector of 3 values: the resulting heap, the computational time, the Hamming distance between the resulting environment and its reference ($\lambda = 0$). They where obtained for each value of λ and for each environment by realizing 10^5 independent simulations. The average and standard deviations are given. We also indicate the percentage of modified environments among the 10^5 ones. Standard deviations are often large, but we still conserve well defined Gaussian distributions. Agent B is clearly more robust in its environment compared to agent A in environment A even if we only add a very low redundancy.

already contained in the destination's past." This can be easily measured in our model of trail system by labeling with the ID label (unique identification number) of each agent, the environmental symbols that they manipulate. For example, a pheromone trail released by TS-agent *N°1* will be marked as *ID-1* while that released by TS-agent *N°2* will be marked as *ID-2*. If TS-agent *N°1* uses the pheromones marked *ID-2* (or reciprocally) it will be considered as an information transfer and scored. Then, from this individual measurements we obtain an average measure called 'transfer entropy' that quantifies "the statistical coherence between systems evolving in time in a directional and dynamic manner".

In real systems, such measurements are not always so easy to realize. First, it is difficult to know if for a given signal, a receiving agent has been able to effectively increase its information. Concerning human trails, the pedestrians can be asked on what they perceived and on how much they found useful the informations (*e.g.* cairns) available in the environment. For other systems, in particular small biological systems (cells, biological assemblies ...) fluorescent staining experiments could be realized in which groups of TS-agents (or individual ones) would be stained with a fluorophore of a first type (*e.g.* in

green) and would emit trail signals of the same type, while other TS-agents would be marked by another sort of fluorophore (*e.g.* a red one). Matter exchanges could then be followed. As an example one could consider two actin comets one stained in red and the other in green. Both are constituted by actin monomers (red or green, depending on their origin). When disassembling, the red actin comet releases red stained actin monomers that can be assembled another time. If they are assembled into the green actin comet, an exchange of matter occurred between the red and the green comets. This means that the red monomers participated to the behavior of the green comet and modified it. Realistic simulations of real systems are also a good way to evaluate such transfer entropy by following matter exchanges (simulations shown in fig. 4 aims to follow the individual trajectory of tubulin-GDP proteins).

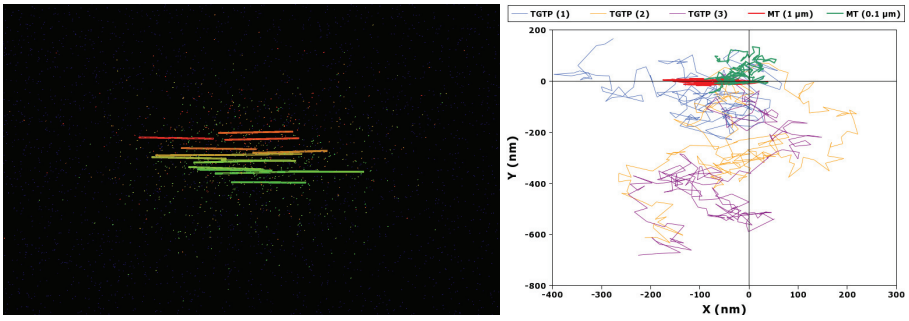


Figure 4: (**left**) On the left side, we show a simulation of microtubule disassembly, where 15 microtubules are marked by different colors. The tubulin-GDP they release is marked by the same color. This allows following the exchanges of matter between microtubules. (**right**) A detail is shown on the right side of the figure: A transversal projection of the 3D trajectory of 2 microtubules and 3 tubulin molecules is followed. As tubulin is quasi-spheric, its diffusion is isotropic. On the contrary, microtubules are rod like shaped and, depending on their size, their rotational and translational diffusion along their 3 axes is not equivalent, that way resulting in an anisotropic diffusion.

The study of the communication between trail systems and of how its efficiency is crucial for the TS-agents to self-organize locally, needs a separate paper. Nevertheless, we would like to mention two important points: the first one concerns the robustness of the process and its relations with the level of description of the environment, the second one the synchronism of agents.

- **(A)** As mentioned in the previous section, the same environment can be described at different scale levels. At the most microscopical one, each event of mutation or transport of matter appears clearly as a dramatic change. If our TS-agents work at such microscopic levels, then they will be very sensible to such changes. On the contrary, if their sensing

surface is extended to a more macroscopic area, small changes have no consequences because they are compensated by the presence of numerous other 'normal' events. Another manner to increase the robustness of the individual behaviors of agents (and to represent large views of the environment in a limited number of areas) is to use multivaluation of the symbols that describe the environment and the degeneracy of the transitions in whose these symbols will be involved. For example, one can decide that our TS-agent covers a surface that can contain up to 1000 molecules and that our TS-agent has three different behaviors depending on two thresholds: less than 10 molecules, more than 10 and less than 100 (a zone of transition), and more than 100 molecules. We will use a representation of the environment in which each subvolume contains up to 1000 values (or symbols) but our TS-agent will need only 3 rules (and not 1000) to know what to do in such an environment. Before, we considered only one TS-agent moving in its own 1D environment. The addition of lateral diffusion in our model (between the neighboring 1D environments) can cause dramatic accidents in the behavior of the agents, depending notably on that point.

- **(B)** The other criterion that will affect the local robustness of individual behaviors and the local information transfer is of temporal nature. The synchronism of the functioning of the agents (and the correlation of their activities), or on the contrary their asynchrony will affect considerably the information transfer during all their lifetime. If the agents are synchronous, like perfect robots, no information transfer will occur until the machines come back on their own steps. By lateral diffusion, the environment modified by TS-agent $N^\circ 1$ can diffuse to environment $N^\circ 2$ (and reciprocally), but even if it was the case, TS-agent $N^\circ 2$ would be yet at its next state and would move before sensing the change in environment $N^\circ 2$ (and reciprocally for TS-agent $N^\circ 1$). If they come back on their own trajectory, they will be able, that time, to sense the changes that occurred in their past. Such 'machines' are not very interesting because they do not communicate easily (and locally in a spatio-temporal sense). Moreover, agents in real systems never work synchronously, even when we think they do. There is always one agent in the populations that is more advanced in its – computational – trajectory and in its decisions than the others. One of our students tall us a very good example, that happened to him, which illustrate exactly that point: he was walking with a friend in our city of Grenoble. Both aimed to go to a precise point in the town, but none of them knew where it was, and both thought that the other knew. So they walked while talking to each other of other things, each of them being sure that the other was going in the

good direction. They walked during at least half and hour before asking them the question of yes or no the other knew. In a very theoretical system, if these two agents were synchronous, this situation could never occur. In reality, although they looked to walk together in a synchronous manner, their trajectory was decided because at each time, the decision to walk forward in a specific direction was made before by one of them, and the other followed this one. In our model, the individual action steps of the TS-agents are clocked (1 action per 1 time step), but for allowing asynchrony, we introduce random waiting steps. In consequence, if a population of individual was realizing the computation showed in figs. 2 & 3, with several environments linked together by a certain relation of neighborhood, the most advanced TS-agent (in its own environment) would modify its environment and this should affect the behavior of the late ones, especially if they are not robust to such changes (see point A).

6 Programming a bioprocessor whom fine architecture is made of trail systems self-organized structures

As in neural networks, 'programming' means 'learning' in such systems.
The conception of classical electronic-based computers took several dozen years to physicists and engineers. They started from very simple circuits that they tried to assemble logically so as to realized controlled calculii. They do not based their conceptions on any theoretical concepts such as Turing machines (a very interesting and puzzling history of the birth of computing machines was given by Burks [43, 44]. After that, a kind of evolutionary process occurred and is still active where engineers try new architectures and complexifications that are then tested successively by them, by benchmarkers and by the consumers (fig. 5 left). There are two types of selection processes: one is purely functional and the other is related to the current preferences of a society in terms of technology. Only the functional architectures persist, but among them, only the most adapted to the needs and preferences of the consuming society survive. Our electronic processors must be able to compute and they must be able to be programmed easily and used by consumers or programmers. That coupling between all protagonists and the process of selection is very long.

The same process is used in a much more shortest time for 'programming' neural networks. The term 'programming' is nevertheless not used for them because indeed we don't conceive that structuring a processor is a kind of programming. However it is ! It needs each time a complete restructuring of the network for obtaining a different processor. It is structural programming. Another time, an evolutionary process (or genetic-type algorithm) is

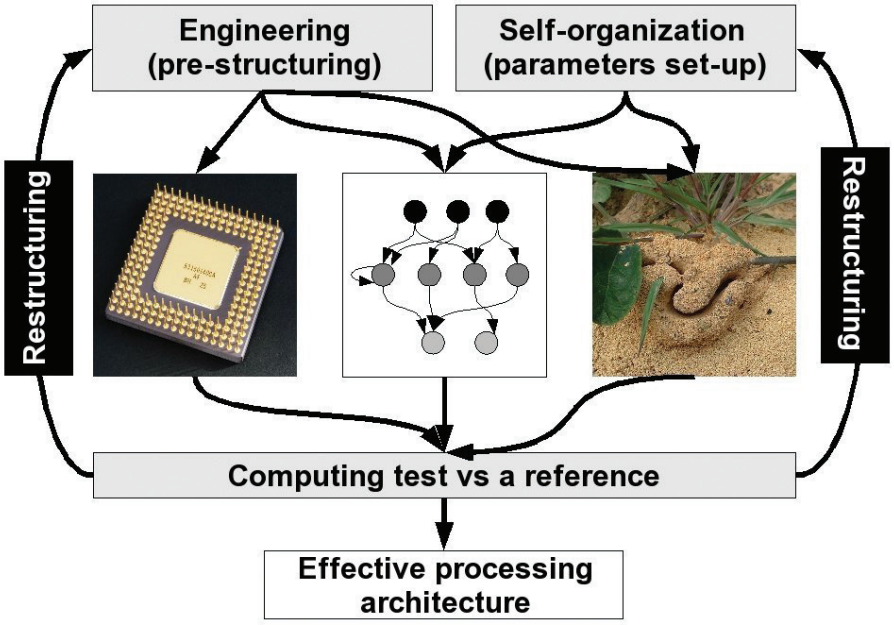


Figure 5: The process of creation of effective processing architectures is always a feedback loop of engineering design or set-up of the parameters implied in self-organization, then structuring of the architecture, and then test of the architecture for a given computation against a reference. Three architectures are shown: a classical electronic processor (**left**), a neural network (**center**) and a trail system (here an ant nest photographed in French Guyana) (**right**)

used: changes of parameters or of structuring (connectivity) are applied to the neural network. Given a set of inputs, it returns a certain output which is compared to the known result (a reference called 'professor'). The parameters are changed until its structure converges to a neuro-processor that effectively computes well the inputs into the expected result. This can be done by considering a population of 'neural networks' and by applying on them different changes. Only the best ones are selected (fig. 5 center). As mentioned before, Pfaffmann and Conrad [5] also applied a similar learning process to their model of self-organized microtubular processor.

Populations of self-organizing trail systems look very close to neural networks, except they are more labile. There are no reasons for not using a learning process on them so as to configure a trail system based processor (fig. 5 right). The only question concerns the nature of the signal injected in the bio-processor: it could be electricity (interfaced by an array of electrodes as for the neurons of Demarse's animats [45]) or light. Concerning the learning pro-

cess itself, three options are available: First one can imagine modifying the parameters that describe the individual behaviors of the agents. In this case, (i) a possibility is for the unconventional computer scientist to design an artificial trail system, for example by synthesizing self-assembling bricks with amino acid sequences or DNA/RNA sequences. Several teams over the world are now expert in realizing incredible structures made of DNA bricks [46, 47]. They are also able now to design quasi-programmable nanotubes [38, 48]. Although their assembly is not yet energy dependent (it occurs at thermodynamic equilibrium) the manner they assemble into tiled planes or tubes is decided by the 'programmer'. His work is to designs specific DNA structures that only can assemble in a certain manner. Once done, DNA nanotubes form. If they could work far from thermodynamic equilibrium like microtubules or actin filaments, such supramolecular assemblies could behave as TS-agents. (ii) The second option is to use an existing trail system and to modify chemically or physically the composition or structure of the TS-agents so as they behave differently. For example, one could use drugs that interact with the cytoskeleton. (iii) Finally, one could use directly an existing trail system and apply on it external factors that will bias their self-organization. For example, microtubules or actin filaments are very sensible to magnetic or electric fields (these fields create a torque that re-orient the fibers, depending on their size and on the density of reticulation of the solution of fibers) [18, 20, 21]. This effect can be strongly enhanced by using magnetic nanoparticle functionalized microtubules [19]. Other external factors such as flows, temperature gradients, vibrations, light (ex: UV light brakes the microtubules into several independent microtubules) can also be used.

In any case, these modifications on TS-agents or bias of the self-organizing conditions necessary to configure correctly the processing architecture has to be determined before engineering our systems. They could be determined by using models such as ours but in at least 2 dimensions for a better corresponding with real systems. We propose as a next step in our work to work with a set of self-organizing population of trail systems. On each population, genetic-type modifications of the parameters (for example addition, deletion, mutations of rules) will be applied and the population to self-organize until reaching a morphological steady state. The morphology constitutes a processor through which we send a – electric or luminous – signal (the set of inputs) via an interface. The returned result (*e.g.* a figure of diffraction or a set of spots of guided light if we use light) is compared to what we expect as a result. And the loop to be applied again and again until it is well structured.

7 Conclusion

The model presented in this paper represents a formalization of trail systems. It constitutes a good tool for studying the robustness of these natural systems not only for biological studies but also for biocomputing developments. In the latter, trail systems are viewed as bio-wires in a soft architecture: the bio-processor.

The development of a trail system based computer can be divided into 3 steps: first, we must understand how work the elements of the system. That is what we did here. Secondly, one must study how a population of elements behave i.e. self-organize in space and time. That study can be realized with – macroscopic or microscopic – natural systems such as chemotactic cells, actin comets and cytoskeleton fibers. Third, since we want to design a processor that can effectively compute something expected, we have to work on the control of its structuring and on the nature of the signals we will use as inputs and outputs, and to design artificial trail systems based on well-controlled elements such as the programmable nanotubes designed by Rothemund et al [38]. This approach has been successfully used for designing logical circuits with Belouzov-Zhabotinskii reactions [22, 23]. In our case, researching well organized processing architectures would correspond to applying a genetic algorithm to a population of TS-agents in a given environment (2D or 3D) and to select the architectures that provide a good correspondence between the output and its inputs. The fitness in this context is the expected result. In other terms, we are talking about learning in trail systems.

In the present article, we described several sources of errors. It is important to note that if the changes that affect the environment are definitive, this is not the case for the errors of the agent. They just temporary modify the deterministic choices and are applied by the program during the process. They do not modify the rule, symbol, state and move tables. Neither they modify the distance matrices. Of course, in parallel to that, an evolutionary process (as in the learning process evoked before) could be applied that would cause different modifications of the deterministic rule, move ... tables and matrices associated to the agents of a population. A heterogeneous population would form that should compute differently their environment, sometimes more efficiently or more robustly. The better agents would then be selected. In the same manner, evolutionary-based algorithms need to be applied on the populations of homogeneous or heterogeneous agents that would constitute a bio-processor so as they self-organize as expected into a given processing architecture [5]. When related to real systems, a learning process applied on the TS-agent parameters corresponds to designing artificial TS-agents. The second case (learning at the level of the population) corresponds to obtaining the expected control of either a natural trail system or an artificial one.

Such researches are often viewed as ambitious and not really profitable in the short term. Nevertheless, one can make analogies between information processing at different scale and time levels in processors and the manner of biological systems function [1, 2]. One can estimate how much information is processed by self-organized biological systems or by artificial systems inspired from nature (such as neural networks, collective robots ...) [2, 5, 42]. As the design of natural computers needs a good understanding of the manner their components behave, this field of research is a good way for increasing our knowledge on self-organizing biological systems, in particular to understand what ensures the coherence of biological systems in their environment.

Acknowledgements

This work was supported as an exploratory project of the Rhone-Alpine Complex System Institute IXXI. It also constitute one of the thematics being developed in the *Bioputing* group, part of the *Epigenomic project*, Genopole, Evry.

References

[1] M. Conrad. On design principles for a molecular computer. *Comm.ACM*, 28:464–480, 1985.

[2] M. Conrad. Scaling of efficiency in programmable and non-programmable systems. *Biosystems*, 35:161–166, 1995.

[3] E. Schroedinger. *Qu'est-ce que la vie ? (What is life ? The physical aspect of the living cell)*. Collection Points-Sciences, Le Seuil (1993), french edition edition, 1944.

[4] O. Bastien, J. C. Aude, S. Roy, and E. Marechal. Fundamentals of massive automatic pairwise alignments of protein sequences: theoretical significance of z-value statistics. *Bioinformatics*, 20:534–537, 2004.

[5] J. O. Pfaffmann and M. Conrad. Adaptative information processing in microtubule networks. *Biosystems*, 55:47–48, 2000.

[6] F. J. Nédélec, T. Surrey, A. C. Maggs, and S. Leibler. Self-organization of microtubules and motors. *Nature*, 389:305–308, 1997.

[7] F. Nédélec, T. Surrey, and E. Karsenti. Self-organisation and forces in the microtubule cytoskeleton. *Curr. Op. Cell. Biol.*, 15:118–124, 2003.

[8] S. R. Hameroff and R. C. Watt. Information processing in microtubules. *J. Theor. Biol.*, 98:549–561, 1982.

[9] S. Hameroff, A. Nip, M. Porter, and J. Tuszynski. Conduction pathways in microtubules, biological quantum computation, and consciousness. *Biosystems*, 64:149–168, 2002.

[10] S. R. Hameroff and J. A. Tuszynski. Search for quantum and classical modes of information processing in microtubules: Implications for the living state. In F. Musumeci and M. W. Ho, editors, *Bioenergetic Organization in Living Systems. Proceedings of the Conference: Energy and Information Transfer in Biological Systems, Acireale, Italy*, pages 31–62. World Scientific, Singapore, 2003.

[11] J. A. Tuszynski, J. A. Brown, and P. Hawrylak. Dieletric polarization, electrical conduction, information processing and quantum computation in microtubules. are they plausible ? *Phil. Trans. R. Soc. Lond. A*, 356:1897–1926, 1998.

[12] J. A. Tuszynski, S. Portet, J. M. Dixon, C. Luxford, and H. F. Cantiello. Ionic wave propagation along actin filaments. *Biophys. J.*, 86:1890–1903, 2004.

[13] D. D. Georgiev, S. N. Papaioanou, and J. F. Glazebrook. Neuronic system inside neurons: Molecular biology and biophysics of neuronal microtubules. *Biomed. Rev.*, 15:67–75, 2004.

[14] J. Faber, R. Portugal, and L. P. Rosa. Information processing in brain microtubules. *Biosystems*, 83:1–9, 2006.

[15] E. Bonabeau, M. Dorigo, and G. Theraulaz. Inspiration for optimization from social insect behaviour. *Nature*, 406:39–42, 2000.

[16] K. Johnson and L. F. Rossi. A mathematical and experimental study of ant foraging trail dynamics. *J. Theor. Biol.*, 241:360–369, 2006.

[17] M. J. Kriger, J. B. Billeter, and L. Keller. Ant-like task allocation and recruitment in cooperative robots. *Nature*, 406:992–995, 2000.

[18] P. M. Vassilev, R. T. Dronzine, M. P. Vassileva, and G. A. Georgiev. Parallel arrays of microtubles formed in electric and magnetic fields. *Biosci. Rep.*, 2:1025–1029, 1982.

[19] M. Platt, G. O. Muthukrishnan, W. Hancock, and M. E. Williams. Millimeter scale alignment of magnetic nanoparticle functionalized microtubules in magnetic fields. *J. Am. Chem. Soc.*, 127:15686–15687, 2005.

[20] N. Glade and J. Tabony. Brief exposure to magnetic fields determine microtubule self-organisation by reaction-diffusion processes. *Biophys. Chem.*, 115:29–35, 2005.

[21] Y. Liu, Y. Guo, J. M. Jr. Valles, and J. X. Tang. Microtubule bundling and nested buckling drive stripe formation in polymerizing tubulin solutions. *Proc. Natl. Acad. Sci. USA*, 103:10654–10659, 2006.

[22] B. De Lacy Costello and A. Adamatzky. Experimental implementation of collision-based gates in belousov–zhabotinsky medium. *Chaos, Solitons & Fractals*, 25:535–544, 2005.

[23] Y. Igarashi, J. Gorecki, and J. N. Gorecka. Chemical information processing devices constructed using a nonlinear medium with controlled excitability. *LNCS*, 4135:130–138, 2006.

[24] A. Lesne. Robustness: confronting lessons from physics and biology. *Biol. Rev.*, 83:509–532, 2008.

[25] A. Wagner. *Robustness and evolvability in living systems*. Princeton University Press, Princeton, New Jersey, first edition, 2005.

[26] N. Glade. Computing with the cytoskeleton : A problem of scale. *Int. J. Unconv. Comp.*, 4:33–44, 2008.

[27] E. D. Salmon, W. M. Saxton, R. J. Leslie, M. L. Karow, and J. R. McIntosh. Diffusion coefficient of fluorescein-labeled tubulin in the cytoplasm of embryonic cells of a sea urchin: Video image analysis of fluorescence redistribution after photobleaching. *J. Cell. Biol.*, 99:2157–2164, 1984.

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

[29] S. Jarau, C. M. Schulz, M. Hrncir, W. Francke, R. Zucchi, F. G. Barth, and M. Ayasse. Hexyl decanoate, the first trail pheromone compound identified in a stingless bee. *J. Chem. Ecol.*, 32:1555–1564, 2006.

[30] M. S. Davies and J. Blackwell. Energy saving through trail following in a marine snail. *Proc. Roy. Soc. B: Biol. Sci.*, 274:1233–1236, 2007.

[31] I. Kroo. Collectives and complex system design. *VKI Lecture Series on Optimization Methods & Tools for Multicriteria/Multidisciplinary Design*, 2004.

[32] D. Helbing, J. Keltsch, and P. Molnár. Modelling the evolution of human trail systems. *Nature*, 388:47–50, 1997.

[33] R. L. Goldstone and M. E. Roberts. Self-organized trail systems in groups of humans. *Complexity*, 11:43–50, 2006.

[34] N. Schulte-Pelkum, S. Wieskotten, W. Hanke, G. Dehnhardt, and B. Mauck. Tracking of biogenic hydrodynamic trails in harbour seals (*Phoca vitulina*). *The J. Exp. Biol.*, 210:781–787, 2007.

[35] A. Bagorda, V. A. Mihaylov, and C. A. Parent. Chemotaxis: Moving forward and holding on to the past. *Thromb. Haemost.*, 95:12–21, 2006.

[36] J. B. Alberts and G. M. Odell. In silico reconstitution of listeria propulsion exhibits nano-saltation. *PLOS Biol.*, 2:e412, 2004.

[37] H. Boukella, O. Campas, J. F. Joanny, J. Prost, and C. Sykes. Soft listeria: Actin-based propulsion of liquid drops. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 69:e061906, 2004.

[38] P. W. Rothemund, A. Ekani-Nkodo, N. Papadakis, A. Kumar, D. K. Fygenson, and E. Winfree. Design and characterization of programmable dna nanotubes. *J. Am. Chem. Soc.*, 126:16345–16352, 2004.

[39] A. M. Bonnot, M. N. Séméria, J. F. Boronat, T. Fournier, and L. Pontonnier. Investigation of the growth mechanisms and electron emission properties of carbon nanostructures prepared by hot-filament chemical vapour deposition. *Diam. Rel. Mat.*, 9:852–855, 2000.

[40] A. Gajardo, A. Moreira, and E. Goles. Complexity of langton's ant. *Discr. Appl. Math.*, 117:41–50, 2002.

[41] E. Mandelkow, E. M. Mandelkow, H. Hotani, B. Hess, and S. C. Muller. Spatial patterns from oscillating microtubules. *Science*, 246:1291–1293, 1989.

[42] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya. Local information transfer as a spatiotemporal filter for complex systems. *Phys. Rev. E*, 77:026110 1–11, 2008.

[43] A. W. Burks. Turing's theory of infinite computing machines (1936-1937) and its relation to the invention of finite electronic computers (1939-1949). In *Proceedings of the Fourth International Conference on Cellular Automata for Research and Industry: Theoretical and Practical Issues on Cellular Automata*, pages 179–197. Springer-Verlag, London, UK, 2000.

[44] A. W. Burks. The invention of the universal electronic computer: how the electronic computer revolution began. *Cellular automata*, 18(7):871–892, 2000.

[45] T. B. Demarse, D. A. Wagenaar, A. W. Blau, and S. M. Potter. The neurally controlled animat: Biological brains acting with simulated bodies. *Autonomous Robots*, 11:305–310, 2001.

[46] W. M. Shih, J. D. Quispe, and G. F. Joyce. A 1.7-kilobase single-stranded dna that folds into a nanoscale octahedron. *Nature*, 427:618–621, 2004.

[47] P. W. K. Rothemund. Folding dna to create nanoscale shapes and patterns. *Nature*, 440:297–302, 2006.

[48] D. Liu, S. H. Park, J. H. Reif, and T. H. LaBean. Dna nanotubes self-assembled from triple-crossover tiles as templates for conductive nanowires. *Proc. Natl. Acad. Sci. USA*, 101(3):717–722, 2004.

Spatial Information and Multivalued Genetic Regulatory Networks[1]

Matthieu Manceny[1,2], Marc Aiguier[1], Pascale Le Gall[1,2],
Joan Hérisson[2], Ivan Junier[2], François Képès[2]

[1] École Centrale Paris, MAS Laboratory, France
[2] Epigenomics Project, Genopole® and University of Evry, F-91000 Evry

Abstract

Modelling frameworks for biological networks are used to reason on the models and their properties. One of the main problems with such modelling frameworks is to determine the dynamics of gene regulatory networks (GRN). Recently, it has been observed in *in vivo* experiments and in genomic and transcriptomic studies, that spatial information is useful to better understand both the mechanisms and the dynamics of GRN. In this paper we propose to extend the modelling framework of R. Thomas in order to introduce such spatial information between genes, and we will show how these further informations allow us to restrict the number of dynamics to consider.

Keywords: Genetic Regulatory Networks, Spatial Information, Multivalued Dynamics, Discrete Mathematical Modelling.

1 Introduction

To understand Genetic Regulatory Networks (GRN), modelling frameworks and simulation techniques are often useful since the complexity of the interactions between constituents of the network (mainly genes and proteins) makes intuitive reasoning difficult. Most of the time, parameters of the model have to be inferred from a set of biological experiments. Formal methods, such as model checking or symbolic execution ([1, 12]), have been proved useful to determine values of parameters leading to valid dynamics of GRN, that is dynamics consistent with biological properties expressed using temporal logic. Nevertheless, these techniques are in practice difficult to manage because biological systems are either large, complex or incompletely known, resulting in a huge number of parameters to consider. Hence, in order to reduce this number, it seems relevant to embed within the model some biological knowledge such as spatial relation between genes.

Recent experiments have shown that both in eukaryotes [6] and in bacteria [2] gene transcription occurs in discrete foci where several RNA polymerases

[1]This work is performed within the European project GENNETEC (STREP 34952).

(the transcribing elements) are co-localized. This suggests that genes also tend to co-localize in space in order to optimize transcription rates. Such a scenario is supported by genomic and transcriptomic analysis [7, 3]. These have revealed that the genes which are regulated by a given transcription factor and the gene which codes for the transcription factor tend to be located periodically along the DNA [7]. In this way, the genes can be easily co-localized in the three-dimensional space according to a solenoidal structure of the DNA/chromatin, even in the presence of several kinds of transcription factors [8]. As a result, the effect of a transcription factor is enhanced due to the spatial proximity of the targets. This phenomenon is reminiscent of the local concentration effect that has been uncovered by Müller-Hill [13] a decade ago. Local concentration simply means that the interaction between molecules that are able to interact with each other is all the more efficient when molecules are close to each other. This straightforward statement is crucial to understand genome organization because genomes seem to have evolved in order to optimize the spatial proximity of reactive groups [8, 13, 9].

In this article, we propose to include spatial information into GRN and to study its effect upon the dynamics of the network. Our approach is based on the discrete modelling of GRN that has been introduced by René Thomas [14]. The spatial information concerns the gene proximity that results from a specific organization of DNA/chromatin. This proximity is modelled through two notions. The notion of *cluster* expresses the notion of co-regulation, that is a set of spatially closed genes that are expressed at the same time due to the expression of a single regulating gene (*i.e.* the presence of a single transcription factor). The notion of *privileged interaction* between genes is an ubiquitous concept in biology; for instance, specific interactions (e.g. between a transcription factor and DNA) in contrast to non-specific interactions, or local concentration phenomena are examples of privileged interactions. The use of privileged interaction is mainly based on the idea that if two interactions lead to contradictory effects, then the privileged interaction is preferred to the non privileged one.

This paper is an extension to multivalued dynamics of our previous work in [10] on Boolean dynamics. Main results of this work are recall, and we will see that whereas it is possible, in a Boolean approach, to determine constraints on the model of GRN to drastically reduce the number of dynamics to consider, this is usually not possible with a multivalued approach.

The paper is structured as follows. Section 2 presents our model of GRN including privileged interactions and clusters. In Section 3, we are interested in the multivalued dynamics of classical GRN. The dynamics is governed by a set of so called threshold and logical parameters, and we present how the structure of the GRN determines the possible values of these parameters.

Nevertheless, the possible dynamics still remain too numerous, and so, Section 4 presents how to use privileged interactions and clusters to reduce the number of dynamics to consider. Section 5 presents a illustrative example, and some numerical simulations. Finally, Section 6 gives some concluding remarks.

2 GRN with Privileged Interactions and Clusters (PCGRN)

Genetic Regulatory Networks are usually represented by an oriented graph, called *interaction graph*, whose nodes abstract the proteins or genes which play a role in the system and edges abstract the known interactions of the GRN. The model of this article is based on Multivalued GRN, that is GRN where gene have a finite set of *expression levels* which discretise their continuous concentration in the cell (see Section 3). An interaction ($a \to b$) can be either an activation or an inhibition: in an *activation*, the increase of the expression level of a leads to an increase of the expression level of b, the edge is labelled by the sign $+$ and a is an activator of b; in an *inhibition*, the increase of a leads to a decrease of b, the edge is labelled by the sign $-$ and a is an inhibitor of b. To this classic representation, we add the notion of *privileged interactions* as a subset of the interactions of the GRN. The notion of *clusters* defines groups of genes which are simultaneously activated or inhibited by a same gene.

Definition 1 PCGRN: GRN with privileged interactions and clusters
A genetic regulatory network with privileged interactions and clusters *(PCGRN) is a labelled directed graph* $G = (V, E, S, P, C)$ *where*

- (V, E, S) *is an* interaction graph *that is*
  - V *is a finite set whose elements are called* variables
  - $E \subseteq V \times V$ *is the set of* interactions
  - $S : E \to \{+, -\}$ *associates to each interaction its* sign *("+" for* activation *and "−" for* inhibition*)*
- $P \subseteq E$ *is the set of* privileged interactions
- C *represents the clusters of* G*, that is for each gene a partition of its target genes: for each* i *in* V*,* $C(i) = \{C_i^1, \ldots, C_i^{p_i}\}$ *where*
  - $\cup_{k=1}^{p_i} C_i^k = \{j | j \in V, (i, j) \in E\}$
  - *for all* k, k'*:* $k \neq k' \Rightarrow C_i^k \cap C_i^{k'} = \emptyset$

For any $i \in V$, $V^-(i)$ (resp. $V^+(i)$) denotes the set of predecessors (resp. successors) of i, that is elements of V which have an action on i (resp. on which i has an action): $V^-(i) = \{j | j \in V, (j, i) \in E\}$, $V^+(i) = \{j | j \in$

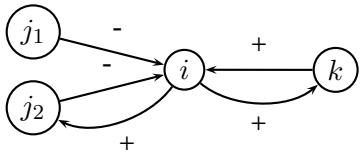


Figure 1: Example of interaction graph

$V, (i,j) \in E\}$; $P(i)$ denotes the set of privileged predecessors of i: $P(i) = \{j | j \in V^-(i), (j,i) \in P\}$.

Definition 2 (Activators and inhibitors) *Let (V, E, S, P, C) be a PCGRN, and let $i \in V$ be a gene. We denote by $A(i)$ (resp. $I(i)$) the set of* activators *(resp.* inhibitors*) of i: $A(i) = \{j | j \in V^-(i), S(j,i) = +\}$ and $I(i) = \{j | j \in V^-(i), S(j,i) = -\}$.*

In the following, a PCGRN will be represented as a graph where nodes are variables, arrows are interactions (dashed arrows for the privileged ones) and signs label arrows (see Fig. 3).

Example 1 (Interaction Graph) *Let us exemplify Definition 1 with the toy interaction graph (that is without any information on privileged interactions nor clusters) from Fig. 1 where a gene i is inhibited by j_1 and j_2 and activated by k, and activates genes j_1 and k.*

Section 3 will present the dynamics of classical interaction graphs (that is PCGRN without privileged interactions nor clusters); the influence of privileged interactions and clusters is presented in Section 4.

3 Multivalued Dynamics of Interaction Graphs

The *dynamics of an interaction graph* consists in the evolution of each gene expression level step by step. Several dynamics can be associated to an interaction graph, and the main problem is to reduce the number of dynamics we have to consider [1]. In reality, the evolution of a given gene's expression level does not depend on all the genes of the interaction graph, but only on the genes which have an action on the given gene, that is its predecessors. More precisely, not all the predecessors of a given gene have an effect on its expression level, but only the predecessors *with a sufficient expression level*, the interaction is then said to be *effective*.

3.1 Threshold Function and Multivalued Dynamic States

When a gene i acts on several targets, on j and k for example, it is often known that the level of i mandatory for an action on j to be is higher than the level

necessary for the action of i on k. This knowledge is modelled through the notion of *thresholds*.

Definition 3 (Thresholds function) *Let $G = (V, E, S, P, C)$ be a PCGRN. A threshold function $T_G : E \to \mathbb{N}^*$ associates to each interaction of a GRN its threshold* thresholds parameters. *T_G is such that such that*

$$\forall (i,j) \in E, T(i,j) \neq 1 \Leftrightarrow \exists k \in E : T(i,k) = T(i,j) - 1$$

In other word, if an interaction outgoing from a variable i is labelled by a threshold α greater than 2, then there exist interactions outgoing from i labelled by $1, \ldots, \alpha - 1$. This well represents the qualitative nature of thresholds in interaction graph, and an interaction (j, i) will be effective if and only if the expression level of j is above the threshold of (j, i). Obviously, several threshold parameters can be associated to a single interaction graph.

Example 2 (Threshold Functions) *In Fig. 1, because j_1, j_2 and k have only one successor, then the threshold of their unique outgoing interaction is* 1. *Because i has two successors, there are three possible threshold functions: $T^1 : (i,k) \mapsto 1, (i,j_2) \mapsto 2$; $T^2 : (i,k) \mapsto 2, (i,j_2) \mapsto 1$; and $T^3 : (i,k) \mapsto 1, (i,j_2) \mapsto 1$.*

In *multivalued dynamics*, genes can attain several levels, called *expression levels* which depend in both the interaction graph, and the associated threshold functions. Indeed, a gene can take as many values as the greatest outgoing threshold. The knowledge of the expression levels of all the genes define a *multivalued dynamic state*.

Definition 4 (Multivalued dynamic states) *Let $G = (V, E, S, P, C)$ be a PCGRN, and let T_G be an associated threshold function. We denote for all $i \in V$: $b_i = max\{T_G(i,j) | j \in V^+(i)\}$). The set of possible level of expression for a gene i is $\mathbb{X}_i(G, T_G) = \{0, 1, ..., b_i\}$.*

We denote[2] *by $\mathbb{X}(G, T_G)$ the set of multivalued dynamic states of G, associated to T_G: $\mathbb{X}(G, T_G) = \prod_{i \in V} \mathbb{X}_i(G, T_G)$.*

For $x = (x_1, ..., x_{|V|}) \in \mathbb{X}(G, T_G)$, x_i is the expression level *of gene i in x.*

Example 3 (Multivalued dynamic states) *In Fig. 1, because j_1, j_2 and k have only one successor, then they have only two expression levels. Because i has two successors, there are three possible threshold parameters T^1, T^2 and T^3 (see example 2) leading to either two expression levels for i (with T^3) or three expression levels (with T^1 or T^2).*

[2] Let us recall that $|V|$ denotes the number of elements in the set V.

3.2 Effective predecessors and Logical Parameters

The *dynamics of an interaction graph* consists in the evolution of each gene's expression level step by step. This evolution for a given gene does not depend on all the genes of the PGRN, but only on the genes which have an action on the given gene, that is its *effective predecessors*.

Definition 5 (Effective predecessors) *Let $G = (V, E, S, P, C)$ be a PCGRN, and let T_G be an associated threshold function. Let $i \in V$ be a gene and let $x \in \mathbb{X}(G, T_G)$ be a dynamic state. We denote by $A^*(i, x)$ (resp. $I^*(i, x)$, $w^*(i, x)$) the set of* effective activators *(resp.* effective inhibitors*,* effective predecessors*) of i in the state x:*

$$A^*(i,x) = \{j | j \in V^-(i), S(j,i) = +, x_j \geq T_G(j,i)\}$$

$$I^*(i,x) = \{j | j \in V^-(i), S(j,i) = -, x_j \geq T_G(j,i)\}$$

$$w^*(i,x) = A^*(i,x) \cup I^*(i,x)$$

Several dynamics can be associated to a given PGRN. These dynamics are described by a set of *logical parameters* which associates the future expression level of a given gene according to its effective predecessors.

Definition 6 (Logical parameters) *Let $G = (V, E, S, P, C)$ be a PCGRN, and let T_G be an associated threshold function. For $i \in V$, we denote by $K_i^{T_G} : 2^{V^-(i)} \to \{0, \ldots, b_i\}$ (with $b_i = max\{T_G(i,j) | j \in V^+(i)\}$) the set of* logical parameters *associated to i, considering T_G.*

For any i in V, if the system is in the dynamic state $x \in \mathbb{X}(G, T_G)$, then i's next expression level is given by $K_i^{T_G}(w^*(i, x))$.

Example 4 (Logical parameters) *In Fig. 1, gene i has three predecessors. Thus, there is 8 logical parameters K_i to consider for any T in T^1, T^2 or T^3: $K_i^T(\emptyset)$, $K_i^T(\{j_1\})$, $K_i^T(\{j_2\})$, $K_i^T(\{k\})$, $K_i^T(\{j_1, j_2\})$, $K_i^T(\{j_1, k\})$, $K_i^T(\{j_2, k\})$ and $K_i^T(\{j_1, j_2, k\})$. We also have to consider $K_{j_2}^T(\emptyset)$, $K_{j_2}^T(\{i\})$, $K_k^T(\emptyset)$ and $K_k^T(\{i\})$. Since j_1 has no predecessor, it remains stable anytime.*

Let us now consider a dynamic state such that $x_i = 1$, $x_{j_1} = 0$, $x_{j_2} = 1$ and $x_k = 1$. Thus, because for any threshold parameters T in T^1, T^2 or T^3 we have $T(j_1, i) = T(j_2, i) = T(k, i) = 1$, we can state that i evolves toward $K_i^T(\{j_2, k\})$. The evolution of j_2 and k depends on the thresholds of (i, j_2) and (i, k). For example, if we consider the threshold function T^1, then, because $T^1(i, j_2) = 2$ and $T^1(i, k) = 1$, j_2's next expression level is given by $K_{j_2}^{T^1}(\emptyset)$ and k's next expression level is given by $K_i^{T^1}(\{i\})$.

Determining the dynamics of an interaction graph consists in the selection of possible threshold parameters, and then the attribution of values to the different logical parameters. The number of the possible attributions is huge: given a gene i with at least one predecessor, there are $2^{|V^-(i)|}$ logical parameters K_i, and each parameter can take at least two values. Thus, we have to consider $\prod_{i \in V} 2^{2^{|V^-(i)|}}$ possible attributions. For example, just for the interaction graph from Fig. 1, there are three possible set of threshold parameters, one leading to $2^{2^3} \times 2^{2^1} \times 2^{2^1} = 4096$ attributions for logical parameters (if i has two expressions levels), the two others leading to $3^{2^3} \times 2^{2^1} \times 2^{2^1} = 26244$ attribution (for i with three predecessors). Nevertheless, the structure of the interaction graph restricts the possible values of logical parameters.

3.3 Valid Logical Parameters

The values of logical parameters of an interaction graph must satisfy some constraints, linked to the graph structure and to the type of interaction. Logical parameters respecting the following constraints are said to be *valid*.

The *Definition constraint* is based on the definition of activation and inhibition. If a gene j which activates a gene i becomes effective, then we cannot be sure that i becomes itself effective (it may be inhibited by other genes), but the expression level of i cannot decrease.

Constraint 1 (Definition) *Let $G = (V, E, S, P, C)$ be a PGRN, and let T_G be an associated set of threshold function. Let i, j in V be two genes such that $j \in V^-(i)$. If $S(j,i) = +$ then $\forall \omega \subseteq V^-(i), K_i^{T_G}(\omega) \leq K_i^{T_G}(\omega \cup \{j\})$. If $S(j,i) = -$ then $\forall \omega \subseteq V^-(i), K_i^{T_G}(\omega) \geq K_i^{T_G}(\omega \cup \{j\})$.*

The *Observation constraint* expresses how we identify that a predecessor is an activator or an inhibitor. If j is an activator of i, then it exists at least one dynamic state where the effectiveness of j leads to an increase of the expression level of i. In other word, at least one of the previous inequalities is strict.

Constraint 2 (Observation) *Let $G = (V, E, S, P, C)$ be a PGRN, and let T_G be an associated threshold function. Let i, j in V be two genes such that $j \in V^-(i)$. If $S(j,i) = +$ then $\exists \omega \subseteq V^-(i), K_i^{T_G}(\omega) < K_i^{T_G}(\omega \cup \{j\})$. If $S(j,i) = -$ then $\exists \omega \subseteq V^-(i), K_i^{T_G}(\omega) > K_i^{T_G}(\omega \cup \{j\})$.*

Finally, the *Maximum constraint* expresses that in a dynamic state where all the activators of a gene are effective and simultaneously none of the inhibitors is effective, then the gene's expression level is maximum. Conversely,

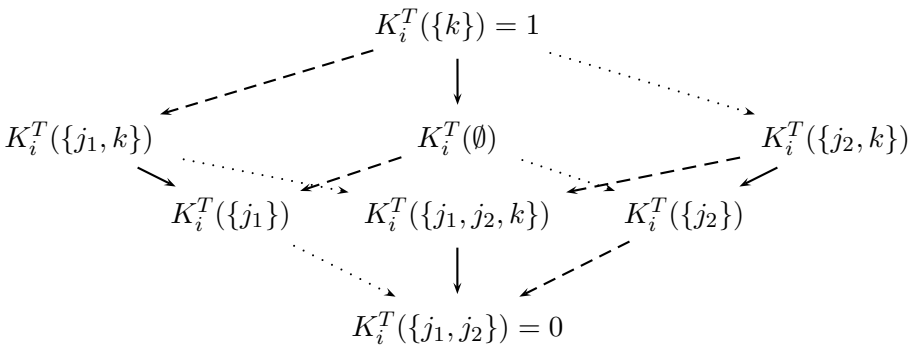


Figure 2: Relation among logical parameters of the interaction graph from Fig. 1 for any T in T^1, T^2 or T^3.

if none of the activators is effective, and all inhibitors are, then the logical parameter is minimum, that is equal to 0.

Constraint 3 (Maximum) *Let $G = (V, E, S, P, C)$ be a PGRN, and let T_G be an associated threshold function. Let i in V be a gene. By denoting $b_i = max\{T_G(i,j)|(i,j) \in E\}$, we have: $K_i^{T_G}(A(i)) = b_i$, and $K_i^{T_G}(I(i)) = 0$.*

Example 5 (Valid parameters) *Let us consider the interaction graph from Fig. 1. The considerations are done for any threshold function T in T^1, T^2 or T^3. The Maximum constraint imposes that $K_i^T(\{k\}) = 1$ and $K_i^T(\{j_1, j_2\}) = 0$. Other relations between parameters are resumed in Fig. 2, where an arrow from a node K to a node K' means $K \geq K'$ (Definition constraint), and this inequality is strict (Observation constraint) for at least one arrow of each type (plain, dashed or doted arrows). All three constraints taking into account, there are only 9 valid sets of parameters.*

4 Toward a reduction of valid dynamics

PCGRN include two new notions within the definition of interaction graph. Clusters help us to reduce the number of threshold functions to consider whereas privileged interactions reduce the number of valid logical parameters.

4.1 Clusters: Reduce the Number of Threshold Functions

The notion of clusters expresses the co-regulation of a set of genes, that is a set of spatially closed genes that are expressed at the same time due to the expression of a single regulating gene (*i.e.* the presence of a single transcription

factor). Thus by definition, clusters allow us to reduce the set of threshold function to consider. Indeed, if two genes j and k are influenced by a gene i, and belonged to a same cluster of i, then the two interactions (i, j) and (i, k) have the same threshold.

Constraint 4 (Clusters and thresholds) *Let $G = (V, E, S, P, C)$ be a PC-GRN. Then the threshold functions T_G to consider are such that: for all i in V, for all k, k' in $V^+(i)$*

$$\exists p \in \mathbb{N}, k \in C_i^p, k' \in C_i^p \Rightarrow T_G(i, k) = T_G(i, k')$$

Example 6 (Clusters and thresholds) *Let us consider the interaction graph from Fig. 1. If j_2 and k belong to a same cluster of i, then there is only one threshold function to consider: T^3 such that $T^3(i, j_2) = T^3(i, k) = 1$. Otherwise, the three possible threshold functions must be considered.*

4.2 Conflicts and Dilemma

Despite the above constraints, valid dynamics of PGRN still remain too numerous. The different dynamics exist due to some dynamics states where the three constraints do not allow us to determine unique values for logical parameters: *Conflicts* occur when a gene is simultaneously activated and inhibited, *Dilemma* occur when all the activators (resp. inhibitors) of a gene are not effective.

Definition 7 (Conflicts and dilemma) *Let $G = (V, E, S, P, C)$ be a PCGRN, and let T_G be an associated threshold function. Let $i \in V$ be a gene and let $x \in \mathbb{X}(G, T_G)$ be a dynamic state.*

- *x is a* situation of conflict *for gene i iff $A^*(i, x) \neq \emptyset$ and $I^*(i, x) \neq \emptyset$*
- *x is a* situation of dilemma *for gene i iff ($A^*(i, x) \neq \emptyset$ and $A^*(i, x) \neq A(i)$) or ($I^*(i, x) \neq \emptyset$ and $I^*(i, x) \neq I(i)$)*

In the following, we will focus on the determination of logical parameters. Thus, conflicts and dilemma will refer to parameters, that is $K_i(w^*(i, x))$ is a conflict (resp. a dilemma) if and only if x is a situation of conflict (resp. dilemma) for gene i. In other words, if $w^*(i, x) = \omega$, then $K_i(\omega)$ is a conflict iff $\omega \cap A(i) \neq \emptyset$ and $\omega \cap I(i) \neq \emptyset$; $K_i(\omega)$ is a dilemma iff $A(i) \not\subseteq \omega \not\subseteq I(i)$ or $I(i) \not\subseteq \omega \not\subseteq A(i)$.

Note that, in this model, $K_i(\emptyset)$ is neither a conflict nor a dilemma, but corresponds to the basal situation, where a gene i is not activated or inhibited.

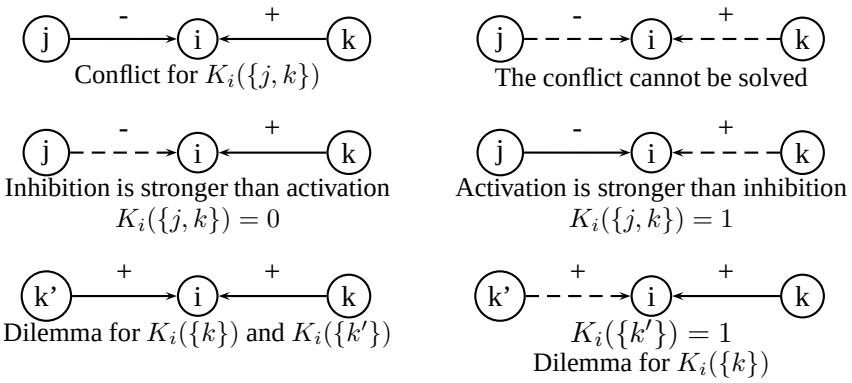


Figure 3: Solving conflicts and dilemma with privileged interactions

Example 7 (Conflicts and dilemma) *Let us consider the* 8 *possible dynamic states and the associated logical parameters for gene* i *for the interaction graph from fig. 1:* $K_i(\{j_1\})$ *and* $K_i(\{j_2\})$ *are dilemma;* $K_i(\{j_1, j_2, k\})$ *is a conflict;* $K_i(\{j_1, k\})$, $K_i(\{j_2, k\})$ *are both conflicts and dilemma.* $K_i(\{k\})$ *and* $K_i(\{j_1, j_2\})$ *are neither conflict nor dilemma: the former correspond to a situation where* i *is fully activated and is not inhibited, the latter corresponds to the reverse situation.*

4.3 Privileged Interactions: Reduce values of Logical Parameters

By definition, privileged interactions are such that their force is higher than the force of non privileged interactions. Figure 3 illustrates how to solve conflicts and dilemma using the privileged interactions: for conflicts, if two interactions occur simultaneously, then the privileged one is preferred; a dilemma is solved if one of the present gene is a privileged one.

This idea is captured through two constraints on logical parameters. The first constraint, called *Direct influence* indicates that if none of privileged activators (resp. inhibitors) is effective, and some privileged inhibitors (resp. activators) of the considered gene are effective, then the expression level cannot be maximum (resp. minimum).

Constraint 5 (Direct influence) *Let* $G = (V, E, S, P, C)$ *be a PCGRN, and let* T_G *be an associated threshold function. Let* $i \in V$ *be a gene and* $x \in \mathbb{X}(G, T_G)$ *be a dynamic state. By denoting* $b_i = max\{T_G(i,j) | (i,j) \in E\}$, *we have:*

- *if* $A^*(i,x) \cap P(i) \neq \emptyset$ *and* $I^*(i,x) \cap P(i) = \emptyset$ *then* $K_i^{T_G}(w^*(i,x)) > 0$
- *if* $I^*(i,x) \cap P(i) \neq \emptyset$ *and* $A^*(i,x) \cap P(i) = \emptyset$ *then* $K_i^{T_G}(w^*(i,x)) < b_i$

The second constraint, called *Relative influence*, states that expression levels of non privileged predecessors is not important compared to the presence or absence of privileged ones. In other words, the value of a logical parameter for a set of effective genes, whose at least one is a privileged predecessor, remains the same whatever non privileged predecessors becoming effective.

Constraint 6 (Relative influence) *Let $G = (V, E, S, P, C)$ be a PCGRN, and let T_G be an associated threshold function. Let $i \in V$ be a gene and let $\omega \subseteq V^-(i)$ be a set of predecessors of i such that $\omega \cap P(i) \neq \emptyset$. Let $j \in V^-(i)$ be a gene such that $j \notin P(i)$. By denoting $b_i = max\{T_G(i,j)|(i,j) \in E\}$, we have:*

- *if $K_i^{T_G}(\omega) < b_i$ then $K_i^{T_G}(\omega \cup \{j\}) < b_i$*
- *if $K_i^{T_G}(\omega) > 0$ then $K_i^{T_G}(\omega \cup \{j\}) > 0$*

Example 8 (Influence of privileged interactions) *Let us suppose that j_1 is the only privileged predecessor in Fig. 1. Then, as soon as j_1 is ineffective, conflict and dilemma appears between other genes, but when j_1 is effective, they are solved. The 9 valid sets of parameters are reduced to 2. If we now suppose that k is the only privileged predecessor, there is no conflict, but some dilemma remains, which reduced the number of dynamics to consider to 2. If j_1 and k are privileged predecessors, there are still conflict and dilemma, but the number of dynamics to consider is to reduced to 2. Finally, if we suppose that both j_1 and j_2 are privileged predecessors, then there is neither conflict nor dilemma, and the dynamics is unique.*

In [10], we study the case of Boolean dynamics, that is interaction graphs where genes have only two levels of expression. In that case, constraints on direct or relative influences are far more restrictive than in multivalued approach. Indeed, for the direct influence, the statement $K_i(w^*(i,x)) > 0$ is equivalent to $K_i(w^*(i,x)) = 1$ (and $K_i(w^*(i,x)) < b_i$ equivalent to $K_i(w^*(i,x)) = 0$); and the formulation of relative influence becomes $K_i(\omega) = K_i(\omega \cup \{j\})$. But, even if these constraints are not constructive in a multivalued approach, they reduce the number of dynamics to consider, and can be added to other systems of constraints, such as the ones we developed in [11] to search GRN with a dynamics verifying a given temporal property.

4.4 Unique Boolean Dynamics

We present here conditions to obtain, given a PCGRN, a unique set of parameters leading to a unique dynamics. We reduce the considered dynamics to Boolean dynamics and recall the result we present in [10]. Such a situation

is obtain when every threshold is equal to 1, which correspond to situations where any gene has only one cluster among its target. For that reason, we do not precise the chosen threshold function in this section. The theoretical results for any threshold function are more difficult to obtain, since we cannot control values of parameters with the constraints on direct or relative influence.

Obviously, if some genes have no predecessor, we cannot determine their expression levels, which in fact do not evolve along the time. A necessary and sufficient condition to have *no conflict* is that the set of privileged predecessors is either equal to activators or inhibitors.

Theorem 1 (No conflict) *Let $G = (V, E, S, P, C)$ be a Boolean PCGRN. The conflict situations of G can be solved iff for all $i \in V$, $P(i) = A(i)$ or $P(i) = I(i)$*

Proof 1 Sufficient. *Let x be a situation of conflict for gene i: $A^*(i, x) \neq \emptyset$ and $I^*(i, x) \neq \emptyset$. Let us suppose that $P(i) = A(i)$ (the proof is similar for $P(i) = I(i)$). Then we have $I^*(i, x) \cap P(i) = \emptyset$ and $A^*(i, x) \cap P(i) = A^*(i, x)$. Thus, due to the constraint of direct influence, $K_i(w^*(i, x)) = 1$ and the conflict is solved.*

Necessary. *Let us suppose that the condition is not verified for a given gene i, that is $P(i) \neq A(i)$ and $P(i) \neq I(i)$. $P(i) \neq A(i)$ iff either it exists $k \in A(i) \setminus P(i)$ or it exists $j \in I(i) \cap P(i)$; $P(i) \neq I(i)$ iff either it exists $j' \in I(i) \setminus P(i)$ or it exists $k' \in A(i) \cap P(i)$. If it exists $k \in A(i) \setminus P(i)$ and it exists $j' \in I(i) \setminus P(i)$, then the situation x where the only effective genes are k and j' is a situation of conflict. If it exists $k \in A(i) \setminus P(i)$ and it exists $k' \in A(i) \cap P(i)$, then two cases must be considered: if $I(i) \cap P(i) = \emptyset$ then, with $j'' \in I(i)$, the situation x where the only effective genes are k and j'' is a situation of conflict; if $I(i) \cap P(i) \neq \emptyset$ then, with $j'' \in I(i) \cap P(i)$, the situation x where the only effective genes are k' and j'' is a situation of conflict.*

Nevertheless, if all privileged predecessors are ineffective, then a situation of dilemma may occur. Dilemmas occur when two genes having the same action (either activation or inhibition) are not effective simultaneously. Thus, a necessary and sufficient condition to have *no dilemma* is that either there is only one gene for a given action, or each predecessor having this type of action is a privileged predecessor of the target.

Theorem 2 (No dilemma) *Let $G = (V, E, S, P, C)$ be a Boolean PCGRN. The dilemma situations of G can be solved iff for all $i \in V$, ($A(i) \subseteq P(i)$ or $|A(i)| = 1$) and ($I(i) \subseteq P(i)$ or $|I(i)| = 1$).*

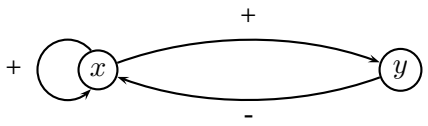


Figure 4: Interaction graph for the mucus production system in *P. aeruginosa*

Proof 2 Sufficient. *Let us consider the case of activation (the proof is similar for inhibition). Obviously, if* $|A(i)| = 1$*, then there is no dilemma. If* $A(i) \subseteq P(i)$*, then: for all* $\omega \subseteq A(i)$*, if* $\omega \neq \emptyset$ *then* $K_i(w) = 1$ *due to the constraint of direct influence; for all* $\omega_a \subseteq A(i)$*, for all* $\omega_i \subseteq I(i) \setminus P(i)$*, if* $\omega_a \neq \emptyset$ *then* $K_i(\omega_a \cup \omega_i) = 1$*, due to the constraint of relative influence; the remaining cases correspond to situations of conflict where both activators and predecessors are privileged predecessors of* i*.*

Necessary. *Let us suppose that the condition is not verified. Let us suppose we have* $|A(i)| > 1$ *and* $A(i) \not\subseteq P(i)$ *(the proof is similar for the inhibition). Then it exists* $a \in A(i) \setminus P(i)$*, and the situation* x *where* a *is the only effective predecessor of* i *is a situation of dilemma.*

Theorem 3 (No conflict nor dilemma) *Conflict and dilemma situations of a Boolean PCGRN* (V, E, S, P, C) *can be solved iff for all* $i \in V$*, (*$A(i) = P(i)$ *and* $|I(i)| = 1$*) or (*$|A(i)| = 1$ *and* $I(i) = P(i)$*)*

Proof 3 *The theorem is a direct consequence of theorems 1 and 2.*

Under the conditions of this theorem, only one dynamics is consistent with all constraints. Obviously, these conditions are difficult to state in practice. Section 5 will nevertheless illustrate that in any case, the consideration of privileged interactions allows us to reduce the set of consistent dynamics.

5 Influence of Clusters and Privileged Interactions on Dynamics

5.1 From a Biological Case Study

Pseudomonas aeruginosa are bacteria that secrete mucus (alginate) in lungs affected by cystic fibrosis, but not in common environment. As this mucus increases respiratory defficiency, this phenomenon is a major cause of mortality. Details of the regulatory network associated with the mucus production by *Pseudomas aeruginosa* are described by Govan and Deretic [4] but a simplified genetic regulatory network has been proposed by Guespin and Kaufman [5], see Fig.4.

It has been observed that mucoid *P. aeruginosa* can continue to produce mucus isolated from infected lungs. It is commonly thought that the mucoid state of *P. aeruginosa* is due to a mutation which cancels the inhibition of gene

x. An alternative hypothesis has been made: this mucoid state can occur by reason of an epigenetic modification, *i.e.* without mutation [5]. The models compatible with this hypothesis are constructed in [1].

5.1.1 Boolean Dynamics

The logical parameters to consider are $K_y(\emptyset)$ and $K_y(\{x\})$ for the gene y and $K_x(\emptyset)$, $K_x(\{x\})$, $K_x(\{y\})$ and $K_x(\{x,y\})$ for gene x, which leads without further consideration, to $2^2 \times 2^4 = 64$ possible dynamics. Obviously, this number is decreased considering the constraints previously presented. $K_y(\emptyset) = 0$ and $K_y(\{x\}) = 1$ due to the observation rule. The maximum rule leads to $K_x(\{x\}) = 1$ and $K_x(\{y\}) = 0$, and then the observation rule leads to two possible dynamics: either ($K_x(\emptyset) = 1$ and $K_x(\{x,y\}) = 1$) or ($K_x(\emptyset) = 0$ and $K_x(\{x,y\}) = 0$).

The two possible dynamics are due to the conflict between x and y, and then the knowledge of privileged interactions among the activation of x by itself or the inhibition of x by y would lead to the determination of a unique dynamics. If both the interactions are privileged ones (or conversely are not privileged ones) then the two dynamics remain valid. If the inhibition is privileged and not the activation, then $K_x(\emptyset) = 0$ and $K_x(\{x,y\}) = 0$. If the activation is privileged and not the inhibition, then $K_x(\emptyset) = 1$ and $K_x(\{x,y\}) = 1$.

5.1.2 Multivalued Dynamics

Given that x has two predecessors, and y only one, there are three threshold functions to consider. Obviously, for each one $T(y,x) = 1$. The first threshold function is such that $T^1(x,y) = T^1(x,x) = 1$, and may seems similar to the Boolean situation, but in fact because the constraints on direct and relative influence are not constructive in multivalued approach, they do not allow us to choose between the different model. The two others are such that $T^2(x,y) = 2$, $T^2(x,x) = 1$ and $T^3(x,y) = 1$, $T^3(x,x) = 2$. The known logical parameters are given in the following table:

x	y	$K_x^{T^2}$	$K_y^{T^2}$	$K_x^{T^3}$	$K_y^{T^3}$
0	0	$K_x^{T^2}(\emptyset)$	0	$K_x^{T^2}(\emptyset)$	0
0	1	0	0	0	0
1	0	2	0	$K_x^{T^2}(\emptyset)$	1
1	1	$K_x^{T^2}(\{x,y\})$	0	0	1
2	0	2	1	2	1
2	1	$K_x^{T^2}(\{x,y\})$	1	$K_x^{T^2}(\{x,y\})$	1

Because of the observation constraint, we cannot have ($K_x^T(\{x,y\}) = 2$ and $K_x^{T^2}(\emptyset) = 0$) or ($K_x^T(\{x,y\}) = 0$ and $K_x^{T^2}(\emptyset) = 2$), which leads to seven

valid dynamics.

5.2 From Artificial PGRN

In order to estimate the reduction in number of models induced by the introduction of privileged interactions, we have randomly generated PGRN, that is PCGRN without any cluster information. The generation is parameterized by three values: n the number of genes, p the number of predecessors of a gene and r a ratio to determine which interactions are privileged. We first generate n genes; for each gene we then randomly select p predecessors among the n genes, each one being a privileged predecessor with a probability r. For each gene, we finally randomly select a maximum threshold (that is a random number between 1 and its number of successors), and define for each outgoing interaction its threshold between 1 and this maximum threshold, verifying that every value between 1 and the maximum threshold is selected at least one time.

Fig. 5 presents some results on artificial PCGRN composed of $n = 10$, 25, 50 and 100 genes. We give one table by hypothesis on the considered number of predecessors: the first two tables correspond to situations where each gene has exactly $p = 2$ or 3 predecessors, and the last table to a situation where each gene has a random number of predecessors between 1 and 3. We chose these rather small values for the number of predecessors per gene to fit a realistic ratio between number of genes and number of interactions.

For each PCGRN we evaluate the number of dynamics without any constraint (row named "*Total*" in each table). We then compute the number of dynamics when all the constraints (definition, observation, maximum, direct and relative influence) are applied, for several ratios of privileged interactions: when there is no privileged interaction (row "0"), when one interaction out of ten is privileged (row "1/10"), one out of five (row "1/5"), one out of two (row "1/2") and when all interactions are privileged ones (row "1"). Let us note that results between row "1" and row "0" may be largely different, since when all predecessors are privileged (row "1"), then the effectiveness of only one of them allows us to solve dilemma unsolved in row "0". All the values in the different tables given in Fig. 5 are the result of an arithmetic mean over 100 tests. The column "100 genes" for the hypothesis "3 predecessors per gene" is left empty, due to the excessive required computation time.

Obviously, the number of dynamics we have to deal with is huge (at least 10^{16}, see row "*Total*"), When considering the constraints of definition, observation and maximum, the number of dynamics is already significantly reduced (see row "0" where none of the interactions is privileged). With the constraints induced by the introduction of privileged interactions (direct and relative influence), the number of dynamics still decreases and the best results are obtained when half of interactions are privileged ones (row "1/2"). Nevertheless, let

Privileged ratio r	Number of genes n			
	10	25	50	100
0	10^{6}	10^{15}	10^{31}	10^{60}
$1/10$	10^{5}	10^{14}	10^{28}	10^{53}
$1/5$	10^{5}	10^{13}	10^{25}	10^{50}
$1/2$	10^{4}	10^{10}	10^{21}	10^{40}
1	10^{4}	10^{11}	$.10^{22}$	10^{42}
$Total$	10^{16}	10^{42}	10^{82}	10^{162}

Each gene has $p = 2$ predecessors

Privileged ratio r	Number of genes n			
	10	25	50	100
0	10^{21}	10^{51}	10^{98}	–
$1/10$	10^{19}	10^{48}	10^{82}	–
$1/5$	10^{19}	10^{42}	10^{70}	–
$1/2$	10^{15}	10^{38}	10^{48}	–
1	10^{17}	10^{44}	$.10^{65}$	–
$Total$	10^{41}	10^{100}	10^{182}	–

Each gene has $p = 3$ predecessors

Privileged ratio r	Number of genes n			
	10	25	50	100
0	10^{11}	10^{30}	10^{48}	10^{82}
$1/10$	10^{9}	10^{22}	10^{39}	10^{74}
$1/5$	10^{9}	10^{22}	10^{37}	10^{67}
$1/2$	10^{8}	10^{18}	10^{26}	10^{57}
1	10^{10}	10^{19}	10^{32}	10^{66}
$Total$	10^{27}	10^{63}	10^{110}	10^{211}

Each gene has between 1 and 3 predecessors

Figure 5: Number of Dynamics for Artificial PCGRN

us point out that the improvement is clearly observed even with small information. For example, when only one interaction out of ten is privileged (row "1/10"). we can observe that in the third table, the number of dynamics is divided by 100 for a ten genes network, by 10^8 for 25 genes, and by 10^8 for 100 genes.

These few simulations illustrate that as soon as spatial information is known, the set of all possible dynamics is really restricted. To go further in this restriction, one can express temporal properties to characterise some knowledge about the behaviour of the GRN. Formal techniques, most of them based on model checking [1], have been applied to select valid dynamics, that is dynamics consistent with biological experiments expressed by temporal properties. The problem is that these formal techniques rapidly become intractable because dynamics associated to the GRN are most of the time very numerous. Thus, from a general point of view, the set of PCGRN dynamics is all the more reduced than all biological knowledge, including spatial information, is taken into account.

6 Concluding Remarks

In this article we have presented a simple way to include spatial information within the René Thomas' framework of GRN. This supplementary information is described as a property of interactions: an interaction is privileged when the source and target genes are known to be spatially close. In the framework of Boolean dynamics, values of logical parameters are weakly constrained, leading to situations of conflicts or dilemmas where several dynamics are possible. With the notion of privileged interactions, we have determined conditions to solve some of these situations.

The spatial oriented framework we have defined is based on René Thomas' Boolean dynamics and presents the two following advantages. Firstly, since the dynamics for our spatial framework are chosen among classical René Thomas' Boolean dynamics associated to the underlying GRN without privileged interaction, then our dynamics are clearly included in the usual dynamics of GRN. Secondly, since spatial information allows us to solve some conflicts and dilemmas, and thus to determine some logical parameters, the number of dynamics is in practice considerably reduced.

In the goal of validating our approach, we are facing to the fact that, although spatial information seams to be central in order to apprehend the complexity of biological networks, experimental data are rare. Indeed, available data mainly concern large GRN, which are for the moment hardly attainable with our approach due to the high number of parameters to consider. Nevertheless our approach seems particularly adapted, since the first results appear even with few information on spatial relation.

Aknowledgments

Authors want thank all the members of the "Groupe Observabilité" from Epigenomics Project for the discussions around distances between genes in genetic regulatory networks..

References

[1] G. Bernot, J.-P. Comet, A. Richard, and J. Guespin. Application of formal methods to biological regulatory networks: Extending Thomas' asynchronous logical approach with temporal logic. *Journal of Theoretical Biology*, 229(3):339–347, 2004.

[2] Julio E Cabrera and Ding J Jin. The distribution of rna polymerase in escherichia coli is dynamic and sensitive to environmental cues. *Mol Microbiol*, 50(5):1493–1505, 2003.

[3] Anne-Sophie Carpentier, Bruno Torresani, Alex Grossmann, and Alain Henaut. Decoding the nucleoid organisation of bacillus subtilis and escherichia coli through gene expression data. *BMC Genomics*, 6(1):84, 2005.

[4] J.R. Govan and V. Deretic. Microbial pathogenesis in cystic fibrosis: mucoid pseudomonas aeruginosa and burkholderia cepacia. *Microbiol rev.*, 60(3):539–74, 1996.

[5] J. Guespin-Michel and M. Kaufman. Positive feedback circuits and adaptive regulations in bacteria. *Acta Biotheor.*, 49(4):207–218, 2001.

[6] D. A. Jackson, A. B. Hassan, R. J. Errington, and P. R. Cook. Visualization of focal sites of transcription whitin human nuclei. *J. Cell Biol.*, 164:515–526, 2004.

[7] François Képès. Periodic transcriptional organization of the e.coli genome. *J Mol Biol*, 340(5):957–964, 2004.

[8] François Képès and Cédric Vaillant. Transcription-based solenoidal model of chromosomes. *Complexus*, 1(4):171–180, 2003.

[9] Christian Lanctot, Thierry Cheutin, Marion Cremer, Giacomo Cavalli, and Thomas Cremer. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet*, 8(2):104–115, 2007.

[10] M. Manceny, M. Aiguier, P. Le Gall, J. Hérisson, I. Junier, and F. Képès. Spatial information and boolean genetic regulatory networks. *BioInformatics and Computational Biology (BICoB)*, LNBI 5462:270 – 281, 2009.

[11] M. Manceny, P. Le Gall, J.-P. Comet, and J.-P. Gallois. Model-checking for parametric transition systems based on symbolic execution. Technical report, Ecole Centrale Paris, MAS Laboratory, 2008.

[12] D. Mateus, J.-P. Gallois, J.-P. Comet, and P. Le Gall. Symbolic modeling of genetic regulatory networks. *Journal of Bioinformatics and Computational Biology*, 2007.

[13] B Muller-Hill. The function of auxiliary operators. *Mol Microbiol*, 29(1):13–18, 1998 Jul.

[14] R. Thomas. Logical analysis of systems comprising feedback loops. *J. Theor. Biol.*, 73(4):631–56, 1978.

Gene Regulatory Networks: Introduction of multiplexes into R. Thomas' modelling

Z. Khalis[1,2], G. Bernot[1,2], J.-P. Comet[1] and Observability Group[2]

[1] I3S, CNRS & Univ. of Nice-Sophia Antipolis, F-06903 Sophia Antipolis, France
[2] Epigenomics Project, Genopole, F-91034 Évry, France

Abstract

When modelling gene regulatory networks, the cornerstone of the modelling process is the search of parameter values which are consistent with the known properties of the system. These parameters drive the dynamics of the system. In this article, we give a formal definition of a slight extension of the R. Thomas' modelling framework, with explicit information about cooperative, concurrent or more complex molecular interactions. It considerably decreases the number of parameters and determining parameter values becomes less time consuming, making possible the study of larger systems.

1 Introduction

To study complex biological systems, formal modelling is often mandatory since the complexity of the interleaved interactions between constituents makes intuitive reasoning error prone. Numerous mathematical modelling frameworks have been proposed to model gene regulatory networks, see for example [7, 13, 20, 8]. Common approaches are quantitative, based on differential or stochastic equations, providing numerical simulations of the system. Nevertheless actual predictions often remain only qualitative because the parameter values of these systems are not precisely known. Several other modelling frameworks are based on a qualitative view, see for example boolean networks and their generalizations [16, 19], Petri nets [4, 6], hybrid modellings [12, 1], and stochastic π-calculus [5]. Each modelling framework highlights some views of models and allows one to detail or to abstract different biological aspects.

We focus here on Thomas' modelling, in which the gene regulatory system is represented by an interaction graph and a set of parameters. The interaction graph is composed, on the one hand, of nodes which abstract genes and their proteins, and on the other hand, of edges which represent the interactions between the genes. The values assigned to the parameters permit one to deduce the dynamics of the system from the interaction graph. Even in a qualitative perspective, the lack of reliable data about the system leads to a typical difficulty of the modelling approach : How to select the parameter values of the model?

For determining values of parameters, we proposed in [3] to test the set of all possible parameterizations against temporal properties. It is finite in the case of Thomas' modelling. This approach can be computer aided [3] using formal temporal logics and systematic model checking. Even if the set of possible parameterizations is finite, it exponentially grows with the size of the interaction graph. Several theorems established in the Thomas' framework considerably reduce the number of generated parameter sets, nevertheless, an entire exploration is not conceivable for large networks.

In order to reduce the time required by this exploration step, it becomes crucial to introduce in the modelling framework more biological information (when available). In this chapter, we propose to take into account information about how constituents of the system act on their targets. For example (Figure 1), if two genes act positively on a common target *via* the formation of a complex (e.g. the transcription factor of the common target contains the complex), then it is obvious that the common target has in fact a unique predecessor (the complex instead of two genes separately) and only two possibilities (instead of four) can occur: The complex is present and the transcription can take place or the complex is not present. Indeed this idea is far from being new but it has never been formalized up to now. R. Thomas remarked that this kind of information can be taken into account in its modelling framework through the valuation of parameters, but he did not explicitely include such information in the interaction graph [18].

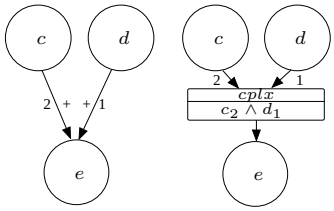


Figure 1: Example of cooperative action

Here, we propose a modelling framework in which the interaction graph makes such cooperative or concurrent biological phenomenon explicit. The decreasing of the number of parameters coupled with the methodology developed in [3], will make possible the study of larger systems.

The chapter is organized as follows. We firstly define our new interaction graph: *Multiplexes* are formally defined to take into account available biological information describing the cooperation or concurrency between constituents acting on a common target. Then we define when a multiplex has an effective action on its targets, and we construct the associated dynamics.

We show that Thomas' and multiplex frameworks have the same power of expression but we illustrate, through the classical example of the lac operon, how multiplexes allow us to be more legible and terse. Lastly we present the benefits of this multiplex modelling.

2 Gene regulatory graphs with multiplexes

Formal modelling frameworks for gene regulatory networks represent interactions between entities (genes, proteins, *etc.*) *via* a *static graph*. Then, dynamics focus on the evolution of entity expression levels and ask for more elaborated mathematical stuff with many parameters.

In our framework, we represent the static part by a directed graph composed of two types of vertices: Variables which correspond to genes and their products, and multiplexes which correspond to interactions between variables. Multiplexes abstract biological phenomena like complex forming or more elaborated phenomena. The predecessors of a multiplex are either variables or other multiplexes brought into play in the interaction; the successors are called the targets of the interaction.

2.1 Formal Definition

The following notation will be useful.

Notation 1 *Given a directed graph G and a node v of G, $G^{-1}(v)$ is the set of all nodes v' of G such that (v', v) is an edge of G (set of predecessors of v).*

A multiplex is provided with a formula in a propositional logic which encodes the situations in which the interaction occurs. For example, if a complex formed with proteins a and b is required in cooperative action and if the complex (a-b) is inactive in the presence of a protein c, then the corresponding formula looks like "$a \wedge b \wedge \neg c$," where the symbols "$\wedge$" and "$\neg$" stand for "and" and "not" respectively.

Definition 1 *A* gene regulatory graph with multiplexes, RG *for short, is a tuple $G = (V, M, E_V, E_M)$ such that:*

1. *$(V \cup M, E_V \cup E_M)$ constitutes a (labelled) directed graph whose set of nodes is $V \cup M$ and set of edges is $E_V \cup E_M$, with $E_V \subset V \times \mathbb{N} \times M$ and $E_M \subset M \times (V \cup M)$.*
2. *V and M are disjoint finite sets. Nodes of V are called* variables *and nodes of M are called* multiplexes. *An edge (v, s, m) of E_V is denoted $(v \xrightarrow{s} m)$ where s is called the threshold.*

3. *Each variable v of V is labelled with a positive integer b_v called the bound of v.*

4. *Each multiplex m of M is labelled with a formula belonging to the language L_m inductively defined by:*

   - *If $(v \xrightarrow{s} m) \in E_V$, then v_s is an atom of L_m, and if $(m' \to m) \in E_M$ then m' is an atom of L_m.*
   - *If ϕ and ψ belong to L_m then $\neg\phi$, $(\phi \wedge \psi)$, $(\phi \vee \psi)$ and $(\phi \Rightarrow \Psi)$ belong to L_m.*

5. *All cycles of the underlying graph $(V \cup M, E_V \cup E_M)$ contain at least one node belonging to V.*

Note: Condition 5 is necessary for the definition of dynamics (see Definition 3).

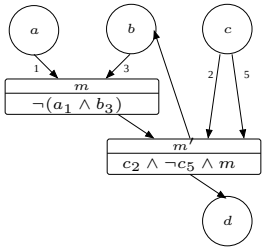


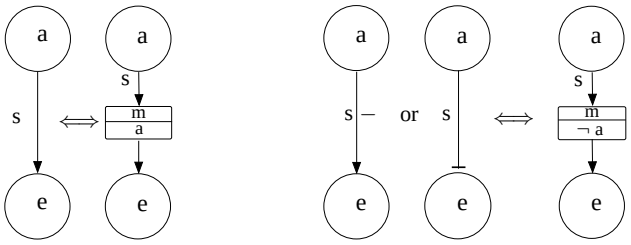
Figure 2: Graphical conventions

Figure 2 provides graphical conventions. In this figure, a, b, c, d are variables; m, m' are multiplexes; m and c are the inputs of m' and b and d are its outputs; the cycle b, m, m' contains the variable b.

In addition to these standard graphical conventions, we allow "light" additional graphical notation abuse:

- If a variable is an input of a multiplex with only one threshold, we then allow to omit the threshold in the formula. For example, in Figure 2, the formula of multiplex m can be simply written as "$\neg(a \wedge b)$." Of course, this light form is not possible for m'.

- Multiplexes with a formula reduced to a unique atom can be removed from the diagram. In figure 3a, removing the multiplex m allows us to retrieve the usual diagrammatic convention of R. Thomas for activations.



(a) Activation. (b) Inhibition.

Figure 3: Light graphical convention for activation and inhibition.

- Similarly, in figure 3b, we retrieve usual inhibitions, either by adding the minus sign, or by using the "inhibition arrow" usual in biology.

In figure 2, we also see that in multiplex formulas the variables are indexed by their thresholds. This is useful when a given variable acts on a multiplex at several thresholds. The multiplex m' means that the expression level of c must be both greater than 2 and lower than 5 in order to participate to the induction of d.

2.2 States and resources

A gene regulatory graph with multiplexes constitutes the static representation of the system. We have now to focus on the dynamics of the system, abstracted by the evolutions of expression levels of variables. Let us first define the states of a system.

Definition 2 *A* state *of a* RG $G = (V, M, E_V, E_M)$ *is a map* $\eta : V \to \mathbb{N}$ *such that for each variable* v *belonging to* V, $\eta(v) \leq b_v$.
$\eta(v)$ *is called the expression level of* v.

A multiplex does not have any expression level because it is a logical composition of variables at a given time. So, we consider only the expression level of all the variables at that time and from this current state it is possible to deduce if the multiplex is active or not *via* the interpretation of its propositional formula.

According to a current state, the set of resources of a variable a is the set of multiplexes which can help a to express its product. More precisely a resource r of a variable a is a multiplex belonging to $G^{-1}(a)$ whose formula is satisfied. So graphically, edges of interaction graphs have no sign but negative actions are taken into account through multiplexes with the operator $\neg$. For example,

in Figure 2 the multiplex m represents an inhibition (the complex a-b inhibits b and d *via* m').

Definition 3 *Given a* RG $G = (V, M, E_V, E_M)$ *and a state* η *of* G*, the set of resources of a variable* $v \in V$ *for the state* η *is the set of multiplexes* m *of* $G^{-1}(v)$ *such that the formula* φ_m *of the multiplex* m *is satisfied. The interpretation of* φ_m *in* m *is inductively defined by:*

- *If* φ_m *is reduced to an atom* v_s *of* $G^{-1}(m)$ *then* φ_m *is satisfied iff* $\eta(v) \geq s$*.*
- *If* φ_m *is reduced to an atom* $m' \in M$ *of* $G^{-1}(m)$ *then* φ_m *is satisfied iff* $\varphi_{m'}$ *of* m' *is satisfied.*
- *If* $\varphi_m \equiv \psi_1 \wedge \psi_2$ *then* φ_m *is satisfied if* ψ_1 *and* ψ_2 *are satisfied; and we proceed similarly for all other connectives.*

We note $\rho(v, \eta)$ *the set of resources of* v *for the state* η*.*

This definition is actually inductive because RG never contain a cycle of multiplex (item 5 of Definition 1). If cycle of multiplexes were allowed then indeterminations or contradictions would be possible. For instance, consider the graph in figure 4. Suppose that the expression level of a is greater or equal to the threshold s:

- If the formula of m' is assumed to be satisfied, then the formula of m must be satisfied and so the formula of m' cannot be satisfied. So, we get an inconstency.
- If the formula of m' is assumed to be unsatisfied, then the formula of m must be unsatisfied and so the formula of m' must be satisfied. So, whatever we assume, we always get an inconsistency.

Let us consider now, the graph in figure 4 where the formula associated with m' is m instead of $\neg m$. Suppose again that the expression level of a is greater or equal to the threshold s. Then, the two interpretations of m' are consistent and compatible with the current state. There is an indetermination which is similar to the notion of schizophrenic cycles of [15].

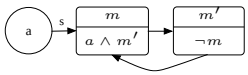


Figure 4: Cycle of multiplexes

To avoid these inconsistencies and indeterminations, cycles of multiplexes are not allowed. This motivates the item 5 of Definition 1.

3 Gene networks with multiplexes

We call network a graph associated with the parameters which determine the dynamics.

Definition 4 *A* gene regulatory network with multiplexes *(*RN*) is a couple* $(G, \mathcal{K})$ *where*

- $G = (V, M, E_V, E_M)$ *is a* RG.
- $\mathcal{K} = \{k_{v,\omega}\}$ *is a family of parameters indexed by* $v \in V$ *and* $\omega \subset G^{-1}(v)$ *such that all* $k_{v,\omega}$ *are integers and* $0 \leq k_{v,\omega} \leq b_v$.

Notice that each variable v admits 2^n parameters of the form $k_{v,\omega}$ where n is the in-degree of v in G.

Additional restrictions for the choice of parameters can be considered. The Snoussi's hypotheses [14] which ensure the consistency of qualitative behaviours with some underlying differential equation system, are well-known: If $\omega \subset \omega'$ then $k_{v,\omega} \leq k_{v,\omega'}$. These hypotheses signify that an effective resource cannot induce the decrease of the expression level of v. Moreover, we can always ignore the parameters $k_{v,\omega}$ such that the conjunction of the formulas associated with the multiplexes in ω is unsatisfied for all states.

3.1 Dynamics

The value of the parameter $k_{v,\rho(v,\eta)}$ (where ρ is defined in definition 3 above), indicates how the expression level of v can evolve from the state η. It can increase (respectively decrease) if the parameter value is greater (respectively less) than $\eta(v)$. The expression level must stay constant if both values are equal. The tendency (increasing, decreasing, unchanging) of variables are given by the directional map associated with each state:

Notation 2 *Given a* RN $N = (G, \mathcal{K})$ *and a state* η *of* $G = (V, M, E_V, E_M)$*, the* directional map $d : V \to \{-1, 0, 1\}$ *is defined by:*

$$\forall v \in V, d(v) = \begin{cases} -1 & \text{if } \eta(v) > k_{v,\rho(v,\eta)} \\ 0 & \text{if } \eta(v) = k_{v,\rho(v,\eta)} \\ 1 & \text{if } \eta(v) < k_{v,\rho(v,\eta)} \end{cases}$$

The probability that two variables change their expression level at the same time is negligible *in vivo*; following the Thomas' approach a state transition of the model modifies only one of the involved variables at a time.

Definition 5 *Let* $N = (G, \mathcal{K})$ *be a* RN*, and let* η *be a state of* G*. A state* η' *of* G *is a* successor *of the state* η *if and only if :*

- *There exists a variable* u *such that* $\eta'(u) = \eta(u) + d(u)$ *and* $d(u) \neq 0$
- *For any other variable* $v \neq u$ *we have* $\eta'(v) = \eta(v)$

In each state transition, at most one variable is modified; this procedure is called *asynchronous update* in Thomas' framework.

Definition 6 *The* asynchronous state graph *of a* RN $N = (G, \mathcal{K})$ *is the graph* S *defined by:*

- *The set of vertices of* S *is the set of possible states of* G *(isomorphic to the Cartesian product* $\prod_{v \in V} [0, b_v]$*).*
- *The set of edges of* S *is the set of couples* (η, η') *such that* η' *is a successor of* η*.*

4 Relative terseness with respect to the classical framework

Obviously our framework with multiplexes embeds the classical Thomas' framework [17] as it is sufficient to translate an activation (resp. an inhibition) with a multiplex whose formula is reduced to the input variable (resp. its negation), see Figure 3. Conversely, a non atomic formula in a multiplex obviously corresponds to a constraint on the parameters [18] following an induction similar to the one of Definition 3.

Our conviction is that this kind of knowledge is a static knowledge and consequently it should be present in the interaction graph (formulas in multiplexes). When we know, for biological reasons, the nature of combined influences, this information should be included in the model as soon as possible because it considerably reduces the number of possible parameters, as shown in the example below. Of course, the nature of combined influences is not always *a priori* known and, in this case, according to our formalism, variables have then several inputs in the regulatory graph.

4.1 Example of lactose operon.

The cell needs carbon. Carbon is preferably obtained from glucose *via* a given catalytic pathway. When glucose is absent, lactose is used *via* an alternative catalytic pathway.

Lactose operon in E.coli is the first genetic regulatory system elucidated, by François Jacob and Jacques Monod [9]. The induction of this system requires two conditions: Absence of glucose and presence of lactose.

An operon is a set of contiguous genes whose transcription is controlled simultaneously by a unique transcription factor. This transcription factor has

an affinity with a DNA area at the beginning of the operon, called operator and denoted O.
The lactose operon is formed by three genes denoted by Z, Y and A. The genes Z, Y and A produce respectively the enzymes β-galactosidase, permease and thiogalactoside transacetylase.

When glucose is absent, the alternative pathway is controlled as follows:

- CAP (Catabolite gene Activator Protein) forms a complex with cAMP (cyclic Adenosine MonoPhosphate), and binds to DNA to increase the transcription of the operon. This is a positive regulation.
- The transcription of the operon is possible only if the DNA area O is free. The regulatory protein lacI binds to O, this is a negative regulation. However, when lactose is present, a lactose isomer binds to lacI and lacI looses its affinity for O. So the operator O becomes free.

When glucose is present, the alternative pathway is inhibited as follows: Glucose inhibits indirectly cAMP and leads to the absence of complex CAP-cAMP. Consequently, there is no transcription even if lacI is present.

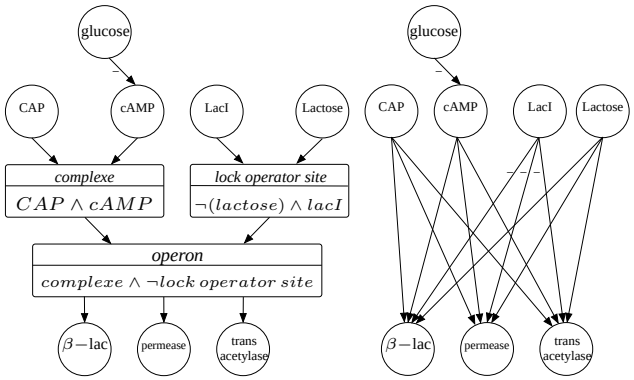


Figure 5: Lactose operon metabolism graph with and without multiplexes.

In Figure 5, the interaction graph of the lactose operon is represented in the multiplex framework (the left part of the figure) and in the classical Thomas' framework (the right part of the figure). The first advantage of the multiplex framework is its legibility: The left hand side of the figure is, to some extent, more legible than the textual description given before. On the contrary, the right hand side of the figure cannot be understood without the textual description.

The second advantage of the multiplex framework is methodological. When we try to elucidate a biological system using Thomas' framework, we do not

know the values of the parameters: The $k_{v,\omega}$ have to be inferred from *in vivo* behaviours. Consequently, models with a small number of parameters allow us to rapidly converge towards the elucidation of the studied biological system. On the contrary, models with large numbers of parameters can be so heavy to manipulate that they obstruct the discovery process. On this small lactose operon example, the total number of parameters according to the multiplex approach is 12, while the total number of parameters according to the classical approach is 54. Putting as much static information as possible explicitly in the graph (instead of putting it later manually in the dynamics) considerably reduces the complexity of the modelling methodology. Indeed, formalizing cooperative actions of several variables on the same target *via* multiplexes enables one to merge into a single multiplex the different acting resources.

The knowledge formalised into multiplexes can lead to reduce even more the number of useful parameters. In figure 6, multiplexes m_1 and m_2 cannot be satisfied for the same state: m_1 is active only if expression level of a is strictly less than 2 whereas m_2 is active when expression level of a is greater or equal to 2. Among the set of formal parameters $\mathcal{K} = \{k_{c,\{\}}, k_{c,\{m_1\}}, k_{c,\{m_2\}}, k_{c,\{m_1,m_2\}}\}$, $k_{c,\{m_1,m_2\}}$ is never used. More generally, when two multiplexes having the same target v have two *mutually exclusive formulas* ϕ_1 and ϕ_2, all parameters of the form $K_{v,\omega\cup\{m_1,m_2\}}$ can be ignored and the number of relevant parameters is reduced.

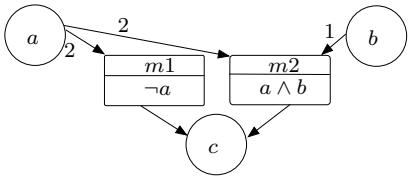


Figure 6: Example of RG which contains mutually exclusive formulas

5 Application

The software SMBionet-3.0 [11] has been designed to facilitate the modelling process of genetic regulatory systems. It allows one to select models of given RG according to their temporal properties. It takes as input a RG and a formula in temporal logic expressing the known or hypothetical temporal properties of the system. It gives as output all the models satisfying the formula.

In both modelling frameworks (with or without multiplexes), we have to give a value to each parameter in order to deduce the dynamics of the system. Because parameter values are not *a priori* known this leads us to consider an

enormous number of parameterizations. Indeed, each variable v admits 2^n parameters of the form $k_{v,\omega}$ where n is the in-degree of v in G ($\omega \subset G^{-1}(v)$). Each of these parameters can take $b_v + 1$ different values where b_v is the bound of v. The number of parameterizations is then given by $\prod_{v \in V} (b_v + 1)^{2^n}$ where n is the in-degree of v. For the example of lactose operon in Thomas' framework, the number of parameterizations is on the order of 2.27×10^8 whereas in our multiplex framework, the number of parameterizations is 1296. For instance, in Thomas' framework, the variable $permease$ has 2^4 parameters, generating $2^{2^4}(65536)$ different parameter settings while in our framework, $permease$ has 2 parameters, generating $2^2(4)$ different parameter settings. The difference resides in the addition of the multiplexes, which reduces the number of inward edges to $permease$ and so the number of possible parameter settings. Consequently, taking into account information about cooperation between variables (through multiplexes) leads to a significant decreasing of the number of possible models: Here, the set of possible models is cut down by a factor of 175000.

We used SMBionet-3.0 to exhibit models which present characteristic alternative catalytic pathway when glucose is absent. Under the Snoussi's hypotheses (see section 3 Biological Regulatory Networks with multiplexes) and for a given logical formula, all possible parameter settings in our framework have been explored in 27 seconds whereas all possible parameter settings in Thomas' framework have been explored in approximately 1000 hours. Notice that the ratio between both time is less than 175000 because SMBionet-3.0 optimizes the exploration of the model set.

6 Conclusion

We rigorously introduced propositional logic elements in the R. Thomas' framework in order to take into account available information concerning the cooperation or concurrency between genes or genes products acting on the same targets.

This idea is rather natural: R. Thomas introduced in [17] a notation that allows the representation of *several* actions of a *unique* gene on another one. Moreover, dozens of articles can be cited which use similar ideas in different frameworks:[2, 10], *etc.* Up to our knowledge, our contribution is the first one which rigorously *formalizes* this more elaborated framework.

The introduction of multiplexes makes models terser because this framework allows the gathering of edges into a single multiplex.

The major advantage of multiplex modelling is methodological: It reduces the number of parameters by formalizing additional biological information. So, the step which searches parameter values consistent with known or hypo-

thetical properties of the system is significantly improved. These advantages open perspectives to study larger gene regulatory networks.

Another advantage of multiplexes is to facilitate manipulations of networks. For example, we may develop graph folding methods in order to reduce the number of variables, at the price of possibly long formulas in multiplexes. However the role of some variables in a path is essentially to delay the global process. Consequently to improve the biological usefulness of such abstractions, it seems necessary to take delays into account. One of our future works will be to introduce delays in multiplexes.

References

[1] J. Ahmad, G. Bernot, J.-P. Comet, D. Lime, and O. Roux. Hybrid modelling and dynamical analysis of gene regulatory networks with delays. *ComPlexUs*, 3(4):231–251, 2006 Cover Date: November 2007).

[2] R. Albert and H. Othmer. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in drosophila melanogaster. *J. Theor. Biol.*, 223:1–18, 2003.

[3] G. Bernot, J.-P. Comet, A. Richard, and J. Guespin. Application of formal methods to biological regulatory networks: Extending Thomas' asynchronous logical approach with temporal logic. *J. Theor. Biol.*, 229(3):339–347, 2004.

[4] C. Chaouiya, E. Remy, P. Ruet, and D. Thieffry. Qualitative modelling of genetic networks: From logical regulatory graphs to standard Petri nets. In *ICATPN 2004*, LNCS 3099, pages 137–156. Springer-Verlag, 2004.

[5] F. Ciocchetta and C. Priami. Biological transactions for quantitative models. *ENTCS*, 171(2):55–67, 2007.

[6] J.-P. Comet, H. Klaudel, and S. Liauzu. Modeling multi-valued genetic regulatory networks using high-level Petri nets. In *ICATPN 2005*, volume 3536 of *LNCS*, pages 208–227, 2005.

[7] H. de Jong. Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.*, 9(1):67–103., 2002.

[8] S. Fujita, M. Matsui, H. Matsuno, and S. Miyano. Modeling and simulation of fission yeast cell cycle on hybrid functional Petri net. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E87-A(11):2919–2928, 2004.

[9] F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of protein. *Journal of molecular biology*, 3:318–356, 1961.

[10] S. Klamt, J. Saez-Rodriguez, J. A. Lindquist, L. Simeoni, and E. D. Gilles. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 7:56, 2006.

[11] A. Richard. *Modèle formel pour les réseaux de régulation génétique et influence des circuits de rétroaction*. PhD thesis, Univ. Evry Val d'Essonne, 2006.

[12] H. Siebert and A. Bockmayr. Context sensitivity in logical modelling with time delays. In LNBI, editor, *Computational Methods in Systems Biology, CMSB 2007*, volume 4695, pages 64–79. Springer, 2007.

[13] P. Smolen, D.A. Baxter, and J.H. Byrne. Modeling transcriptional control in gene networks: Methods, recent results, and future directions. *Bulletin of Mathematical Biology*, 62(2):247–292, 2000.

[14] E.H. Snoussi. Necessary conditions for multistationarity and stable periodicity. *Journal of Biological Systems*, 6:3–9, 1998.

[15] O. Tardieu and R. de Simone. Curing schizophrenia by program rewriting in Esterel. *Formal Methods and Models for Co-Design*, pages 39–48, 2004.

[16] R. Thomas. Boolean formalization of genetic control circuits. *J. Theor. Biol.*, 42:563–585, 1973.

[17] R. Thomas. Regulatory networks as seen asynchronous automata: A logical description. *J. theor. Biol.*, 153:1–23, 1991.

[18] R. Thomas and R. d'Ari. *Biological Feedback*. CRC Press, 1990.

[19] R. Thomas, D. Thieffry, and M. Kaufman. Dynamical behaviour of biological regulatory networks. *Bull. Math. Biol.*, 57(2):247–276, 1995.

[20] D.J. Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman & Hall/CRC, 2006.

LIST OF ATTENDEES

(February 23rd, 2009)

AMAR	Patrick (pa@lri.fr)
BALLET	Pascal (pascal.ballet@univ-brest.fr)
BARRETT	Steven (steven.j.barrett@gsk.com)
BARBRY	Pascal (barbry@ipmc.cnrs.fr)
BERMUDEZ	Olga (bermudez@unice.fr)
BERNOT	Gilles (bernot@unice.fr)
BEURTON-AIMAR	Marie (beurton@labri.fr)
CLÉMENT	Frédérique (Frederique.Clement@inria.fr)
COZE	Fabien (fabien.coze@igmors.u-psud.fr)
DEL REAL	Alejandro (hadeshell@gmail.com)
FICHERA	Davide (fichera.davide@gmail.com)
FOURMENTIN	Éric (fondation@fourmentinguilbert.org)
GANDRILLON	Olivier (gandrillon@cgmc.univ-lyon1.fr)
GUESPIN	Janine (janine.guespin@wanadoo.fr)
KAUFMAN	Marcelline (Marcelle.Kaufman@ulb.ac.be)
KÉPÈS	François (Francois.Kepes@epigenomique.genopole.fr)
KEPSEU DJAMPOUOP	Wilfred (kepseuwilfred@yahoo.fr)
LE GALL	Pascale (pascale.legall@epigenomique.genopole.fr)

GOUZÉ	Jean-luc (gouze@sophia.inria.fr)
LIVOLANT	Françoise (livolant@lps.u-psud.fr)
LOBRY	Claude (claude.lobry@inria.fr)
MAZAT	Jean-Pierre (jpm@u-bordeaux2.fr)
NORRIS	Vic (victor.norris@univ-rouen.fr)
OVADI	Judit (ovadi@enzim.hu)
PARISEY	Nicolas (parisey@labri.fr)
PROST	Jacques (jacques.prost@espci.fr)
REICHL	Udo (ureichl@mpi-magdeburg.mpg.de)
RIVAS	German (grivas@cib.csic.es)
RUET	Paul (ruet@iml.univ-mrs.fr)
SARDET	Christian (christian.sardet@obs-vlfr.fr)
SÀNCHEZ JIMÉNEZ	Francisca (kika@uma.es)
TAHI	Fariza (fariza.tahi@ibisc.univ-evry.fr)
THOMAS	Randy S. (randy.thomas@ibisc.univ-evry.fr)
VALLÉE	François (vallee@labri.fr)
VIROLLE	Marie-Joëlle (marie-joelle.virolle@igmors.u-psud.fr)
ZEMIRLINE	Abdallah (zemirlin@univ-brest.fr)