Proceedings of
The Bordeaux Spring School
on

Modelling Complex Biological Systems in the Context of Genomics

April 3th - 7th, 2006

Edited by

Patrick Amar, François Képès, Vic Norris
Marie Beurton-Aimar, Jean-Pierre Mazat

"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 five subgroups. The *membranes and intracellular structures* subgroup focuses on membrane deformations involved in the functionning of the Golgi, in cell division or in attachment to surfaces, on the dynamics of the cytoskeleton, and on the dynamics of *hyperstructures* (which are extended multi-molecule assemblies that serve a particular function). The *organisation* subgroup has adopted a systems biology approach with the application and development of new programming languages to describe biological systems which it has been applying to problems in the growth and differentiation of plants and in the structure and functioning of mitochondria. The *observability* subgroup addresses the question of which models are coherent and how can they best be tested by applying a formal system, originally used for testing computer programs, to an epigenetic model for mucus production by *Pseudomonas aeruginosa*, the bacterium involved in cystic fibrosis. The *G cube (Genomic Graphonomy Group)* subgroup works on networks of molecular interactions. Questions pertaining to the topology, dynamics and partitioning of molecular networks, and statistical inference of networks from post-genomic data, are discussed on a regular basis. The *bioputing* group works on new approaches proposed to understand biological computing using computing machine made of biomolecules or bacterial colonies.

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

This book gathers overviews of the talks, discussions and roundtables, original articles contributed by speakers, and abstracts from attendees. We thank the sponsors of this 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, Marie Dutreix, Jean-Louis Giavitto, Christophe Godin, Janine Guespin, François Képès, Jean-Pierre Mazat, Franck Molina, Victor Norris, Vincent Schächter, Philippe Tracqui.

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 group 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 *RelaiSoleil village "Les Bruyères"* at Carcans Maubuisson 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.

We would like to thank Paul Hossenlopp (Responsable de la formation, CNRS) for his encouragements to this 5th edition of the spring school.

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:

- Genopole® Evry: http://www.genopole.fr
- Centre National de la Recherche Scientifique (CNRS): http://www.cnrs.fr
- Institut National de Recherche en Informatique et en Automatique (INRIA): http://www.inria.fr
- Institut National de la Santé et de la Recherche Médicale (INSERM): http://www.inserm.fr
- Laboratoire de Méthodes Informatique (LaMI), Université d'Evry Val d'Essonne: http://www.lami.univ-evry.fr
- Fondation Scientifique Fourmentin-Guilbert: http://www.fourmentinguilbert.org
- Ministère de l'Éducation Nationale Enseignement Superieur Recherche: http://www.education.gouv.fr
- Universités de Bordeaux-1 et Bordeaux-2: http://www.u-bordeaux1.fr - http://www.u-bordeaux2.fr
- Institut de Mathématiques de Bordeaux (IMB): http://math.u-bordeaux.fr
- Région Aquitaine: http://aquitaine.fr

THE EDITORS

CONTENTS

Petri nets for qualitative modelling of biological networks

C. Chaouiya[1]

[1] IBDML, Institut de Biologie du Développement de Marseille-Luminy
UMR 6216, Case 907 - Parc Scientifique de Luminy
13288 Marseille Cedex 9 - France

Abstract

After an introduction to the basics on Petri nets (PNs), I present a brief overview of their use for the modelling of biological networks, in particular for the qualitative modelling of metabolic networks by means of standard PNs. At a higher level of abstraction, the genetic regulatory networks are qualitatively modelled using the logical approach initially introduced by R. Thomas. I summarise the proposal of our group for a standard PN based representation of logical regulatory graphs.

1 Introduction

Given the huge amount of genetic and molecular data, there is a pressing need for mathematical tools to develop dynamical models which integrate these data and provide means to understand the dynamical behaviour of biological systems. It is now commonly agreed that most biological functions are directed by complex networks of genes, proteins and biochemical reactions.

One can distinguish two main classes of dynamical models: on the one hand *quantitative* models, essentially based on systems of differential equations and, on the other hand, *qualitative* models defined through discrete formalisms or piecewise linear differential systems (for a review see [6]). Quantitative methods aim at representing the system in a detailed way and produce quantified results. They require accurate kinetic data, which are generally lacking and, because of the size and the preciseness of the models, most of the results are obtained by numerical integration methods. These do not capture general and systematic information on the properties of the models under study. Therefore they are helpfully complemented by qualitative approaches which are often more suitable for the induction of dynamical properties of systems which are too complex or for which few data are accessible. Indeed, qualitative modelling should allow the investigation of all relevant situations (because of a restricted number of parametrical values, and/or of qualitatively different situations). However, qualitative approaches still confront problems of a combinatorial nature, but some formalisms provide analytical methods which can circumvent this difficulty, at least partially (for example, the regulatory circuit analysis in the logical modelling [23]).

Petri nets (PNs) and their various extensions allow the definition of both qualitative and quantitative models. They have recently emerged as a promising tool among the various methods employed for the modelling and analysis of biological networks. This document constitutes a preliminary version of a forthcoming survey on PN modelling of biological networks. Section 2 comprises a general introduction to PN basics. Section 3 follows with a brief overview of PN modelling of biological networks, including the qualitative modelling of biochemical networks. Then, in Section 4, the proposal of our group for a standard PN based representation of logical regulatory graphs is summarised. Finally, Section 5 provides some discussion and prospects.

2 Petri net basics

Petri nets define a graphical and mathematical formalism suitable for the modelling and the analysis of concurrent discrete events dynamical systems. A PN is a directed bipartite graph with

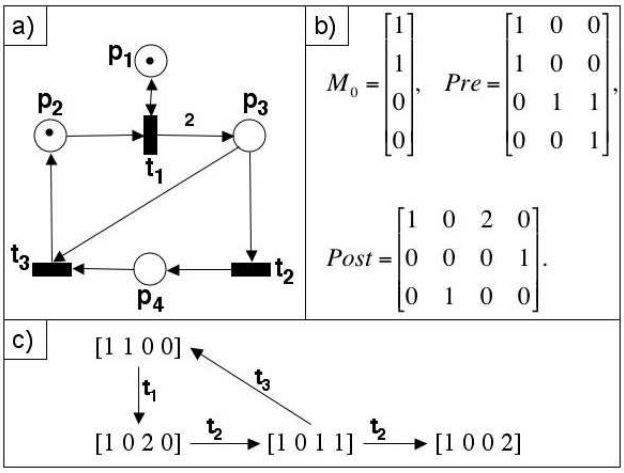


Figure 1: a) An example of Petri net; b) The initial marking M_0 and matrices Pre and $Post$; c) The corresponding marking graph $\mathcal{R}(M_0)$.

weighted arcs connecting *places* with *transitions*. At any time of the evolution of a PN, its places hold zero or a positive number of *tokens*. The state of the system is represented by the allocation of tokens over the places and is called a *marking*. The arcs connected to a transition define its input places and output places. More formally,

Definition 1 *A Petri net is a 5-tuple* $< P, T, Pre, Post, M_0 >$*, where:*
P *is a finite set of places,*
T *is a finite set of transitions, with* $P \cap T = \emptyset$ *and* $P \cup T \neq \emptyset$*,*
$Pre : P \times T \rightarrow \mathbb{N}$ *defines weighted arcs between places and transitions,*
$Post : T \times P \rightarrow \mathbb{N}$ *defines weighted arcs between transitions and places,*
$M_0 : P \rightarrow \mathbb{N}$*, is the initial marking (which associates an integer number of tokens to each place).*

An example of PN is given in Figure 1.a, together with its initial marking M_0 and matrices Pre, $Post$ (Figure 1.b). Graphically, places are depicted by circles, while rectangles represent transitions. If $Pre(p,t) \neq 0$ (respectively $Post(t,p) \neq 0$), it is represented by a weighted arc from p to t (resp. from t to p) and its weight is the value of $Pre(p,t)$ (resp. $Post(t,p)$). The weight is omitted when its value is 1. The black dots (tokens) represent the marking; within the initial marking M_0 which puts one token in places p_1 and p_2, transition t_1 is enabled. The dynamical behaviour from the initial marking is described by the *marking graph* in terms of transitions between states (markings) and is denoted $\mathcal{R}(M_0)$ (Figure 1.c).

Formally, the firing rules define the enabling of transitions and the evolution of the markings:

Definition 2 *A transition* $t \in T$ *is enabled by a marking* M *if:* $\forall p \in P, M(p) \geqslant Pre(p,t)$*. An enabled transition* t *may fire, and its firing then modifies the marking of its input and output places: it removes* $Pre(p,t)$ *tokens from each input place* p *of* t*, and adds* $Post(t,p')$ *tokens to each output place* p'*.*

One further defines the *incidence matrix* as $C = Post^T - Pre$. Then, the *state equation* defines the kth marking obtained from marking M_{k-1}, after the firing of t_i (and thus, we suppose that $M_{k-1}(p) \geqslant Pre(p,t_i), \forall p \in P$):

$$M_k = M_{k-1} + C.u_k\,, \tag{1}$$

where u_k is the firing unit vector with $|T|$ components, all being zero but the ith position being equal to one, indicating that transition t_i fires at the kth firing.

For example, in the Figure 1, $M_0 = [1, 1, 0, 0]^T$ enables transition t_1; the firing of t_1 leads to the new marking $M_1 = [1, 0, 2, 0]^T$, which in turn enables t_2.

Note that a side-condition is modelled with a loop (often depicted as a bidirectional arc) as shown in Figure 1 where the firing of transition t_1 removes the token in place p_1 and restores it afterwards. Such arcs, called *test arcs*, may have a particular semantic in the case of *timed* PNs, where time-delays are attributed to the firings of transitions. In the case of a *pure* PN (with no loop), C^T is the classical incidence matrix of a directed graph.

With standard PNs as defined in what precedes, one can check main qualitative properties which basically can be checked using algebraic methods (with the state equation), analysing the structure of the net or investigating the whole marking graph [15]. Hereafter, the main properties are briefly described, together with their possible interpretation in the context of biological networks:
- *Boundedness* insures that, whatever the initial marking and the evolution of the net, the number of tokens in each place is bounded. For biochemical models, it means that no product can accumulate;
- *P-invariants* are sets of places for which the weighted sum of tokens is constant independently of the sequence of firings (x, a $|P|$-vector of integers defines a P-invariant if $C^T.x = 0$). In biological terms, these sets may define some conservation relations;
- *T-invariants* are firing sequences which would reproduce a marking (y, a $|T|$-vector of integers defines a T-invariant if $C.y = 0$). In biological terms, T-invariants may represent cyclical behaviours and can also be related to the 'elementary flux modes' defined by Schuster *et al.* [18].
- *Reachability* of a marking M defines that there exists an evolution (a sequence of firings) from the initial marking to the marking M (*i.e.*, M is in the marking graph $\mathcal{R}(M_0)$). This property may be relevant for biological networks, as it ensures the existence of a trajectory leading the system to a desired state.
- *Liveness* insures that, no matter the evolution of the net, it is always possible to ultimately fire any transition. There are other weakened classes of liveness. In other words, liveness is related to the guarantee that an event (a reaction for example) can eventually occur.
- *Reversibility* means that no matter the evolution of the net, it is always possible to ultimately reach the initial marking again.

Standard PNs have been extended to increase their expressiveness (in particular to allow quantitative analyses). In summary, among the main extensions, stochastic PNs include randomness (enabled transitions fire with exponentially distributed time delays); colored PNs distinguish different kinds of tokens; and hybrid PNs take into account both discrete and continuous processes. However, note that generally, the more expressive the formalism, the more difficult the analysis. In particular, the methods which are valid to check the above properties in standard PNs, are generally no more valid in the context of high level PNs.

3 A brief overview of Petri net modelling of biological networks

This section intends to give a flavour of the many ways in which Petri nets can be used to model biological networks. It is far from being an exhaustive review. In the sequel, I chose to classify the applications upon the kind of PNs they employ. Most of the PN models are related to metabolic networks. When gene regulatory networks are modelled, the mechanisms related to transcription and translation are represented (as, for example in [14]), with the noteworthy exception of *logical* regulatory graphs presented in Section 4.

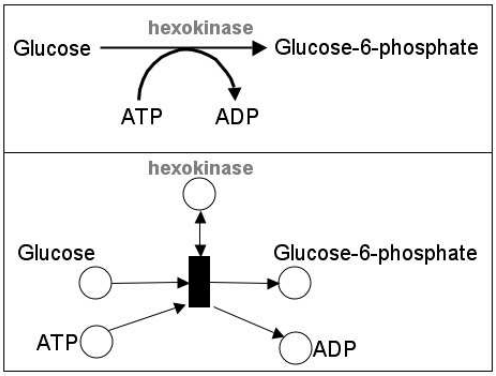


Figure 2: Top: an enzyme-catalyzed reaction. Bottom: the corresponding Petri net.

Standard Petri nets: An early attempt of PN modelling of biochemical reaction systems has been presented by V. Reddy and M. Mavrovouniotis [17]. In particular, they showed how PNs allow the representation of the essential components in biological pathways, and how such PN models can be used to perform a qualitative analysis. Metabolic pathways are generally seen as interconnected networks of enzymatic reactions, where the product of one reaction is a reactant of (or an enzyme that catalyzes) a subsequent reaction. Figure 2 illustrates the PN modelling of a catalyzed reaction with places representing reactants and products and transition representing the reaction.

Since then, it has been shown that several concepts arising in structural pathway analysis of biochemical networks have their counterparts in PN theory (*e.g.* [19]). In particular, it has been demonstrated that T-invariants correspond to 'elementary flux modes'. Indeed, the topology of the PN matches the topology of the metabolic network it represents, and one can draw extensive relationships between the traditional biochemical modelling and PNs (see *e.g.* [26]). In particular, the stoichiometry matrix of a metabolic network corresponds to the PN incidence matrix. In [11], the authors demonstrate how PNs provide a mean for model validation. All these qualitative analyses assume that the system has reached a steady state.

Stochastic Petri nets: SPNs associate random (exponentially distributed) time delays to transitions [13]. Stochastic models are useful to take into account uncertainty attached to data, but also to describe external noise (generated by fluctuations of the environment), or intrinsic noise (due to low molecular concentrations). So far, SPN models have mainly been used for stochastic molecular interactions [9, 22] (these models essentially reproduce the Gillespie's algorithm [7]).

Colored Petri nets: CPNs assign categories to the tokens defining *color sets*, thus allowing reduced models of complex systems (see [12] and references therein). In [25], the authors used colors to differentiate molecules of the same species (according to the paths along which they are produced and consumed) and performed a qualitative analysis of the erythrocyte combined glycolysis and pentose phosphate pathways (a refined version of the original model proposed by Reddy *et al.* in [17]).

CPN models of genetic networks are developped in [5], according to the logical approach developped by R. Thomas [23], providing means to check the model under various hypothesis, using model-checking techniques (see Section 4 for further details).

Hybrid Petri nets: HPNs allow the markings of places to take either (positive) discrete or continuous values [1]. They provide a good mean to represent protein concentration dynamics being coupled with discrete switches as shown in [14], where Matsuno *et al.* presented a HPN model

of the λ phage genetic switch mechanism. Since then, this group considered a further extension, called Hybrid Functional PNs and developed the dedicated software Genomic Object Net (GON[1]) [16].

In Functional PNs, primarily defined by R. Valk as *self-modified* PNs [24], the flow relations between places and transitions depend on the marking. Hofestädt and Thelen applied this extension for the quantitative modelling of biochemical networks [10].

4 Qualitative Petri net modelling of genetic networks

When a biological network has to be studied, it is crucial to consider the relevant level of abstraction, depending on the question to be addressed, but also on the knowledge and data which are available. In the case of the regulation of gene expressions, it is often sufficient to represent the fact that a particular regulatory product activates or inhibits a gene or set of genes to convey the role of this product in the network. In such networks, the semantics associated with the interactions between components varies: while in a chemical reaction the reactants are consumed, the expression levels of regulators do not change during the regulatory process. One successful approach to qualitatively model such regulatory networks is the *logical* approach initially developped by René Thomas and collaborators [23].

In a *logical* regulatory graph, the nodes represent genes which are associated with discrete levels of expression, and arcs represent interactions between genes. Each interaction is associated with a threshold from which the gene source of the interaction has an effect onto the targeted gene. For each gene, a discrete logical function defines to which qualitative level tends a gene when submitted to a given combination of interactions. GINsim is a software which implements this formalism and allows the user to define a model and analyse it[2]. Further details on the formalism and GINsim can be found in [8] and references therein.

Figure 3 illustrates our proposal for a standard PN representation of logical regulatory graphs, called *Multi-level Regulatory Petri Nets* (MRPNs). Note that this example is far from being realistic (in the top case, B would have no chance to reach its highest level). But the objective here is to illustrate the rules governing the rewriting. For further details and applications, see [2] for the Boolean case, and [3] for the multi-level case. This systematic rewriting has been implemented in GINsim and should be available soon. In [3] we demonstrated our rewriting rules for a multi-level logical model of the genetic switch controlling the lysis-lysogeny decision in the bacteriophage lambda (the PN model can be downloaded at the GINsim web site, as text files in the INA[3] format or in PNML).

We took advantage of the MRPNs to develop an integrated model of a biochemical pathway and its regulation [21]. More precisely, we defined a qualitative modelling of the biosynthesis of tryptophan (Trp) in *Escherichia coli*, taking into account two regulatory feedbacks: the direct inhibition of the first enzyme of the pathway by the final product of the pathway, and the transcriptional inhibition of the Trp operon by the Trp-repressor complex.

Recently, Comet *et al.* developed a correspondance between the logical approach and Colored PNs providing a compact CPN modelling (a unique place and a unique transition where the marking of the place represents the state of the whole system and the guard associated to the transition implements the logical rules) [5]. The objective here is to provide the biologist with a tool which systematically verifies the coherence of the model under various hypotheses (accounting for observed biological behaviours such as homeostasis and multistationarity, or even more precise temporal properties).

[1] http://www.genomicobject.net/

[2] GINsim web page: http://www.gin.univ-mrs.fr/GINsim/

[3] Integrated Net Analyzer, tool for the analysis of (Coloured) PNs: www.informatik.hu-berlin.de/s̃tarke/ina.html

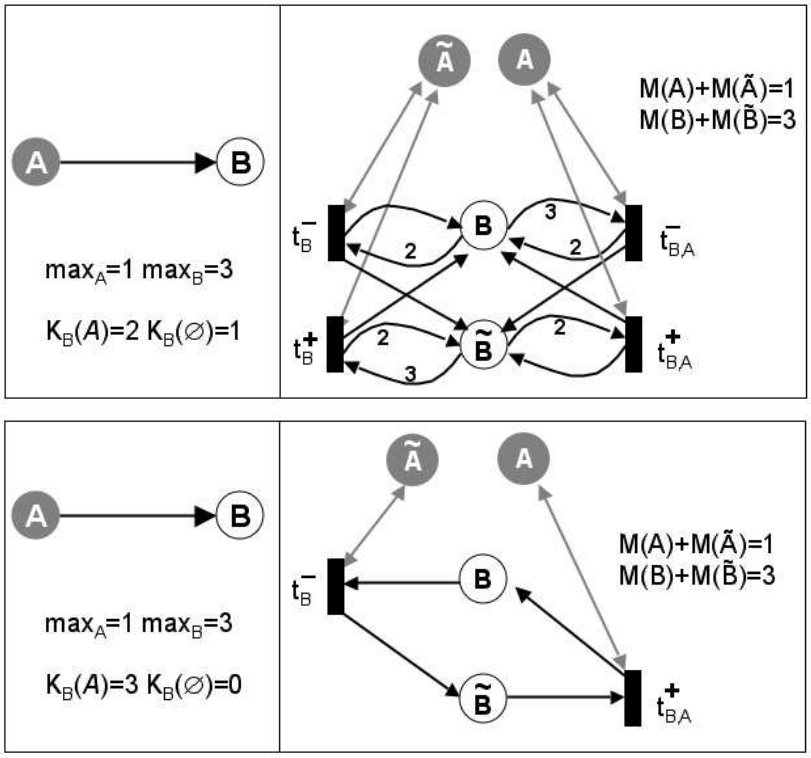


Figure 3: *Top-left:* a toy example with two regulatory products, A (2 qualitative levels) and B (4 levels) and an interaction from A to B. The effect of A alone onto B is described by the *logical parameter* $K_B(A) = 2$ (when A is present, B tends to level 2), while the *base value* of B denoted $K_B(\emptyset)$ (B submitted to no regulation) is 1. *Top-right:* the corresponding *Multi-level Regulatory Petri net*. For each regulatory product, two *complementary* places are defined (A and $\widetilde{A}$, B and $\widetilde{B}$). For each parameter, two transitions are defined (one for the increase order, the other for the decrease order). For instance, the parameter $K_B(A) = 2$ is represented by $t^+_{B,A}$ and $t^-_{B,A}$. If A is marked (product A at level 1) and B not marked (B at level 0, 3 tokens in $\widetilde{B}$), then $t^+_{B,A}$ is enabled and its firing leads to an increase of the marking of B; if A is marked and B has 3 tokens (its higher level), then $t^-_{B,A}$ is enabled and its firing leads to a decrease of the marking of B. *Bottom-left:* the same toy example, with other logical parameter values. *Bottom-right:* the corresponding MRPN is simpler as there is only one transition for each parameter (because the new parameter values are extremal, $t^-_{B,A}$ and t^+_B are of no use here).

5 Discussion

This document is certainly not exhaustive and only aims at giving a general view on PN modelling applied to biological networks. The increasingly use of PNs for the modelling of biological networks can be explained by their underlying graphical representation, their suitability to model concurrent distributed systems, their well founded mathematical theory, and the dedicated tools already available[4]. A variety of models have been already developed, from purely qualitative to sophiticated hybrid models. These different modelling approaches led to different kinds of analyses, from structural analyses to pure simulation. Some efforts still have to be done to unify the concepts, notations and terminologies [4].

In our group, we are mainly interested in the qualitative modelling of biological networks,

[4]see the Petri Nets Tool Database on the Petri Nets World web site:
http://www.informatik.uni-hamburg.de/TGI/PetriNets/tools/db.html

as they proved to be useful to get insights into the dynamical behaviour of the system. They are also valuable to further delineate more precise quantitative models. It has been demonstrated in the literature that PNs are well suited for the qualitative modelling of biochemical networks. By developing a PN representation of logical regulatory graphs, we have now a way to systematically and qualitatively model regulated metabolic networks by connecting the biochemical network model to the logical part of its regulation. This approach can now provide a good basis for more quantitative analysis, in particular taking advantage of stochastic extensions.

It is interesting to observe that PNs allow a step by step modelling procedure if one considers the refinements provided by the several existing extensions. Much work still has to be done to integrate PN models accounting for different levels of abstraction. PNs could be useful to face the problem of the composition of models. In this context, some groups are proposing to define PN unities or modules which would be the building blocks of biological network models. Finally, some attempts are made to map other existing formalims to PN models. More specifically it would be valuable to define a systematic correspondance between ODEs and PN models. In this respect, Shaw *et al.* propose a correspondance between SBML[5] and PNML[6] models [20].

References

[1] H. Alla and R. David. Continuous and hybrid petri nets. *Journal of Circuits, Systems and Computers*, 8(1):159–88, 1998.

[2] C. Chaouiya, E. Remy, P. Ruet, and D. Thieffry. Qualitative modelling of genetic networks: From logical regulatory graphs to standard petri nets. In *ICATPN'04*, volume 3099 of *Lecture Notes in Computer Science*, pages 137–56. Springer-Verlag, 2004.

[3] C. Chaouiya, E. Remy, and D. Thieffry. Petri net modelling of biological regulatory networks. Technical report, IML, 2005. PrePrint.

[4] M. Chen and R. Hofestädt. Quantitative petri net model of gene regulated metabolic networks in the cell. *In Silico Biol.*, 0030, 2003.

[5] JP. Comet, H. Klaudel, and S. Liauzu. Modeling multi-valued genetic regulatory networks using high-level petri nets. In P. Darondeau G. Ciardo, editor, *ICATPN'05*, volume 3536 of *Lecture Notes in Computer Science*, pages 208–27. Springer-Verlag, 2005.

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

[7] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J.Phys.Chem.*, 81:2340–61, 1977.

[8] A. González, A. Naldi, L. Sanchez, D. Thieffry, and C. Chaouiya. Ginsim: A software suite for the qualitative modelling, simulation and analysis of regulatory networks. *Biosystems*, in press.

[9] PJ. Goss and J. Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets. *Proc Natl Acad Sci U S A.*, 95(12):6750–5, 1998.

[10] R. Hofestädt and S. Thelen. Quantitative modeling of biochemical networks. *In Silico Biol.*, 1(1):39–53, 1998.

[5] Systems Biology Markup Language web site: sbml.org

[6] Petri Net Markup Language web site: www.informatik.hu-berlin.de/top/pnml

[11] I. Koch, BH. Junker, and M. Heiner. Application of petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. *Bioinformatics*, 21(7):1219–26, 2005.

[12] L.M. Kristensen, S. Christensen, and K. Jensen. The practitioner's guide to coloured petri nets. *Int Journal on Software Tools for Technology Transfer*, 2(2):98–132, 1998.

[13] M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. *Modelling with Generalized Stochastic Petri Nets*. John Wiley, 1994.

[14] H. Matsuno, A. Doi, M. Nagasaki, and S. Miyano. Hybrid petri net representation of gene regulatory network. In *Pac Symp Biocomput.*, pages 341–52, 2000.

[15] T. Murata. Petri nets: Properties, analysis and applications. In *Proceedings of the IEEE*, pages 541–580, April 1989.

[16] M. Nagasaki, A. Doi, H. Matsuno, and S. Miyano. Genomic object net: a platform for modeling and simulating biopathways. *Applied Bioinformatics*, 2:181–184, 2004.

[17] V.N. Reddy, M.N. Liebman, and M.L. Mavrovouniotis. Qualitative analysis of biochemical reaction systems. *Comput Biol Med.*, 26(1):9–24, 1996.

[18] S. Schuster, DA. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol.*, 18(3):326–32, 2000.

[19] S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to mycoplasma pneumoniae. *Bioinformatics*, 18:351–61, 2002.

[20] O.J. Shaw, A. Koelmans, L.J. Steggles, and A. Wipat. Applying petri nets to systems biology using xml technology. In *Proceedings of the Workshop on the Definition, Implementation and Application of a Standard Interchange Format for Petri Nets*, pages 11–25, 2004.

[21] E. Simao, E. Remy, D. Thieffry, and C. Chaouiya. Qualitative modelling of regulated metabolic pathways: application to the tryptophan biosynthesis in e.coli. *Bioinformatics*, 21(21 Suppl 2):190–6, 2005.

[22] R. Srivastava, M. S. Peterson, and W. E. Bentley. Stochastic kinetic analysis of the escherichia coli stress circuit using σ^{32}-targeted antisense. *Biotechnol Bioeng*, 2001.

[23] R. Thomas, D. Thieffry, and M. Kaufman. Dynamical behaviour of biological regulatory networks–i. biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull Math Biol.*, 57(57):247–76, 1995.

[24] R. Valk. Self-modifying nets, a natural extension of Petri nets. In G. Ausiello and C. Böhm, editors, *ICALP'78*, volume 62 of *Lecture Notes in Computer Science: Automata, Languages and Programming*, pages 464–76. Springer-Verlag, 1978.

[25] K. Voss, M. Heiner, and I. Koch. Steady state analysis of metabolic pathways using petri nets. *In Silico Biol*, 3(3):367–87, 2003.

[26] I. Zevedei-Oancea and S. Schuster. Topological analysis of metabolic networks based on petri net theory. *In Silico Biol.*, 1, 2003.

Machine Learning Biochemical Networks from Temporal Logic Properties

L. Calzone[1], N. Chabrier-Rivier[1], F. Fages[1], S. Soliman[1]

[1] Projet Contraintes, INRIA Rocquencourt, France

Abstract

One central issue in Systems Biology is the definition of formal languages for describing complex biochemical systems at different levels of abstraction. In this talk, we show how temporal logic, possibly with numerical constraints, can be used to formalize the experimental knowledge from which biological models are built. In return, the temporal logic properties constitute a specification which can be checked automatically with model-checking techniques, or used to curate the model with machine learning techniques. We present two algorithms for inferring reaction rules and kinetic parameter values from a temporal specification formalizing the biological experiments. We illustrate how these machine learning techniques, implemented in the Biochemical Abstract Machine BIOCHAM, can be useful to the modeler through an example on the cell cycle control.

References

[1] Chabrier-Rivier, N., Fages, F., Soliman, S., (2003-2006) BIOCHAM's user manual, `http://contraintes.inria.fr/BIOCHAM/`.

[2] Fages, F., Soliman, S., Chabrier-Rivier, N., (2004) Modelling and Querying Interaction Networks in the Biochemical Abstract Machine BIOCHAM, *Journal of Biological Physics and Chemistry* **4:2**, 64-73.

[3] Calzone, L., Chabrier-Rivier, N., Fages, F., Soliman, S., (2006) Machine learning biochemical networks from temporal logic properties, *Transactions on Computational Systems Biology*, to appear.

En route to semi-synthetic minimal cells

P L Luisi,[1] P Stano,[2] G Murtas,[2] Y Kuruma,[2] and T Ueda[3]

[1] Biology Department, University of RomaTre, Viale G. Marconi, 446; 00146 Rome, Italy
[2] "Enrico Fermi" Centre, Compendio del Viminale; 00184 Rome, Italy
[3] Department of Medical Genome Science, Graduate School of Frontier Sciences, University of Tokyo, Bldg. FSB-4015-1-5 Kashiwanoha, Kashiwa, Chiba Prefecture 277-8562, Japan

Abstract

In this programmatic paper, we present a possible road map to the realization of minimal cells. Starting from consideration of the state-of-the-art in the use of lipid vesicles as compartments for biochemical reactions, we describe the most significant experimental steps that are needed for the realization of self-reproducing cells. This approach consists in the controlled assembly of informational (DNA) and functional (enzymes, ribosomes) macromolecules in a vesicle compartment, in order to construct a semi-synthetic minimal cell. Emphasis is given to the use of reconstituted (i.e., from purified components) biochemical machineries for the achievement of cellular functions. Synthetic biology, a new emerging discipline that is recently attracting considerable scientific interests, represents the framework for the development of this research program.

1 Premise

The question "what is life?", and the idea of constructing simple forms of life in the laboratory, have been in the agenda of science for a very long time. The scientific analysis of life is made difficult by the many layers of believes, biases and prejudices that have accumulated over the centuries over this term; in order to tackle the question in a proper scientific way, we have to go back to the roots of simplicity, and consider the simplest level of natural life, at the level namely of microbes and simpler unicellular organisms.
We have therefore to recognize that there is no other form of life on earth other than the cellular life. This means that life is based on compartments that permit a high local concentration of reagents, protection from the environment, containment and control of structural changes. These compartments are defined by semi-permeable membranes that permit the selection of chemicals as nutrients or other co-adjuvant of the cell functions.
A first assessments to the question "what is life", as well as the approach to the construction of simple life forms in the laboratories, have to start from such basic considerations.

2 The notion of minimal cell

The choice to look at unicellular organisms as our models for understanding life certainly simplify the analysis. However, when we look inside the cell of even very simple microbes, we are bewildered by a tremendous complexity: even the smallest unicellular organisms contains many hundreds of genes, and contains therefore several hundreds of enzymes, and families of nucleic acids, and a total of several thousand of reactions occurring inside each tiny compartment.
But precisely this staggering complexity elicits the question, whether all this is really necessary for life, or whether instead cellular life can also be realized by a much smaller degree of complexity.

The rational for this statement comes from two different kinds of considerations: on the one hand, the biochemical complexity of a cell is certainly in large part the result of millions of years of evolution based on competition struggles, that produced plenty of defence mechanisms, enzymes and nucleic acid redundancies, security loops and the development of a series of reactions that would have probably not occurred in a more permissive environment; the other argument is based simply on the consideration that the first early cells, that started the origin of life, could not have been possibly so complex from the very start, they must have been, conceivably, much simpler.
All this brings to the notion of minimal cell: this is the cell containing the minimal and sufficient number of components to be defined as alive. What does "alive" means? The definition of life is a complex matter, but at this stage we can accept a very general description: cellular life implies the concomitance of three properties: self-maintenance (metabolism), self-reproduction, and evolvability. When these three properties are all realized and work simultaneously, we will have full-fledged cellular life. In very early cells, as well as in synthetic constructs, we may have only two out of three properties working, and thus we will have several kinds of approximations to life, various forms of "limping" life. These approximations are very important for understanding the origin and development of cellular life-as again, one cannot assume that life started on Earth as the perfect machinery that is now. It is also clear that there will not be just one type of minimal cell, but instead the term minimal cell describe a large family of possibilities.
All these considerations permit to conceive a research line that focuses on the construction and study of minimal artificial constructs that are designed and realized with the aim of achieving a minimal living cell.
In the following, we discuss some aspects related to the minimal cell project, as sketched in Figure 1. Such topics will be not discussed separately but within a general discourse on our experimental approach.

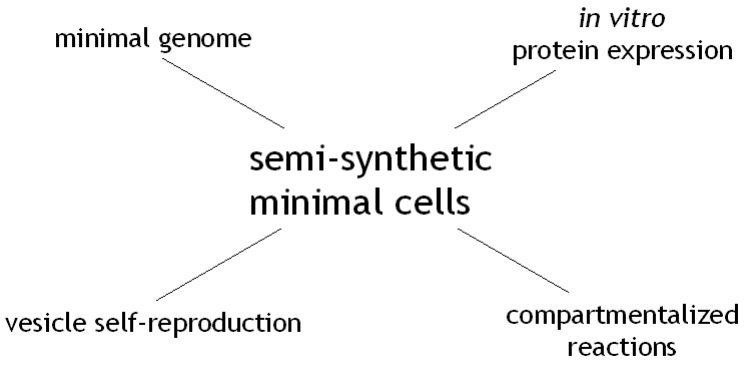


Figure 1. Some important issues related to the concepts and realization of semi-synthetic minimal cells.

3 The approach to the construction of the minimal cell: literature stand

The first thing needed for implementing in the laboratory the construction of a minimal cell is a suitable compartment. Lipid vesicles (liposomes) represent a proper model for cellular membrane. The strategy consists in the insertion of the minimal number of extant genes and enzymes inside a synthetic vesicle, so as to have an operational viable cell. This is represented in Figure 2.

Of course, from a certain point of view, one may say that is not fair to utilize extant macromolecules for studying the origin of cellular life. The point of this investigation however is not to clarify the origin of macromolecular sequences and life in general – but to focus on the earliest forms of cellular life, at a time in which such functional macromolecules were already present – and to have under laboratory control simple forms of cellular life; and to qualify the lowest degree of complexity compatible with cellular life. The fact that for this kind of construct natural macromolecules are used, brings to the terminology of "semi-synthetic minimal cells", meaning that part of this construct is synthetic (the membrane, and the assemblage), whereas some other parts (enzymes and nucleic acids) are natural.

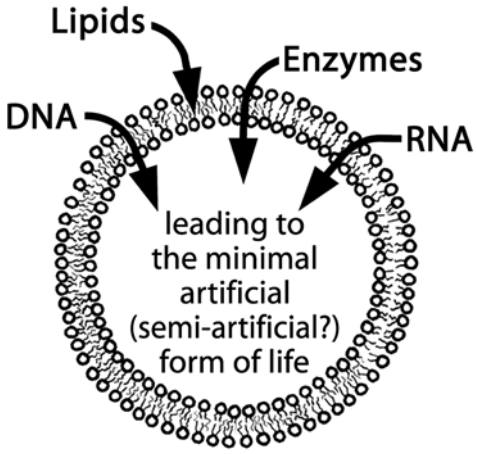


Figure 2. The construction of semi-synthetic minimal cells.

With these premises, the next point is to clarify what do we mean by "a minimal number of extant genes and enzymes"; the question of *minimal genome* is strictly connected to this issue. In order to be considered alive, a cellular construct must satisfy some conditions, that have been indicated in section 2. These theoretical considerations find a realistic scenario in the studies of several researcher that point out what is the minimal number of genes in a living organism. Thanks to the available data on minimal genome it is possible to evaluate a minimal gene-set of 200-300 genes. The interested reader can refer to a recently published review [1]. Here we would like to comment one of the most recent published study, a contribution of the Moya's lab (University of Valencia), who, on the basis of a comparative and functional approach, indicates 206 genes as the genomic core required for minimal living forms. The analysis was carried out by a comparative approach that considers the genomes of five endosymbionts and other microorganisms [2]. Such small number of genes codifies for the proteins that perform the essential cell functions, as basic metabolism and self-reproduction.

A pictorial representation of the Moya's minimal genome is shown in Figure 3.

Notice first of all the large number of genes that are associated to RNA metabolism, such as genes involved in the transcription, t-RNA synthesis and modifications, rRNA synthesis, ribosome-related functions, translation factors and RNA degradation. In other words, the biochemical machinery of a minimal cell is strongly biased to the transcription-translation processes. Of course, also other important processes are implemented by the minimal genome: protein processing, cellular processes, energetic pathways and – of course – DNA replication.

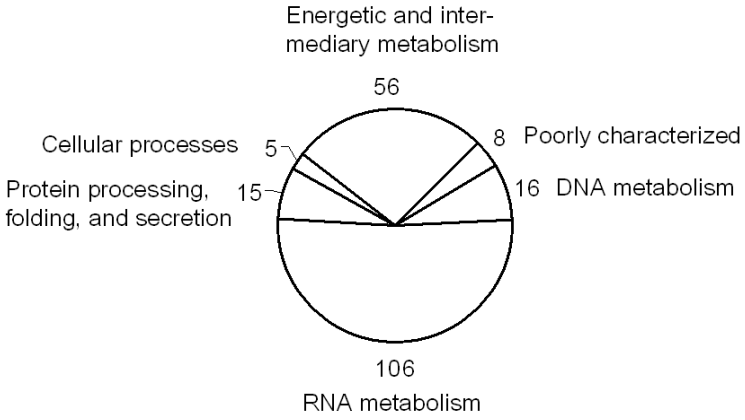


Figure3. The functional classification of a minimal genome composed of 206 genes. Data from reference [2].

When we consider the construction of artificial minimal cells, there are some additional points that can reduce the number of genes. First of all, the appropriate choice of "environment" conditions, i.e. the presence in the medium of specific compounds. In fact, the biosynthesis of low-molecular weight compounds and the cellular requirements for *in situ* energy production can be bypassed by an external supply of such compounds. This will correspond to an extremely favourable environment, that considerably simplifies our experimental approach and does not affect the definition of minimal life. In fact, the very blueprint of cellular life is not related to the synthesis of low molecular weight compounds but to the synthesis and replication of high molecular weight (and information carrying), as well as to the reproduction of the entire cellular construct. In addition, fine regulation of cellular processes are often characterized by a complex feed-back mechanisms that involve also the production of small molecules. At the present level of experimental complexity, these kind of loop are essentially neglected, since we are focusing on simplified cells, as discussed above in the section 2.

A survey of literature data indicates that the large majority of the researchers in this field have approached the construction of minimal semi-synthetic cells by performing simple and complex compartmentalized biochemical reactions into compartments. At this aim, vesicles are formed in an aqueous medium that contains all the reactants for a determined reaction, or – in an alternative approach – a substrate is externally added to a pre-formed and pre-loaded vesicles. Both the strategies correspond to the exploitation of passive compartmentation ability of vesicles, since the solutes are statistically entrapped in the inner aqueous volume of vesicles during the process of vesicle formation. As a consequence, the external (in bulk) reaction must be somehow inhibited. In contrary, an active process of entrapment would consists in the direct injection of the required reagents into a vesicle, that in this case must be necessarily a giant vesicle [3].

Pioneering work (see Table 1) on compartmentalized reactions was done in the Nineties by Oberholzer, Walde and Luisi at the ETH, in Zurich. A recently published review describes this work in detail [1].

Let us discuss now the most recent advancements in the works on minimal cell. In the last years, several papers have appeared in the literature, describing the insertion of complex biochemical systems in vesicles. Several groups all around the world, in particular those lead

by T. Yomo [4,5], T. Ueda [6-8], and T. Sugawara [9] in Japan, P. L. Luisi in Italy [10-19], D. Deamer [20,21], S. Rasmussen [22], as well as A. Libchaber [23,24] in the United States, E. Szathmáry in Hungary [25], G. Ourisson and Y. Nakatani in France [26,27]. In particular it has been possible to express some proteins by inserting the whole ribosomal synthetic apparatus into vesicles. Some reviews have already made their appearance in the literature [1,19,21,24,25], while the most representative results have reported in Table 1.

Table 1. Compartmentalized reactions in lipid vesicles.

Nucleic Acid Metabolism	Ref.	Protein Expression	Ref.
Enzymatic poly(A) synthesis inside vesicles	[28,29]	Ribosomal synthesis of polypeptides into vesicles	[14]
RNA replication into vesicles	[13]	GFP production into vesicles	[4]
DNA amplification by the PCR inside the liposomes	[12]	Expression of EGFP into small vesicles	[17]
Transcription of DNA in giant vesicles	[26]	EGFP expression into giant vesicles	[27]
Production of mRNA inside giant vesicles	[3]	Encapsulation of a genetic network into vesicles	[5]
		Expression of two proteins into a long-lived giant vesicle	[23]

From all these data – especially from those concerning the expression of protein within vesicles, (Figure 4 and right-hand column of Table 1) it is possible to depict the state of art of the field, also evidencing the difficulties, and possibly to design future working plans.
One first, important limit in all these studies is given by the fact that in order to express a protein inside the vesicles (mostly the green fluorescence protein, GFP), entire commercial kits have been used. Those commercial preparations are usually "black boxes", in the sense that the composition of the mixture is not given, and therefore also the number and the relative concentration of the enzymes involved is unknown. Since all these preparations are cellular extracts (from *E. coli*), a rigorous synthetic approach is not possible.

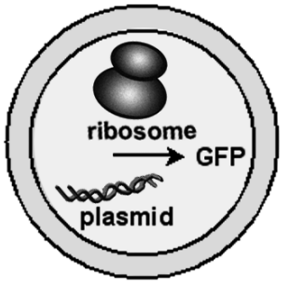


Figure 4. Schematic drawing representing the expression of green fluorescent protein (GFP) into lipid vesicles. In particular, a plasmid that contains the gene of interest is co-entrapped into the aqueous core of a vesicle together with all the components of the protein expression machinery (RNA polymerase, ribosomes, tRNAs, aa-tRNA synthetase, amino acids, ATP, GTP, etc.). The production of GFP is generally followed by batch fluorimetry, confocal microscopy or flow cytometry.

Another limit, clearly arising from the analysis of literature data, is that only proteins have been expressed, which corresponds to an active internal metabolism; however there is no result pointing to self-reproduction of these semi-synthetic cells.
Our work stems from the effort of overcoming these two important limits of present day literature. In particular, we wish to carry out the research by employing a minimal, and well known, set of enzymes, and focusing in the self-reproduction.

4 Our approach to the construction of the semi-synthetic minimal cells

Few years ago, the group of Takuya Ueda at the Tokyo University reported on the creation of a new *in vitro* protein expression kit composed by purified components [6]. This kit, now commercially available with the trademark of PURESYSTEM® (Post Genome Institute Co., Ltd. – Japan), is composed by 36 purified enzymes and ribosomal components; each one of these compounds is present at a known concentration, see Table 2 [8]. Originally developed to synthesize proteins *in vitro*, this tool appears to be perfectly suitable for a pure synthetic biology approach to the minimal cell projects. In fact, it allows the construction of an artificial cell that contains a minimal number of components in order to perform some function, i.e., the synthesis of a functional protein – in this case. From the point of view of minimal genome, taking into account the ribosomal proteins (each one having a corresponding gene) and the t-RNAs/r-RNAs, it can be concluded that by using PURESYSTEM a genome of about 100 genes is actually involved a protein-expressing minimal cell.

With this premises, it is evident that the next relevant steps in minimal cell research deal with the insertion of PURESYSTEM inside vesicles, therefore allowing the protein synthesis with a controlled (and subject to modulation) expression tool. In further steps, this cellular construct must be oriented to the production of lipids, in order to achieve a vesicle self-reproduction. Finally, we intend to reach the complete reproduction (core-and-shell) of such construct. The road map for this approach is illustrated below in some detail.

Table 2. Composition of the PURESYSTEM (adapted from [8])

Translation factors		*aa-tRNA synthetases*		*Other components*	
1	IF1	17	AlaRS	37	ribosomes
2	IF2	18	ArgRS		
3	IF3	19	AsnRS	*Energy sources*	
4	EF-G	20	AspRS	38	ATP, GTP,CTP,UTP
5	EF-Tu	21	CysRS	39	creatine phophate
6	EF-Ts	22	GlnRS		
7	RF1	23	GluRS	*Buffers/others*	
8	RF2	24	GlyRS	40	HEPES-KOH pH 7.6
9	RF3	25	HisRS	41	K glutammate
10	RRF	26	IleRS	42	Mg acetate
		27	LeuRS	43	spermidine
Other enzymes		28	LysRS	44	DTT
11	MTF	29	MetRS	45	20 amino acids
12	creatine kinase	30	PheRS	46	20 tRNA mix
13	myokinase	31	ProRS	47	formyl-tetrahydrofolic acid
14	pyrophosphatase	32	SerRS		
15	nucleosidePP kinase	33	ThrRS		
16	RNA polymerase	34	TrpRS		
		35	TyrRS		
		36	Val RS		

4.1 Incorporation of PURESYSTEM into liposomes, and optimisation of the GFP expression inside them.

The main target in this phase is to put under control a PURESYSTEM-containing liposome system that is capable of expressing one standard protein in optimal yield and in relatively short time. The novelty in this approach will be: (*a*) the expression of a protein with a purified and reconstituted transcription-translation machinery; (*b*) the consideration of the feeding problem, i.e. the continuous renewal of external phase (biochemically rich phase), by solving the problem of membrane permeability; (*c*) the possible removal of some components of the PURESYSTEM in order to further reduce the complexity of the system (in this case, a lower protein yield is expected). The point (*b*) of the list deserves a special comment. In fact, following the report of Noireaux and Libchaber [23], the introduction of α-hemolysin pores into vesicles allows the free entrance of externally given small substrates – as amino acids, ATP, etc – into the internalised water pool of liposomes. If this strategy is coupled with a "continuous reactor system", developed some years ago by Luisi and co-workers [30], one can achieve a long-time bioreactor that can be almost continuously fed.

From the point of view of liposome physical-chemistry, there are in principle two main size regions that can be probed: the sub-micrometric size region (0.1-1 μm) and the giant-vesicle size region (>1μm). Working with smaller vesicles means utilizing simpler vesicle formation techniques, but with the drawback of involving only indirect analysis methods. Vice versa, working with giant vesicles can be advantageous from the detection viewpoint, since giant vesicles can be easily studied in real time light microscopy, but unfortunately the preparation methods are often difficult and not very reproducible. Intermediate sized vesicles (i.e. 0.4-0.8 μ m) offer the possibility of joint analysis by flow cytometry and confocal microscopy, as well as classical batch techniques, as spectrofluorimetry, in addition they are readily prepared. On the other hand, the multiple lamellae that often characterize such structures can limit the theoretical understanding of these systems.

4.2 Development of a PURESYSTEM-containing liposome system capable of shell self-reproduction.

Self-reproduction in the case of semi-synthetic cells is a difficult problem that should be addressed in various degrees of approximation. The first level that we propose is that of a system that expresses in its core those enzymes which catalyse the synthesis of the membrane. Preliminary studies have been carried out several years ago in the group of Luisi, aimed at producing inside lecithin liposomes the synthesis of lecithin [10]. The synthetic route that, also according to these early results and also to some new developments [31], appears to be promising is the so-called lipid salvage pathway, indicated in Figure 5.

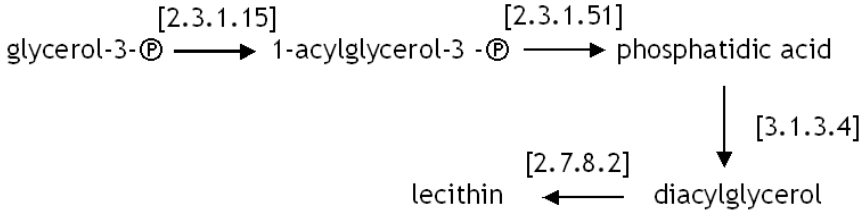


Figure 5. Lecithin biosynthesis by the salvage pathway. The four enzymes needed to accomplish the trasformation of glycerol-3-phosphate into phospatidylcholine are: *sn*-glycerol-3-phosphate *O*-acyl-transferase [2.3.1.15], 1-acyl-*sn*-glycerol-3-

phosphate *O*-acyl-transferase [2.3.1.51]; phosphatidate phosphatase [3.1.3.4]; diacylglycerol cholinephosphotransferase [2.7.8.2]. The reagents needed in each step are omitted.

The advantage of this pathway is also related to the chemical nature of precursors. In fact, in addition to glycerol-3-phosphate, the long-chain acyl-CoA, needed to carry out steps 1 and 2 of the route, as well as CDP-choline in step 4 are all water-soluble compound.[1] In other words, the salvage pathway transforms small water-soluble molecules into a membrane-forming lipid molecule, i.e., lecithin.
As additional possibility, we must consider that in principle it is possible to stop the salvage pathway after the second acylation step, obtaining phosphatidic acid, a molecule that forms *per se* lipid bilayers and vesicles. We recently investigated the formation and the properties of dioleoylphosphatidic acid vesicles, and its interaction with oleoyl-CoA [32].

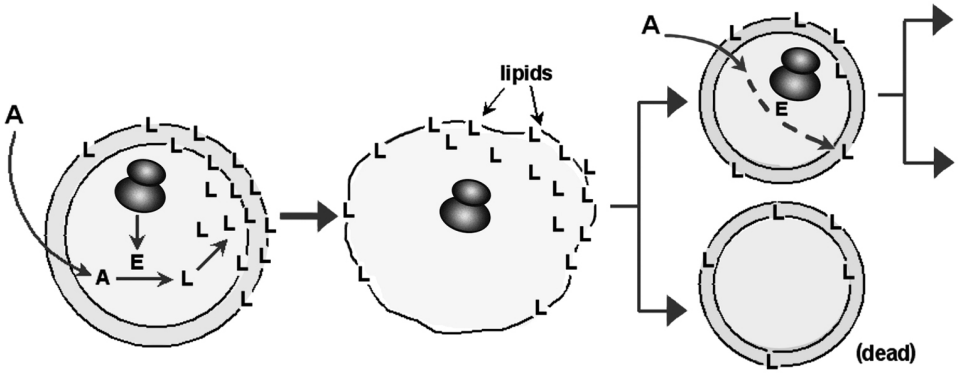


Figure 6. A cell that makes its own boundary. The complete set of biomacromolecules needed to perform protein synthesis (genes, RNA polymerases and ribosomes) is indicated as a couple of spheroids. The product of this synthesis (indicated as E) is the complete set of enzymes for lipid (L) synthesis, that start from the set of precursors A. After growth and division, some of the 'new' vesicles might undergo 'death by dilution'.

In contrary to the approach of Schmidli et al. [10], who started by a proteoliposome containing the four enzymes of the salvage pathway, in the modern approach, the endogenous biosynthesis of lipid should be obtained after the *in situ* expression of the enzymes indicated in Figure 5.
Therefore, the practical approach to such study involves the entrapment of the PURESYSTEM kit together with the four genes codifying for the four enzymes of lipid biosynthesis, e.g., the expression of lipid-synthesis enzyme-battery (Figure 6).
In alternative to a self reproducing lecithin shell from within vesicle compartment, based on the salvage pathway, we can envisage the introduction into vesicle of a Fatty Acid Synthetase (FAS) enzyme for fatty acid synthesis. From the point of view of compartment chemistry, phospholipids provide a stable and quite inert membrane which is very convenient to handle. However, if we focus our attention to the use of minimal cells as model prebiotic structures, a more simple surfactant must be used and most scientist involved in the origin of life research agree that fatty acid vesicles are better candidate. The use of fatty acid vesicles, however, may

[1] Long-chain fatty acid-CoA molecules are probably soluble in water in form of micelles.

represent a challenge because of potential chemical incompatibilities, magnesium ions above all, but also working pH and high-pI enzymes. To accomplish this result, we are focusing on the self-reproduction of vesicle shell by the endogenous (within vesicle) synthesis of fatty acids. Two strategies can be in principle pursued: (*i*) incorporating first the enzyme/s that synthesize the Fatty acids or, (*ii*) introducing the corresponding gene/s, and expressing those enzyme/s within the vesicles. Work is in progress to clone the gene for a mammalian Fatty Acid Synthetase Type-I enzyme into a protein expression vector. Type I FAS is a multifunctional enzyme consisting of a dimmer with single long multifunctional polypeptides and is water soluble. The human FAS has been successfully expressed in *E.coli* and proven to catalyzes the synthesis of palmitate [39] (Figure 7).

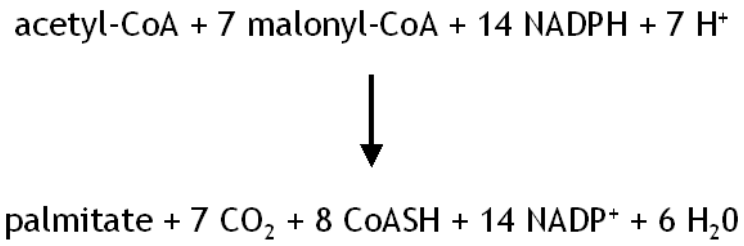


Figure 7. Type-I FAS enzymatic system is capable to catalyze all the biochemical reactions that, from acetyl-CoA and malonyl-CoA, lead to long-chain fatty acids, principally palmitate.

Palmitate represents the principal product of the reaction and, in the case of the animal FAS, it is spontaneously released by the enzyme when ready, as it is without the need of any further modification before to be incorporated in the vesicle membrane.
If the production of lipids is significant (in terms of chemical yield), it follows that the membrane surface can grow, and eventually the liposomes will divide up so as to maintain a constant surface/volume ratio. Of course, with progressive generations, most of the liposomes will become inactive, due to "death by dilution", as the new created liposomes will not necessarily contain, by statistical laws, all components which are necessary for protein synthesis.

The prediction of this behaviour originates from the results on the vesicle self-reproduction, carried out mainly by Luisi and co-workers in Zurich [33-37], and extended recently by other groups [9,38]. In that studies, a membranogenic surfactant is added to pre-formed vesicles, with the result that original vesicles take up the freshly surfactant molecules, grow and divide in vesicle-daughters. It has been also postulated that such mechanism could have had a role in prebiotic vesicle proliferation [36].

The system depicted in Figure 6 (a system that makes its own boundary from within) corresponds to an autopoietic system, and it would be the first time that one such system is realized in the laboratory. The system should be so optimised, to contain in the original cell many numeral copies of the ribosomes and of the synthesizing enzymes, so as to increase the probability that the first and second generation of newly produced liposomes also contain all the battery for shell-self-reproduction.

4.3 Self-reproduction of semi-synthetic cells.

A real self-reproduction of biological cells is based on the formation of cellular copies, each containing the full genetic apparatus. The experimental implementation of this principle in a semi-synthetic construct is a very ambitious goal that requires a careful and detailed evaluation of the system under study. In particular, it is necessary to find conditions under which the ribosomal system and all polymerase enzymes self-replicate. This is in principle possible by enriching the PURESYSTEM with additional components that carry out the requested functions.

In particular, in addition to the replication of genetic material, the critical point is related to the production – from the within – of all the components of the PURESYSTEM itself, ribosomes included. Of course, and in contrast with the two above mentioned cases (expression of GFP and expression of the enzymes of a simple biosynthetic pathway), for the complete core-and-shell self-reproduction, the minimal genome itself should be inserted into vesicles, in order to allow the transcription of all the genes that will originate the whole cellular machinery (RNA polymerases, DNA polymerases, enzymes for the protein expression, ribosomal proteins, lipid biosynthesis enzymes, etc…).

It is self-evident that the achievement of this complexity level requires a very high control over the compartmentalized reactions and over the interconnected pathways of protein biosynthesis, DNA replication, lipid production and whole enzymatic-machinery replication. We are confident that in the following years the possibility of achieving such challenging goals will significantly increase.

5 Acknowledgements

We thank the Study Centre “Enrico Fermi” (Rome) for financial support.

6 References

[1]Luisi PL, Ferri F, Stano P (2006) Approaches to semi-synthetic minimal cells: a review. Naturwissenschaften 93, 1-13

[2]Gil R, Silva FJ, Peretó J, Moya A (2004) Determination of the core of a minimal bacteria gene set. Microbiol Mol Biol Rev 68:518-537

[3]Fischer A, Franco A, Oberholzer T (2002) Giant vesicles as microreactors for enzymatic mRNA synthesis. ChemBioChem 3:409-417

[4]Yu W, Sato K, Wakabayashi M, Nakatshi T, Ko-Mitamura EP, Shima Y, Urabe I, Yomo T (2001) Synthesis of functional protein in liposome. J Bioscience Bioengineering 92:590-593

[5]Ishikawa K, Sato K, Shima Y, Urabe I, Yomo T (2004) Expression of a cascading genetic network within liposomes. FEBS Letters 576:387-390

[6]Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T (2001) Cell-free translation reconstituted with purified components. Nature Biotechnology 19: 751-755

[7]Kuruma Y, Nishiyama K, Shimizu Y, Mueller M, Ueda T (2005) Development of a minimal cell-free translation system for the synthesis of presecretory and integral membrane proteins. Biotechnology Progress 21, 1243-1251

[8]Shimizu Y, Kanamori T, Ueda T (2005) Protein synthesis by pure translation systems. Methods 36:299-304

[9]Takakura K, Sugawara T (2004) Membrane dynamics of a myelin-like giant multilamellar vesicle applicable to a self-reproducing system. Langmuir 20:3832-3834

[10]Schmidli PK, Schurtenberger P, Luisi PL (1991) Liposomemediated enzymatic synthesis of phosphatidylcholine as an approach to self-replicating liposomes. J American Chemical Society 113:8127-8130

[11]Walde P, Goto A, Monnard PA, Wessicken M, Luisi PL (1994) Oparin's reactions revisited: enzymatic synthesis of poly (adenylic acid) in micelles and self-reproducing vesicles. J American Chemical Society 116:7541-7544

[12]Oberholzer T, Albrizio M, Luisi PL (1995) Polymerase chain reaction in liposomes. Chemical Biology 2:677-682

[13]Oberholzer T, Wick R, Luisi PL, Biebricher CK (1995) Enzymatic RNA replication in self-reproducing vesicles: an approach to a minimal cell. Biochemical Biophysical Research Communications 207:250-257

[14]Oberholzer T, Nierhaus KH, Luisi PL (1999) Protein expression in liposomes. Biochemical Biophysical Research Communications 261:238-241

[15]Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. Nature 409:387-390

[16]Luisi PL (2002) Toward the engineering of minimal living cells. Anatomical Record 268:208-214

[17]Oberholzer T, Luisi PL (2002) The use of liposomes for constructing cell models. J Biological Physics 28:733-744

[18]Islas S, Becerra A, Luisi PL, Lazcano A (2004) Comparative genomics and the gene complement of a minimal cell. Origins Life Evolution Biosphere 34:243-256

[19]Stano P, Murtas G, Luisi PL (2006) Semi-Synthetic Minimal Cells: New Advancements and Perspectives, in "Protocells: Bridging Nonliving and Living Matter", Rasmussen S and Bedau M (ed) The MIT Press.

[20]Pohorille A, Deamer D (2002) Artificial cells: prospects for biotechnology. Trends Biotechnology 20:123-128

[21]Deamer D (2005) A giant step towards artificial life? Trends in Biotechnology 23, 336-338

[22]Rasmussen S, Chen L, Deamer D, Krakauer DC, Packard NH, Stadler PF, Bedau MA (2004) Transitions from nonliving to living matter. Science 303:963-965

[23]Noireaux V, Libchaber A (2004) A vesicle bioreactor as a step toward an artificial cell assembly. Proceedings National Academy Science USA 101:17669-17674

[24]Noireaux V, Bar-Ziv R, Godefroy J, Salman H, Libchaber A (2005) Toward an artificial cell based on gene expression in vesicles. Physical Biololgy 2:P1-P8.

[25]Szathmary E, Santos M, Fernando C (2005) Evolutionary Potential and Requirements for Minimal Protocells, in "Prebiotic Chemistry", Walde P (ed). Topics Current Chemistry 259:167-211

[26]Tsumoto K, Nomura SM, Nakatani Y, Yoshikawa K (2001) Giant liposome as a biochemical reactor: transcription of DNA and transportation by laser tweezers. Langmuir 17:7225-7228

[27]Nomura SM, Tsumoto K, Hamada T, Akiyoshi K, Nakatani Y, Yoshikawa K (2003) Gene expression within cell-sized lipid vesicles. ChemBioChem 4:1172-1175

[28]Chakrabarti AC, Breaker RR, Joyce GF, Deamer DW (1994) Production of RNA by a polymerase protein encapsulated within phospholipid vesicles. J Mol Evol 39:555-559

[29]Walde P, Goto A, Monnard PA, Wessicken M, Luisi PL (1994) Oparin's reactions revisited: enzymatic synthesis of poly(adenylic acid) in micelles and self-reproducing vesicles. J Am Chem Soc 116:7541-7544

[30]Zepik HH, Bloechliger E, Luisi PL (2001) A chemical model of homeostasis. Angewandte Chemie International Edition 40:199-202

[31]Luci P (2003) Gene cloning expression and purification of membrane proteins. ETH-Z Dissertation No. 15108, Swiss Federal Institute of Technology (ETH) Zurich

[32]Dubois J, Stano P, Luisi PL (2005) Unpublished data

[33]Berclaz N, Mueller M, Walde P, Luisi PL (2001) Growth and transformation of vesicles studied by ferritin labeling and cryotransmission electron microscopy. J Phys Chem B 105:1056-1064

[34]Bloechliger E, Blocher M, Walde P, Luisi PL (1998) Matrix effect in the size distribution of fatty acid vesicles. J Phys Chem 102:10383-10390

[35]Lonchin S, Luisi PL, Walde P, Robinson BH (1999) A matrix effect in mixed phospholipid/fatty acid vesicle formation. J Phys Chem B 103:10910-10916

[36]Luisi PL, Stano P, Rasi S, Mavelli F (2004) A possible route to prebiotic vesicle reproduction. Artif Life 10:297-308

[37]Rasi S, Mavelli F, Luisi PL (2003) Cooperative micelle binding and matrix effect in oleate vesicle formation. J Phys Chem B 107:14068-14076

[38]Hanczyc MM, Fujikawa SM, Szostak JW (2003) Experimental models of primitive cellular compartments: encapsulation, growth, and division. Science 302:618-622

[39]Jayakumar A, Huang WY, Raetz B, ChiralaSS, Wakil SJ (1996) Cloning and expression of the multifunctional human fatty acid synthase and its subdomains in Escherichia coli. Proceedings National Academy Science USA 93:14509-14514

Creative Patterning of Mathematical Bacteria

Eshel Ben Jacob[1]

[1]School of Physics and Astronomy, Tel Aviv University, 69978 Tel Aviv, Israël

Abstract

Under natural growth conditions, bacteria can utilize intricate communication capabilities(e.g. quorum-sensing, chemotactic signaling and plasmid exchange) to cooperatively form (self- organize) complex colonies with elevated adaptability the colonial pattern is collectively engineered according to the encountered environmental conditions. Bacteria do not genetically store all the information required for creating all possible patterns. Instead, additional information is cooperatively generated as required for the colonial self organization to proceed.

We describe how complex colonial forms (patterns), emerge through the communication-based singular interplay between individual bacteria and the colony. Each bacterium is, by itself, a biotic autonomous system with its own internal cellular informatics capabilities (storage, processing and assessment of information). These afford the cell plasticity to select its response to biochemical messages it receives, including self-alteration and the broadcasting of messages to initiate alterations in other bacteria.

Hence, new features can collectively emerge during self-organization from the intracellular level to the whole colony. The cells thus assume newly co-generated traits and abilities that are not explicitly stored in the genetic information of the individuals.

Prologue: Our best friends and worst enemies

Eons before we came into existence, bacteria inhabited the then hostile planet Earth. Being the first form of life here, they had to devise ways to counter the spontaneous course of increasing entropy and convert high-entropy, inorganic substances into low-entropy, organic molecules. Acting jointly, these tiny organisms also paved the way for other forms of life by changing its harsh conditions into the life-sustaining environment we know. With their impressive engineer- ing skills, bacteria changed the atmosphere above us to be oxygen rich, and the water and soil to be loaded with nutrients, resulting in the Biosphere that supports all life on Earth [1-5].

Four billion years have passed, and the existence of higher organisms still depends on the unique bacterial know-how that converts between inanimate and living matter. With all our scientific knowledge and technological advances, the ways that bacteria act as Maxwell demons against the second law of thermodynamics is still a mystery. This makes bacteria are our best friends on Earth, indispensable friends we simply cannot do without. If we seek a future for the human race in space, we must take bacteria along for the ride, as none other can prepare the setting for us. They will quickly learn how to thrive in any new environment, and make use of whatever it offers to synthesize life-sustaining organic molecules and to recycle waste products for further use.

But, as we know, the same best friends are also our worst enemies. In our rush to free the human race from deadly bacterial diseases, we created a major health problem worldwide: bacteria are becoming increasingly resistant to antibiotics. Unaware of bacteria's cooperative behavior and social intelligence, which allow them to learn from experience to solve new problems and then share their newly acquired skills; we recklessly used, and still use, antibiotics to fight them. As

a result, we are now witnessing the resurgence of strains of disease-causing bacteria believed to have been vanquished long ago; only now they come armed with multiple drug resistance, and we can't invent new drugs fast enough.

The engineering skills of bacteria

The idea that bacteria act as unsophisticated, solitary creatures stems from years of laboratory experiments in which they are grown in Petri dishes in benign conditions. They can be tempted to reveal their tricks by, for example, growing them on nutrient-poor hard surfaces. The bacteria you see in fig. 1 coped with this situation by collectively producing a lubricating layer of fluid, which allowed them to swim on the hard surface. As they swim, the individual bacteria at the front push the layer forward so as to pave the way for the colony to expand. By carefully adjusting the lubricant viscosity, the bacteria stick together and keep the colony dense enough for protection [4,6,15-17].

Under conditions somewhat more favorable to motion, such as softer substrate, the bacteria engineer radically different classes of colony patterns. In this situation, the branches exhibit macroscopic chirality, always curling in the same direction (handedness). Accompanying the colonial structure is a designed genome change; the bacteria are now programmed to become much longer, which helps them to move in a coordinated motion within the branches [4,6,15-17].

To achieve even greater efficiency, bacteria invented the clever mechanism of chemotactic signaling, in which the individual bacteria send chemical messages to tell their peers in which directions to move. For example, when detecting a rich source of food they call their peers to join the meal by sending attractive chemotactic signals. On the other hand, bacteria that detect regions of low food or harmful chemical imbalances send out a repulsive chemical to signal the others to stay away [4,6,15-17].

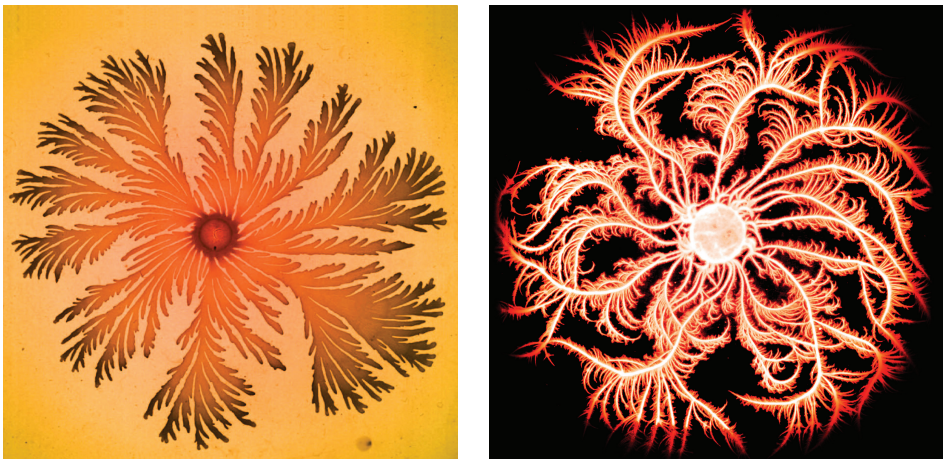
Using these self-engineering strategies, the individual cells collectively manipulate the overall colony organization for the group benefit, as is reflected by the tantalizing colonial patterns shown in Fig.1.

Clearly, bacteria cannot contain in their genes the information for creating all the patterns they might need to survive in unexpected situations. Well, they don't need to; they only need to have coded genetic information to provide them with the strategic design principles and the tools for communication, for information processing, and for changing themselves accordingly. Using these tools, they can design new creative shapes [4,6,15-17].

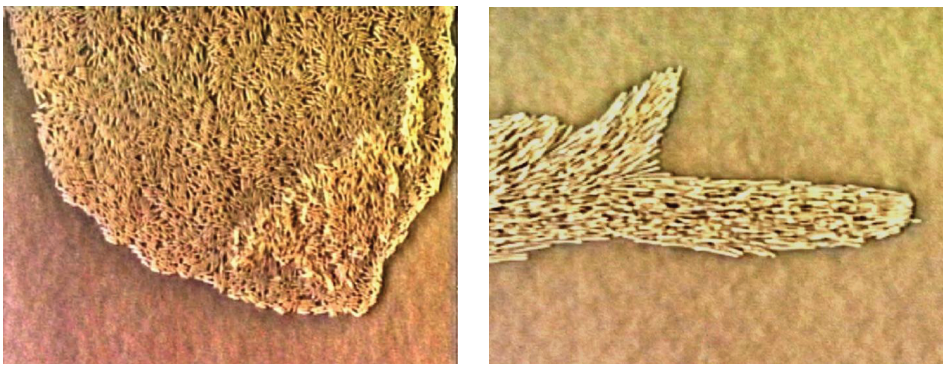
Bacterial communities

Bacterial engineering creativity is further manifested when forced to grow on very hard surfaces. The colony is now formed from new building blocks the vortices shown in Fig.2. It becomes much like multi-cellular organisms, with cell differentiation and distributed tasks.

In fact, bacteria can go a step higher and form a community (biofilm) of many cooperating colonies [8,9]. Each colony in the community acts as an organism that communicates with the other colonies for coordination and distribution of tasks, for the benefit of the community as a whole. To have an idea of the complexity involved, let us look at our oral cavity, which hosts a biofilm composed of hundreds colonies of different bacteria species, each consisting of tens of billions of bacteria.



1-a 1-b

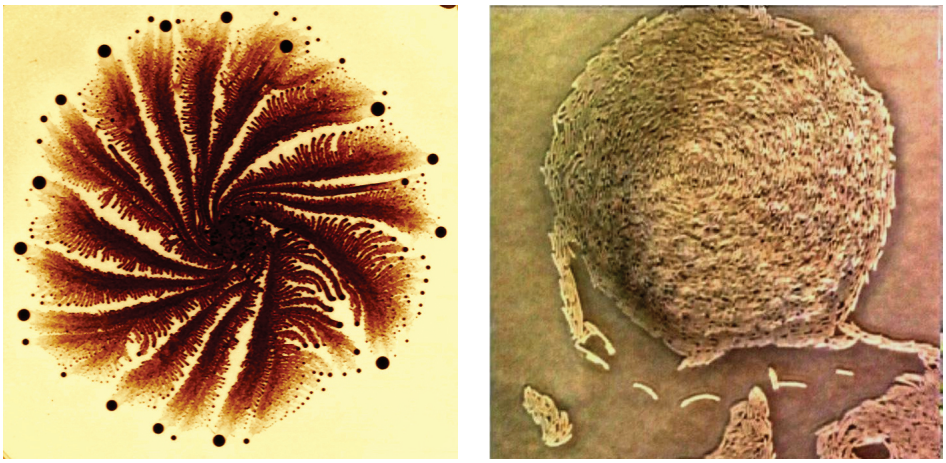


1-c 1-d

Figure 1: Patterns of Paenibacillus dendritiformis bacteria form when grown on nutrient-poor, hard substrate. Far from being shapes of mere aesthetic beauty, these colonial structures reflect the self- engineering skills of bacteria. The spreading patterns help the colony access more of the scarce food in the most efficient way under the given conditions. Ordinary branching pattern is shown on the left (a), and the chiral one (with broken left-right symmetry) is shown on the right (b). The top pictures show the colony patterns. Each colony is a few inches in size and has more bacteria than the number of people on Earth. The bottom pictures (c) and (d) show the individual bacteria (the small bars) at the branch tips with x500 magnification for (a) and (b) respectively.

Yet bacteria of all those colonies communicate for tropism in shared tasks, coordinated activities, and even exchange of relevant genetic bacterial information. For that to happen, cells should be able to talk and make sense of chemical messages they receive within a chattering of a huge crowd that is about thousand times larger than the number of people on Earth. In linguistic terms, the cells have multi-lingual skills, and each cell should be able to identify messages from its peers to the colony but at the same time also understand some of the messages from other colonies.

For that, bacteria have developed intricate chemical signaling mechanisms using a broad repertoire of biochemical messages - from simple molecules to ”cassettes of genetic materials” (plasmids). More recently, it was realized that to conduct social life, bacteria use the chemical messages much like a language, including the semantic (the assignment of meaning or interpretation of messages) and pragmatic (conduction of a dialogue) levels of linguistics [15].



2-a 2-b

Figure 2: Patterns of the Paenibacillus vortex are formed during growth on very hard surface. In these colonies (a), foraging vortices of rotating bacteria shoot out to conquer the hard agar, lubricating the way for their followers. The dynamics is fascinating: a vortex (b) grows and moves, producing a trail of bacteria and being pushed forward by the very same bacteria left behind. At some point, the process stalls and this is the signal for the generation of a new vortex behind the original one; the latter leaves home (the trail) as a new entity toward the colonization of new territory.

Harnessing the art of generic modeling

With present computational power, it is only natural to use computer models to study bacterial self-engineering in attempt to reveal their secret strategies we can learn from. In physics and chemistry, models have long served as an indispensable research tool. But in the life sciences the predictive power of models is still largely questioned. The usefulness and crucial role of modeling in the study of bacterial complex organization as is illustrated in the communicating walker model presented in Box B. We try to demonstrate that generic modeling allows in a natural way for both the physics (represented by chemical fields such as the nutrient concentration, any chemical signals, the lubricating fluid flow) and the biology (the response of individual cells to their sensed environment) to be represented in a computationally tractable format. Yet models should never be viewed as precise analogs of the actual system, but instead as tools for unraveling cause and effect and for helping us search for as yet unknown biological phenomena at the single bacterium scale.

One can easily fall into the ‘reminiscence trap’ - the tendency to devise a set of rules that will just mimic aspects of observed phenomena. Nevertheless, the other extreme the ‘realistic trap’ - where the model becomes swamped with too many details (and, usually, unknown parameters) and hence loses all predictive power must also be avoided. Model building, indeed, is ”an art in its own right” (the skills can only be acquired by practice), the challenge being how to elicit the generic features and basic principles needed to explain behavior from experimental observations and biological knowledge.

Harnessing the genome to shape the mathematics

One of the most important unanswered questions regarding these branching patterns involves the branching-to-chiral transitions themselves. The change is heritable (epigenetic) in the sense that inoculating a new agar plate with bacteria form a chiral region will create a colony which is immediately chiral. Hence this new structure is not just a passive response to an altered environment, but instead is connected to some switch that has been flipped in the bacterium’s internal circuitry.

The nature of this switch is not known and the signal responsible for the switching is not known. One interesting speculation is that the transition requires cell-cell communication, as is known to be the case in the switch that results in bacteria deciding to become dormant spores in response to poor growth conditions.

Merry go round

The patterns discussed so far depend on the ability of individual cells to push outward at the edges of the colony, of course with the help of the lubricating fluid. If the agar is made harder, another strategy can emerge. The bacteria can first organize themselves into coherent rotating vortices with highly coordinated motions of the individual cells. Some recent work on the conditions necessary for the emergence of these structures is discussed in Box D. These vortices can then move on the surface, under the influence of nutrient and chemotactic chemical signals and organize the colony in their wake. This type of pattern-formation is best thought of as hierarchical - instead of directly forming the colony scale structure, the cells first create mescoscale functional units which interact among themselves and with the cells left in the trails behind the vortices as they move outward. Building more highly-organized structures by creating a hierarchy appears to be a common motif in biological systems.

Shaped to survive antibiotic stress

Both model calculations and direct experiments have taught us a lesson that needs to be stressed throughout, namely that the patterns are highly reproducible; the interplay of these coordination mechanisms creates stable colony-scale responses to fixed environments even in the presence of obvious large-scale fluctuations in individual cell behavior. Adaptable self-engineering can be viewed as the solution to a challenging self-consistency mathematical problem at the forefront of optimization and control in non-linear dynamics. Thus it is fair to conclude that collectively, bacteria can glean information from the environment and from other organisms and interpret the information in an existential ”meaningful” way, i.e. by building an appropriate colony structure. It is perhaps not so far-fetched to imagine that the bacteria even develop common knowledge and learn from past experience. As we already mentioned, the recent findings about shaped to survive as response to antibiotic stress appear to be pushing us in this direction.

Linguistic communication and bacterial intelligence

Finally, we have recently argued that the way the bacteria coordinate their response to changing environments and to their own dynamical history is perhaps not very different from semantic communication as we usually understood it to occur between higher-level organisms (such as us). After all, an individual cell must make sense of a set of complex chemical messages it receives regarding the state of the colony and the features of the environment and act accordingly. This interpretation clearly depends not only on the message but also on the internal state of the receiving cell, hence the use of the term semantic. The cell must then transmit its decisions to other cells by sending out its own chemical messages. Should this be called intelligent behavior? Perhaps the bacteria deserve to be given more credit for inventing some of the basic approaches that all living organisms use to survive in the face of an unforgiving world.

References

[1] Lovelock, J. Gaia: A new look at Life on Earth. *Oxford University Press*, 1995

[2] Margulies, L. and Dolan, M.F. Early life: Evolution on the Precambrian Earth. *Jones and Bartlett*, 2002

[3] Liebes, S., Sahtouris, E., Swimme, and Liebes, S. A Walk Through Time: From Stardust to Us. *Wiley*, 1998

[4] Ben-Jacob, E. Bacterial self-organization: co-enhancement of complexification and adaptability in a dynamic environment. *Phil. Trans. R. Soc. Lond. A*, 361:1283-1312, 2003

[5] Ben Jacob, E., Shapira, Y. and Tauber, A.I. Seeking the foundations of cognition in bacteria: From Schrödinger's negative entropy to latent information. *Physica A*, 359:495-524, 2006

[6] Ben-Jacob, E. *et al.* Generic modeling of cooperative growth patterns in bacterial colonies. *Nature*, 368:46-49, 1994

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

[8] Wirth, R. *et al.* The Role of Pheromones in Bacterial Interactions. *Trends Microbiol.*, 4:96-103, 1996

[9] Rosenberg, E. (Ed.) Microbial ecology and infectious disease. *ASM Press Washington DC*, 1999

[10] Velicer, G.J. *et al* Developmental cheating in the social bacterium Myxococcus xanthus. *Nature*, 404:598-601, 2000

[11] Crespi, B.J. The evolution of social behavior in microorganisms. *TrendsEcol. Evol.*, 16:178-183, 2001

[12] Bassler, B.L. Small talk: cell-to-cell communication in bacteria. *Cell*, 109:421-424, 2002

[13] Kolenbrander, P.E. *et al* Communication among oral bacteria. *Microbiol. Mol. Biol. Rev.*, 66:486-505, 2002

[14] Velicer, G.J. Social strife in the microbial world. *Trends Microbiol.*, 7:330-337, 2003

[15] Ben Jacob, E. *et al* Bacterial Linguistic Communication and Social Intelligence. *Trends in Microbiology*, 12 (8):366-372, 2004

[16] Ben Jacob, E., Aharonov, Y. and Shapira, Y. Bacteria harnessing complexity. *Biofilms*, 2005

[17] Ben Jacob, E. and Levine, H. Self engineering capabilities of bacteria Interface. *Published online.* Rsif, 2006

[18] Ben-Jacob, E. Bacterial wisdom, Godel's theorem and creative genomic webs. *Physica A*, 248:57-76, 1998

[19] Ben Jacob, E. and Shapira, Y. Meaning-Based Natural Intelligence vs. Information-Based Artificial Intelligence.. *The Cradle of Creativity* Edited by H. Nen Nun, Saarei Tzedk Jerusalem, 2004

Additional relevant publications, pictures of bacterial colonies and video clips of bacterial movements can be found at my homepage **`http://star.tau.ac.il/~eshel/`**

Structure, Dynamics and Modularity of Signal Transduction Networks

R Heinrich[1]

[1]Department of Theoretical Biophysics, Institute of Biology, Humboldt University Berlin
D-10115 Berlin, Germany

1 Signalling times in protein kinase cascades

An enormous amount of information has accumulated about the components of various signalling pathways, their interplay, and their final output. However, the complex nature of these pathways renders it extremely difficult to understand how they are regulated and which parameters determine their dynamics. For example, how do the magnitudes of signal output, of signal propagation time and of signal duration depend on the kinetic properties of pathway components, such as kinases or phosphatases? Questions of that kind are of high biological relevance since differences in the time characteristics of signalling, leading for example to a sustained or a transient response, can have dramatically different consequences [1]. Here we show that mathematical modeling [2] may give interesting answers to this and other related problems of signalling pathway dynamics.

As a very simple example we consider a linear signalling cascade in which stimulation of a receptor leads to the consecutive activation of several downstream protein kinases (Fig. 1). The "signal output" of this pathway is the phosphorylation of the last kinase which, in turn, can elicit a cellular response (e.g., activation of a transcription factor). Signalling is terminated by phosphatases, which dephosphorylate the kinases, and by inactivation of the receptor, which can involve receptor dephosphorylation, internalization of the receptor-ligand complex, and/or degradation of the receptor or ligand. This general scheme is representative of many signalling pathways stimulated, for example, by growth factors such as EGF, PDGF, or NGF [3]. Typically, real signalling pathways are more complicated than this scheme, due to cross-talk between signalling pathways, binding of kinases to scaffolding proteins, multiple phosphorylation and to participation of G-proteins [4, 5]. We describe the dynamics of a pathway depicted in Fig.1 by the following set of differential equations

$$\frac{dX_i}{dt} = \tilde{\alpha}_i X_{i-1} \tilde{X}_i - \beta_i X_i \ , \qquad (1)$$

where X_i and $\tilde{X}_i$ denote the concentrations of phosphorylated forms and unphosphorylated forms of kinases, respectively. The $\tilde{\alpha}_i$'s are second order rate constants for phosphorylations whereas β_i's denote rate constants for dephosphorylation by phosphatases. Defining $C_i = \tilde{X}_i + X_i$ as the total concentration of kinase *i* and using $\alpha_i = \tilde{\alpha}_i C_i$ as a pseudo-first order rate constant, Eqn. (1) becomes

$$\frac{dX_i}{dt} = \alpha_i X_{i-1}\left(1 - \frac{X_i}{C_i}\right) - \beta_i X_i \ . \qquad (2)$$

For the first kinase (X_1), activation occurs via the stimulated receptor R, that is, X_0 should be replaced by $R(t)$. For simplicities sake we assume an exponential decay of the receptor activity, $R(t) = R\exp(-\lambda t)$ where $1/\lambda$ is the "characteristic life time" of the active receptor.

To treat key questions for the regulation of signalling cascades such as: (1) How fast does the signal arrive at its destination? and (2) How long does the signal last?, we introduced the following two quantities:

a) *Signalling propagation time* τ_i: It represents the average time to activate kinase *i* and can be defined as follows

$$\tau_i = \frac{T_i}{I_i}\ , \quad \text{where} \quad I_i = \int_0^\infty X_i(t)dt\ , \quad \text{and} \quad T_i = \int_0^\infty t\,X_i(t)dt\ ; \tag{3}$$

b) *Signal duration* ϑ_i. This characterizes the average time during which a kinase remains activated and is defined by

$$\vartheta_i = \sqrt{\frac{Q_i}{I_i} - \tau_i^2}\ , \text{ where } \ Q_i = \int_0^\infty t^2 X_i(t)dt\ . \tag{4}$$

(for detailed justification, see [2]). The quantity I_i is called the integrated response of X_i [6]. For weakly activated pathways in which all kinases are phosphorylated to a low degree $(X_i << C_i)$ Eqn. (2) simplifies to a linear differential equation system

$$\frac{dX_i}{dt} = \alpha_i X_{i-1} - \beta_i X_i\ , \tag{5}$$

and the key parameters can be calculated explicitly [2]. For the signal propagation time through the entire pathway, and for the signal duration, one derives

$$\tau = \frac{1}{\lambda} + \sum_{j=1}^{n}\frac{1}{\beta_j}\ , \qquad \vartheta = \sqrt{\frac{1}{\lambda^2} + \sum_{j=1}^{n}\frac{1}{\beta_j^2}}\ , \tag{6}$$

respectively, where *n* denotes the total number of kinases within the cascade. Interestingly, both τ and ϑ depend only on the receptor life time and on the rate constants of phosphatases. Accordingly, in a weakly activated pathway, the kinases do not regulate these two key quantities. It has been derived that for strongly activated pathways the kinases do have effects on τ and ϑ which are, however, small compared to the effects of phosphatases.

The leading role of phosphatases in controling the time characteristics of MAPK pathway has been demonstrated also experimentally [5]. Applying sodium orthovanadate as protein tyrsosine inhibitor to fibroblasts stimulated by EGF, inhibition resulted in a much broader peak. Kinase

inhibitions (inhibition of MEK) did mainly effect the amplitude of signalling (for the corresponding theoretical prediction see [2]).

In the case that kinases are not completely specific the differential equation systems (2) has to be replaced by

$$\frac{dX_i}{dt} = \sum_{k \neq i}^{n} \alpha_{ik} X_k \left(1 - \frac{X_i}{C_i}\right) - \beta_i X_i \qquad (7)$$

where the α_{ik}'s are rate constants describing phosphorylation of kinase *i* by kinase *k*.

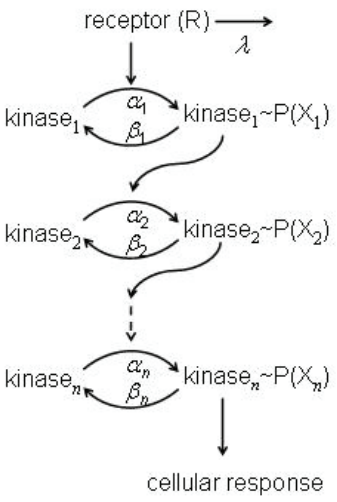


Fig.1. Most simple scheme of a protein kinase signalling cascade

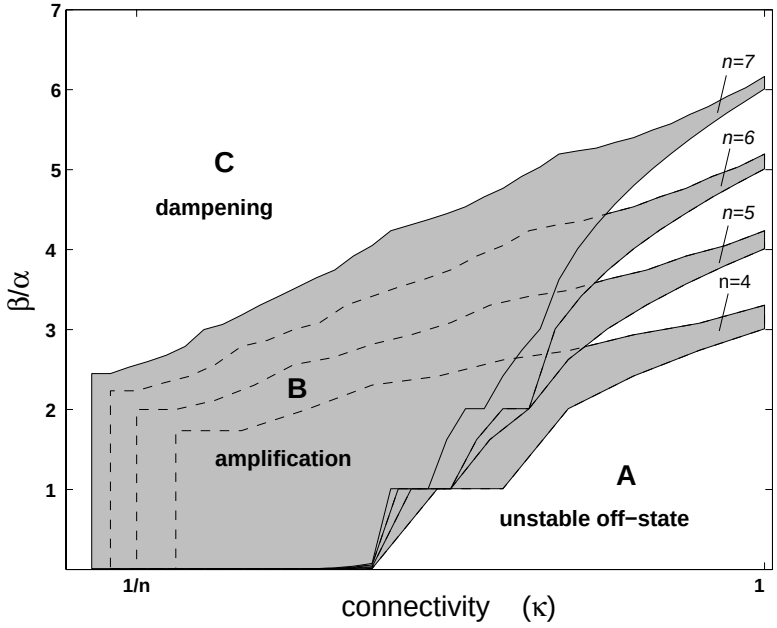


Fig.2. Effect of network connectivity and phosphatase activity on stability and signal amplification for networks of different size

It is easy to see that such an unspecificity may result in positive feedback loops and in this way to dynamic instability and autoactivation of a pathway. However, proper functioning of a signal transduction pathway will generally require that the signalling off-state is dynamically stable. Accordingly, it is a challenging task to analyse how pathway stability depends on the structure of the network as well as on the kinetic properties of the involved enzymes. Stability analysis can directly be performed by considering the spectrum of the eigenvalues of the Jacobian of the differential equation system (7). The results are represented in Fig. 2 showing for all possible structures for kinase/phosphatase networks different regions depending on the network connectivity (abscissa), on the phosphatase activity (ordinate: normalized β values) as well as on the network size *n*. The number of possible pathways increases strongly with increasing *n* (e.g. 9364 pathways for $n = 5$). It is seen that for each family of networks having the same size *n* there are three distinct regions A, B, and C leading to different dynamic properties of the networks. In region A all networks of a given family have an unstable signal off-state, in region B (shaded) stable networks exist showing amplification properties, and in region C all networks of the given family have a stable signal off-state and show a dampening in the signalling amplitudes.

The borders between regions A and B and between B and C are shifted towards higher β/α-values when n increases, that is, networks with higher number of kinases tend to be more unstable. Keeping in mind that biological cells contain several hundreds different types of kinases, one may draw the conclusion that real kinase networks must exhibit only a low connectivity to avoid autoactivation.

2 Structure and dynamics of the Wnt pathway

Signalling pathways may display a completely different structure and different dynamics when compared with MAPK pathways. This concerns, for example, the Wnt signal transduction pathway regulating cell fate in embryonal development. The same pathway plays a crucial role in the formation of cancer by controlling the concentration of β-catenin [8, 9]. The main components of the Wnt signalling pathway are the frizzled receptor (Frz), the scaffold proteins Axin and adenomatous polyposis coli (APC), the glycogen synthase 3 kinase (GSK3), the protein Dishevelled (Dsh), the phosphatase PP2A, as well as the transcription factors β-catenin and TCF. The pathway controls the concentration of β-catenin via assembly and disassembly of a destruction complex consisting of Axin, GSK3, APC, and PP2A. In the absence of the Wnt signal β-catenin binds to this complex, and becomes, after phosphorylation, a substrate for ubiquitination leading to its destruction by proteasomes. In this way the concentration of newly synthesized β-catenin is kept low. In the presence of the Wnt signal the Dsh protein is activated, resulting in an inhibition of GSK3 and in turn in a reduction of β-catenin phosphorylation and degradation. The build-up of β-catenin in the presence of a Wnt signal leads to transcription of specific genes. Recently, a first mathematical model of the Wnt-signal transduction pathway has been developed [10, 11].

The mathematical model of the Wnt pathway is based on the reaction scheme shown in Fig. 3. The core of this signalling network is the destruction complex to which unphosphorylated β-catenin binds (reaction 8). After phosphorylation (reaction 9) β-catenin is released from this complex (reaction 10) and degraded by proteasomes (reaction 11). GSK3 also phosphorylates the two scaffold proteins Axin and APC (reaction 4). The latter process is counteracted by PP2A (reaction 5). The destruction complex is formed by binding of Axin to APC (reaction 7) and subsequent binding of GSK (reaction 6). Dsh is activated upon stimulation of the frizzled receptor by Wnt (reaction 1). Reaction 2 denotes inactivation of Dishevelled. Inactivation of GSK3 is described by a release of GSK3 from the destruction complex (reaction 3). Reaction 14 and 15 denote the synthesis and degradation of Axin, respectively. Newly synthesized β-catenin (reaction 12) is not only degraded after binding to the destruction complex but also via non-Axin dependent proteolysis (reaction 13). In addition, the model takes into account complex formation of β-catenin with TCF (reaction 16).

The model consists of a set of differential equations governing the time dependent changes of the concentrations of proteins, either in their free form or as protein complexes, see [9, 10]. First, it has been used to define an unstimulated reference state in the absence of the Wnt signal. It is a stationary state where Dsh is inactive and does not affect the degradation complex. β-catenin concentration is kept low by continuous phosphorylation and degradation. Using the reference state as a starting point, other steady states can be calculated when the pathway is permanently stimulated.

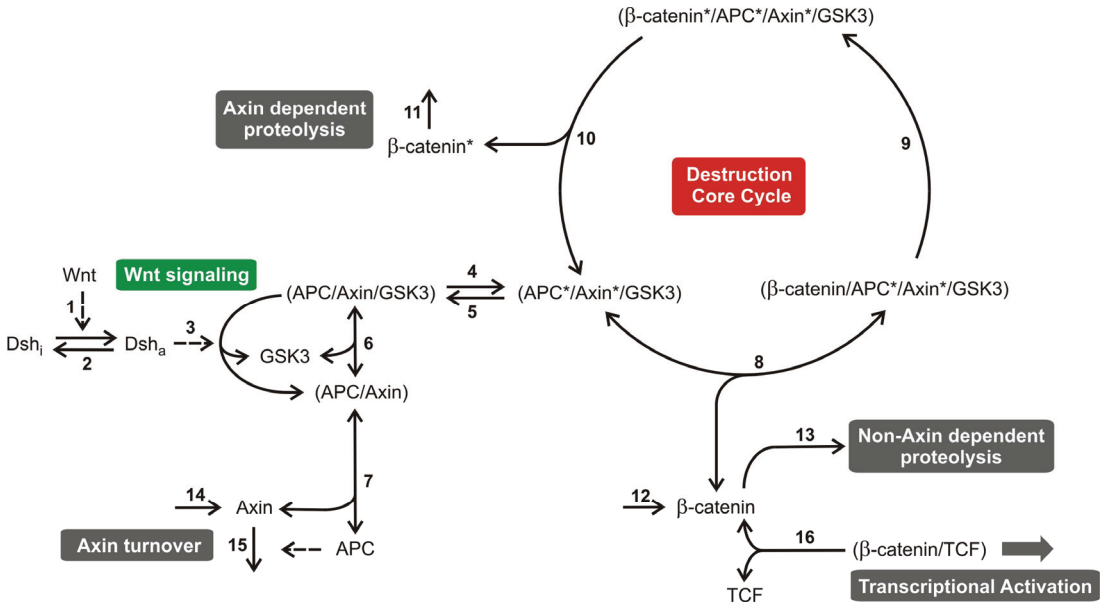


Fig.3 Reaction scheme of the Wnt-pathway model.

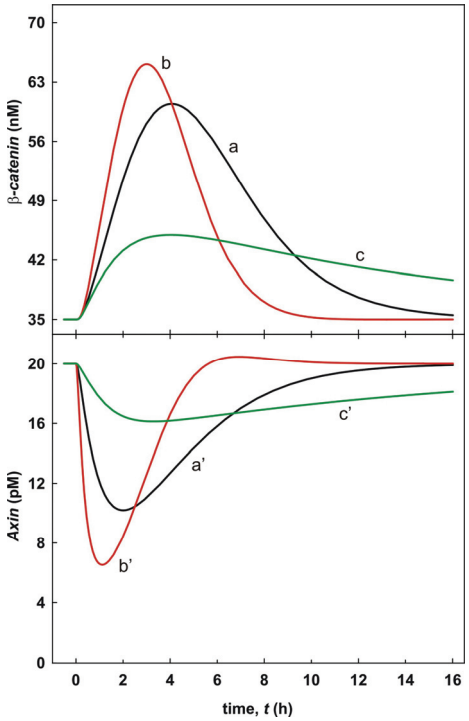


Fig.4 Time course of β-catenin and Axin concentrations resulting from a transient Wnt stimulation. The various curves differ in the turnover rate of Axin (see text).

In addition to steady states the systems equations allow to determine time dependent states. For example, to test, whether the mathematical model may represent correctly the dynamical properties of the Wnt pathway, it was used to simulate experimental data for the time courses for β-catenin degradation under a variety of *in vitro* conditions [9]. Of particular interest for *in vivo* conditions are transient stimulations of the receptor. Fig. 4 shows the time dependent behaviour

of the total concentration of β-catenin and the total concentration of Axin upon transient stimulation by Wnt. The concentration of β-catenin increases temporarily and then returns to its initial value. In contrast, the concentration of Axin is temporarily downregulated.
Curve a for β-catenin and curve a' for Axin were calculated by using the reference set of parameter values given in [9]. The other curves were obtained by changing turnover rates of Axin. It is seen that the rate of Axin turnover sharply affects the dynamics of the response of Wnt signalling. The curves b and b' and the curves c and c' are obtained for the case where both the synthesis rate and the degradation rate constant of Axin are increased by a factor 5 and decreased by a factor 5, respectivley. Interestingly, an increase in th turnover rate of Axin leads to higher amplitudes and shorter durations of the β-catenin signal. This effects can be explained by the fast degradation of Axin after its Dsh mediated release from the destruction complex [9,10].

Experimental analysis of this pathway revealed that the concentration of Axin is extremely low [9]. This very low concentration of a scaffold protein may indicate a very general and important design feature in the modularity of signalling pathways. Axin is a critical node point for the control of β-catenin levels through the regulation of β-catenin but it also interacts with components that are shared with many other essential systems. As the binding of these components such as GSK3 fluctuate due to Wnt signals, other components important in other pathways would also have to fluctuate. Yet, because, the concentration of Axin is so low there will be no appreciable changes in the overall levels of GSK3, Dsh, or APC. Hence, the very low Axin concentration isolates the Wnt-pathway from perturbing other systems, a simple mechanism to achieve modularity. One may conclude that quantitative and kinetic data may be important in detecting moduls. Inspection of circuit diagrams of signal transduction may, therefore, be not enough to identify modules in signalling networks.

References

[1] Marshall, C.J. (1995). Specificity of receptor tyrosine kinase signalling: Transient versus sustained extracellular signal-regulated kinase activation, Cell 80, 179-185.
[2] Heinrich, R., Neel, B.G., and Rapoport, T.A. (2002) Mathematical models of protein kinase signal transduction. Mol. Cell 9, 957-970.
[3] Cobb, M. H. (1999). MAP kinase pathways. Progr. Biophys. Mol. Biol. 71, 479-500.
[4] Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D., and Müller, G. (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat. Biotechnol. 20, 370-375
[5] Hornberg, J., Bruggeman, F., Binder, B., Geest, C., Marjolein Bij de Vaate, A.J., Lankelma, J., Heinrich, R., and Westerhoff, H. (2005) Principles behind the multifarious control of signal transduction ERK phosphorylation and kinase/phosphotase control. FEBS Journal 272, 244-258.
[6] Asthagiri, A.R., Reinhart, C.A., Horwitz, A.F., and Lauffenburger, D.A. (2000). The role of transient ERK2 signals in fibronectin- and insulin-mediated DNA synthesis. J. Cell Sci. 113, 4499-4510.
[7] Binder, B. and Heinrich R. (2004) Interrelations between dynamical properties and structural characteristics of signal transduction networks. Genome Informatics 15, 13-23.
[8] Wodarz, A. and Nusse, R. (1998) Mechanisms of Wnt -signaling in development. Annu. Rev. Cell Devl. Biol. 14, 59-88.
[9] Polakis P. (2000). Wnt signaling and cancer. Genes Dev. 14, 1837-1851.
[10] Lee, E., Salic, A., Krüger, R., Heinrich, R., and Kirschner, M. W. (2003) The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway. PLOS Biology 1, 116-132.
[11] Krüger, R. and Heinrich R. (2004) Model reduction and analysis of robustness for the Wnt/β-catenin signal transduction pathway. Genome Informatics 15, 138-148.

Genome-wide prediction and analysis of transcription factor repertoires

Sarah A. Teichmann[1]

[1] MRC Laboratory of Molecular Biology, Cambridge, UK.

Abstract

Regulation of gene expression influences almost all biological processes in an organism; sequence-specific DNA-binding transcription factors are critical to this control. These transcription factors are involved in complex circuits of regulation between transcription factors and target genes (Babu et al., 2004). However, for most genomes, the repertoire of transcription factors is only partially known. Hitherto transcription factor annotation has been largely based on genome annotation pipelines that use pairwise sequence comparisons, which detect only those factors similar to known genes, or on functional classification schemes that amalgamate many types of proteins into the category of 'transcription factor'. To fill this void, we have developed a novel transcription factor identification method, providing genome-wide transcription factor predictions for organisms from across the tree of life, available at www.transcriptionfactor.org (Kummerfeld and Teichmann, 2006).

We applied this procedure to several multi-cellular eukaryotes, including the human, mouse and insect genomes. In mouse, we used a comprehensive set of full-length cDNAs as well as microarray expression data to study regulation of the transcription factor repertoire of this organism through splicing and regulatory control. We find that transcription factor proteins preferentially undergo alternative splicing and that this affects the resulting proteins by altering the number and nature of DNA binding domains, a change that is likely to have a direct impact on the regulatory functionality. Through analysis of distinctive expression patterns from mouse microarray data, we identify sets of transcription factors that are ubiquitously expressed in mouse, and sets that are present in groups of related tissues. The 99 ubiquitously expressed mouse transcription factors include well-characterized proteins as well as hypothetical proteins. We have further investigated the role of about a dozen of these ubiquitous transcription factors in embryonic development using in situ hybridization to identify regional expression patterns.

The wealth of data and the experimental tractability of mouse makes it an excellent model organism, but the human genome remains of central interest. Starting from the repertoire of human transcription factors, we have traced the phylogenetic conservation of these proteins across the spectrum of eukaryote genomes. This allows us to identify the transcription factors that are specific to human and chimpanzee only, specific to mammals, vertebrates, metazoa, and finally, those that are common to all multi-cellular and uni- cellular eukaryotes. We will integrate the conservation patterns with microarray expression data and functional annotation of the human genes in order to investigate the origins of the phylogenetic distributions of transcription factors.

The invertebrate model organism Drosophila melanogaster has been used to the study of development and differentiation for decades. Transcription factor regulation plays an important role in both processes. Therefore, we aimed to accurately identify the repertoire of sequence-specific DNA-binding transcription factors (DBTFs) in twelve closely related fly species. Our approach was to first review the literature on D. melanogaster TFs to extract those that bind DNA in a sequence specific manner. Then we identified previously uncharacterized DBTFs using DNA-binding domain predictions, as described in the DBD database. This study was complemented by

a phylogenetic analysis of the evolutionary conservation of individual DBTFs across invertebrates, and of transcription factor families. In order to characterize the D. melanogaster DBTF repertoire in terms of spatio-temporal expression during development, we computationally integrated gene expression information from several sources using an anatomical ontology. We now aim to link the transcription factors expressed during development of particular tissues to their cognate DNA-binding sites, and thus build up networks of transcriptional regulatory interactions in tissue development. We are pursuing this by computationally predicting regulatory motifs from our integrated expression data set, as well as testing these by DNA-affinity chromatography followed by mass spectroscopic identification of the bound proteins.
Through these surveys of the phylogenetic conservation and dynamic expression of transcription factor repertoires we hope to gain insight into the role of gene expression in the evolution of organismal complexity.

References

Kummerfeld and Teichmann (2006)
DBD: a transcription factor prediction database. Nucleic Acids Res., 34, D74-81.

Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M. and Teichmann, S.A. (2004)
Structure and evolution of transcriptional regulatory networks. Curr. Op. Struc. Biol., 14, 283-291

EvolOMICS: Reconstructing the evolution of biological entities using comparative OMICS, from molecules to networks.

Erich Bornberg-Bauer[1]

[1] Univ. de Münster Botanical Inst. (Germany)

Abstract

While evolutionary research so far has mostly focused on the understanding of how single molecules evolved, the availability of large amounts of diverse OMICS data allows us to get a glimpse on the evolutionary history of biological entities.
We use a multitude of data from genomics, proteomics and trancriptomics to reconstruct the encestry and the evolution of networks and pathways. For example, we have investigated the domain wise evolution of several families of eukaryotic transcription factors (bHLH, bZIP, NR and MIKC-type MADS) and the associated emergence of theri interaction networks. This enables us to deiscern principles of processes whic are driven by random evolutionary events from those which are clearly adaptive in response to ecological changes.

Modelling Metabolic Pathways and Identification of the Control Steps

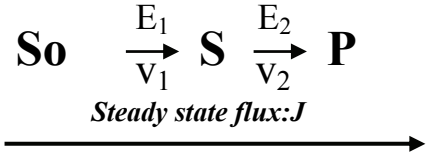
Marta Cascante

Department of Biochemistry and Molecular Biology, Faculty of Chemistry and CeRQT- Parc Científic de Barcelona. University of Barcelona, c/ Martí i Franquees, 1, 08028 Barcelona, Spain.
E-mail: martacascante@ub.edu
Webside: http://www.bq.ub.es/bioqint/arecerca.html

Metabolic Control Analysis (MCA) has been developed to describe quantitatively the distribution of control of fluxes and metabolite concentrations among the enzymes of the system under a defined steady-state (*As a review see* Fell, D. Understanding the control of Metabolism,Portland Press, London; 1997 and Cascante *et al*. Metabolic Control Analysis in Drug Discovery and Disease, Nat. Biotechnol. 20, 243-249; 2002). Recently the theory has been extended to include pathways with high enzyme concentrations and moiety conservation and metabolic channelling. One of the most important conclusions emerged from MCA has been to demonstrate that in some cases a democracy exist concerning the control of the flux by the enzymes instead of a single rate limiting step.

The most important magnitudes defined in MCA are the flux control coefficients. They have been defined to quantify how the flux change when the enzyme concentration (or enzyme activity) change. A high control coefficient indicates that the flux of the system is high sensitive to changes in the concentration of this enzyme.

To illustrate how control coefficients are defined we will use a simple two steps exemplary pathway:



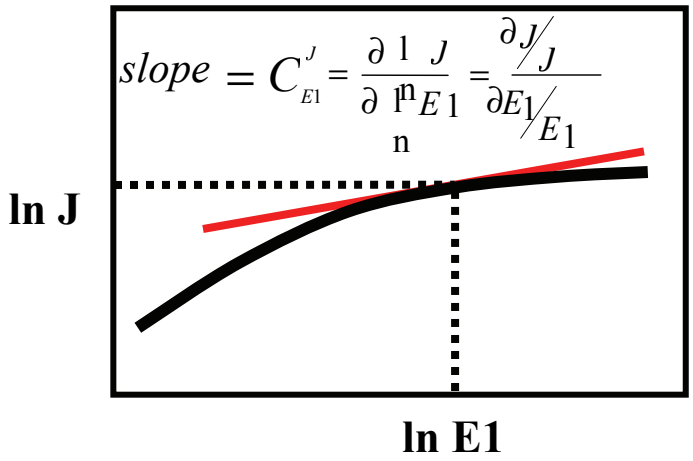
Where So and P are respectively the initial substrate and final product and S the intermediary metabolite. E1 and E2 are the enzymes that catalyze the two metabolic reactions respectively, v1 and v2 the corresponding local reaction rates and J is the steady-state pathway flux.

Flux Control Coefficients for enzyme 1 and 2 are defined as the fractional change in steady state flux produced by a fractional change in an enzyme concentration (or enzyme activity) and can be expressed as:

$$C_{E1}^{J} = \frac{\partial \ln J}{\partial \ln E1}$$

$$C_{E2}^{J} = \frac{\partial \ln J}{\partial \ln E2}$$

Graphically a flux control can be visualized as the slope of the tangent to the curve of ln J against ln E. In particular for enzyme 1:



Analogously control coefficients of intermediate metabolite concentrations with regard the enzymes have been defined to quantify the variation of metabolite levels with regard the enzymes.

These coefficients can be very useful in biotechnology and biomedicine to quantitatively predict changes in flux and concentration of metabolites after a genetic manipulation of the concentration of any enzyme of the pathway.

In general, for any systemic variables (Y), control coefficients can be defined with respect to any parameter (α) as the fractional change in the system variable (Y) over the fractional change in the parameter (α):

$$C^Y_\alpha = \frac{\partial Y / Y}{\partial \alpha / \alpha}$$

Several methods can be used to experimentally measure control coefficients. For instance, experiments can be performed in which active enzyme concentrations are changed, using either an inhibitor that specifically decrease the activity of a given enzyme or overexpressing an enzyme (i.e. transfecting cells with the appropriate adenovirus), and the corresponding flux at each different active enzyme concentration are measured. From this data control coefficients can be graphically estimated from the slope of the log-log plot of the fluxes versus the corresponding enzyme concentrations.

The values of flux control coefficients range between 0 and 1 for the enzymes in a linear chain metabolic pathway.

Moreover some simple relationships apply for flux and concentration control coefficients named summation theorems:

$$C^J_{E1} + C^J_{E2} = 1 \quad \text{(Flux summation theorem)}$$

The flux summation theorem, in the case of a linear chain where all the flux control coefficients are positive, permit us to conclude that if one enzyme have a control coefficient close to 1 all the other enzymes are practically not controlling the flux of the metabolic chain.

The concentration summation theorem, for this simple two step pathway, permit us to conclude that if one enzyme has a positive control coefficient on the intermediary substrate the other enzyme exert exactly the same control but in the opposite direction.

$$C_{E1}^{S} + C_{E2}^{S} = 0 \quad \text{(Concentration summation theorem)}$$

In rat hepatocytes, for instance, flux control coefficient of glucokinase on hepatic glycogen síntesis have been measured in rat hepatocytes using adenovirus-mediated enzyme overexpression and a valu close to 1 has been obtained (Agius et al., 1996) indicating a true rate-limiting step. A control coefficient with respect to one enzyme close to 0 indicates that the flux has a very low sensitive to changes in the concentration of this enzyme. An example, is the control coefficient of triose phosphate isomerase in erythrocytes which have been reported to be lower than 0.1 *(Schuster and Holzhutter, 1995, Eur. J. Biochem. 229: 403-418*). The fact that this control coefficient is so low explain that individuals deficient in more than 90% of the activity of this enzyme not present clinical symptoms whereas a 50% decrease of glucokinase activity (control coefficient close to 1) can result in clinical symptoms.

The complete estimation of control coefficients allows us to predict the response of the metabolic pathway to perturbations in any of its component enzymes.

The distribution of the control among the different enzymes in a metabolic pathway is a consequence of the structure of the pathway and the kinetic properties of each component enzyme. In MCA the influences of metabolite concentrations on the individual enzyme activities are measured in terms of the so-called elasticity coefficients, defined as the fractional change in rate of the isolated enzyme (vi) for a fractional change in a substrate (S), with all other effectors of the enzyme held constant at the value they have in the metabolic pathway:

$$\varepsilon_{S}^{vi} = \frac{\partial vi / vi}{\partial S / S} = \frac{\partial \ln vi}{\partial \ln S}$$

It should be noted that elasticities are positive for substrates and metabolites (activators) that stimulate the rate of a reaction and negative for products or inhibitors that decrease the reaction rate.

The relationships between elasticity coefficients (enzyme kinetic properties) and control coefficients (overall metabolic pathway properties) have been expressed in terms of the so-called connectivity theorems. In particular, for a simple two steps metabolic pathway where S is the intermediate metabolite, the connectivity theorem is written as:

$$C_{E1}^{J}\varepsilon_{s}^{v1} + C_{E2}^{J}\varepsilon_{s}^{v2} = 0$$

Rearranging this equation we can see that there is a tendency of large elasticities to be associated with small flux control coefficients, and vice versa:

$$\frac{C_{E1}^{J}}{C_{E2}^{J}} = \frac{\varepsilon_{s}^{v2}}{\varepsilon_{s}^{v1}}$$

Control coefficient distribution among the different steps of a pathway give us useful initial directions for manipulation of a metabolic network genetically or with drugs to achieve

the desired values of a flux or any other systemic variable. In spite that application of MCA to biotechnology is yet a young field some promising examples have already been described with biotechnological interest. One of the areas of application of metabolic control analysis in biomedicine is the rational design of combined drug therapy, that is the identification particularly suitable sites for the manipulation of the metabolism with drugs (*As a review see Cascante et al. 2002, Nature Biotechnology, 20, 243-249*). Thus, even every enzyme in a sequence is essential for the metabolic process to work, the effects on metabolism are likely to be obtained with lower concentration of drug if an enzyme with high flux control coefficient is being inhibited, rather than one with a low coefficient. Recently, we successfully applied this strategy to inhibit tumor proliferation through the inhibition of ribose phosphate synthesis, an essential pathway in cancer cell. Thus, specific inhibitors (such as dehydroepiandrosterone-sulfate and oxythiamine) of these key enzymes with high flux control coefficients in the pentose cycle can be expected to decrease nucleic acid synthesis and cell proliferation of tumour cells. We reported a maximum inhibitory effect of 80% *in vivo* in mice hosting Ehrlich's ascites tumour cells and a 60% maximum inhibitory effect *in vitro* in cultured Mia pancreatic adenocarcinoma cells when both inhibitors were administered together *(See Cascante et al.2000, Nutrition and Cancer, 36, 150-154*).

Multi-Agent Modelling of Nuclear Structures

G. Beslon[1]

[1] INSA - Lyon. France

Abstract

Although the eukaryotic nucleus is a major actor of the cellular process, its own dynamic and structure are still barely known. However, during the last five years, lot's of studies have shown that, in the nucleus, structure and dynamic are intimately coupled and that it is not possible to understand the former without a complete understanding of the later. Yet, to understand how a complex dynamic leads to such an organized system, we need specific modelling tools. The "Multi-Agent Systems" (MAS) approach will be discussed as an efficient tool to model nuclear dynamic. However, the use of MAS requires a precise methodology to choose the best modelling level (i.e. the description level) depending on the question one wants to address (i.e. depending on the observation level). We will present 3DSPI, a modelling software under development, based on molecular-agents. Our aim is to use 3DSPI to study the emergence of organized multi-protein structures.

Multi-Formalism Modelling and Model-based Interpretation of Biomedical Signals: Applications in Cardiology and Epileptology

Alfredo I. Hernández[1], Antoine Defontaine[1], Guy Carrault[1]

[1]LTSI - INSERM U642, Université de Rennes1, Campus de Beaulieu, F-35042 Rennes Cedex, France

Abstract

Many models of the cardiovascular system (e.g. cardiac electrical activity, mechanical activity, vascular system, autonomic nervous system,...) have been proposed during the last decades at different levels of detail (cell, tissue, organ,...) and using different formalisms (continuous, discrete, ...). Research is now focusing on a multiscale integration of these different models, in order to study more complex physiopathological phenomena and to derive new knowledge that can be useful in clinical practice.

This contribution is focused on the application of multiscale modelling methods to the interpretation of clinical observations and, in particular, of biomedical signals. In the first part, different modelling approaches of the cardiovascular system, from the cellular to the system level, are presented and their limitations for a direct clinical application are pointed out. The second part presents our current efforts towards the application of these physiological models in a clinical environment. They concern two major research subjects: i) the development of a new multiformalism modelling framework, which is a necessary step towards a multiresolution representation of physiological models, and ii) the proposition of a model-based interpretation method combining model-based reasoning and evolutionary computing. Finally, the application of these methods to different clinical problems, such as the interpretation of cardiac beats, the analysis of the autonomic regulation of heart rate and the characterization of electroencephalographic signals, will be presented.

1 Introduction

A model can be defined as a convenient representation of the knowledge about the structure and function of a given system [23]. This representation can be formalized in several ways: “mental” or “heuristic”, qualitative, graphic or mathematical models. Modelling and simulation have proven useful to analyze a set of observations obtained from a physical/physiological system; to understand the underlying mechanisms governing certain phenomena, or to predict future states of the system in order to prepare an appropriate action. Additionally, during the last decade, an important effort has been devoted to take advantage of the knowledge embedded into these models, to interpret the state of an observed physical/physiological system. This approach is known as model-based diagnosis (MBD) or model-based interpretation (MBI) [20]. MBI presents several advantages over previously proposed methods for the automatic interpretation of observed data, which have been classically developed as rule-based systems:

- Models are easier to maintain than a set of rules,
- The knowledge is represented in a compact manner,
- Models can generalize (i.e. interpretation of “new” cases),
- Time is represented intrinsically, avoiding a specific processing of temporal events,
- Models can provide detailed explanations,

- The same model can be used for other purposes, such as simulation, education, ...

However, new problems arise with the MBI approach: the development of an appropriate model can be a difficult task and the interpretation process in model-based systems is more complex than in their rule-based counterparts (mainly because more complex aspects are taken into account). This contribution is focused on the application of multiformalism and multiscale modelling methods to the interpretation of clinical observations and, in particular, of biomedical signals. Two key aspects of model-based interpretation will be presented:

- The design and development of an appropriate physiological model,
- The adaptation of the model parameters to reproduce and interpret clinical observations.

The discussion will be mainly focused on the model-based interpretation of cardiovascular signals, but the proposed approach will also be applied to the interpretation of electroencephalographic data.

2 Design and development of an appropriate physiological model for MBI

Physiological modelling of the cardiovascular system has been a subject of important research during last four decades. Since the pioneer works of Noble and Beeler and Reuter [18, 3], a number of electrophysiological models have been proposed for the main types of cardiac myocytes [1, 12]. These cellular models can be coupled in the form of 1D, 2D or 3D objects to represent a given part of the cardiac tissue, or to reproduce the whole cardiac anatomy[8, 12]. Other modelling approaches have been proposed to represent the mechanical activity of the heart, the vascular network or the autonomic modulation of the cardiovascular system. A number of projects are currently focused on the integration of these different models, in order to study more complex physiopathological phenomena and to derive new knowledge that can be useful in clinical practice. However, the definition and the simulation of such models become increasingly difficult, due in part to the complexity of the systems studied. Indeed, the underlying physiopathological processes imply:

- a wide diversity of spatial and temporal scales: from the gene to the organ levels and from ionic currents to the whole life;
- a high level of interdependence: for example, a given physiological function at the tissue level depends on the nature of cells constituting the tissue, and these cells are regulated by complex control systems such as the autonomic nervous system;
- a diversity of physical and chemical processes: i.e. regulation, growth, metabolism...
- a variety of energy domains: hydraulics, mechanics, electrical.
- an important amount of non-linear components.

Current research in “integrative modelling” seeks to cope with this complexity by means of a multiscale modelling approach [2]. This approach takes into account, in the same model, different physiological phenomena occurring on various scales, by using a common representation defined comprehensively at the most detailed level. Multiscale models of the cardiac function have been proposed in the literature by considering the interactions between the sub-cellular level, the electrical activity and the mechanical activity of the cells [4, 19]. These models of the global cardiac function have shown to be useful in a number of applications. However, their comprehensive definition makes them difficult to use in an MBI approach, because they require significant

data-processing resources. Moreover, none of the existing multiscale models allows a complete consideration of the whole cardiovascular system and, choice and compromise have to be done, depending on the expected application.

McCulloch and Huber [16] proposed a graph representing the integrative modelling on physiology, based on three different axes (Figure 1): i) spatial integration, ii) the integration of different sources and physiological systems (i.e. electrical activity, mechanical activity, regulation, ...) and iii) the integration of physical/physiological knowledge (one end corresponds to "superficial" –observational– models, which are limited to reproducing the observations, and the other end corresponds to the models integrating the most detailed physical knowledge).

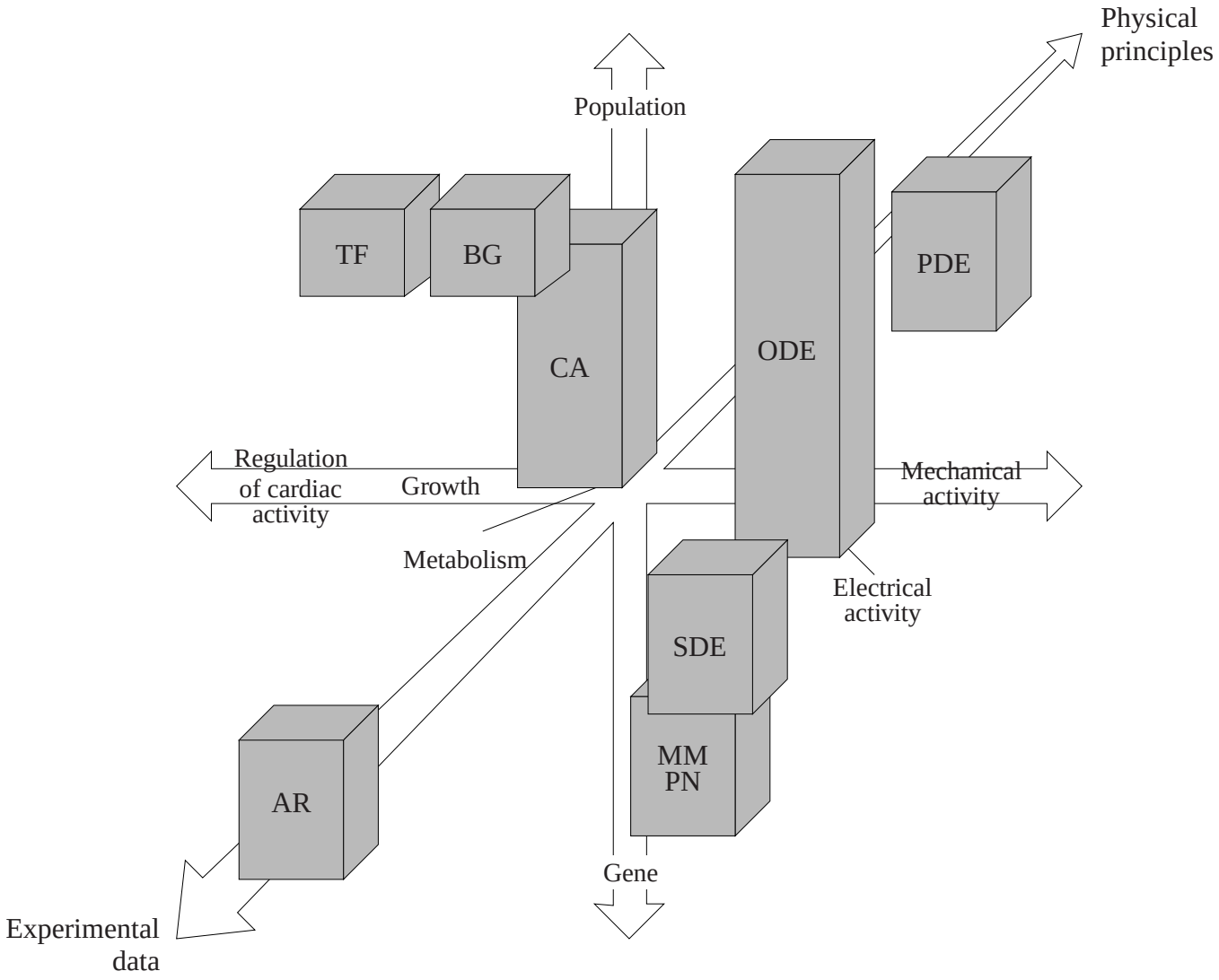


Figure 1: 3D space constituted by the three principal axes of integrative modelling proposed by McCulloch et al. The vertical axis corresponds to spatial integration, the diagonal axis represents the integration of knowledge and the horizontal axis represents the various physiological phenomena. We have projected on this space various formalisms used in the modelling of the electrical and mechanical cardiac activities and the regulation of the cardiovascular activity by the autonomic nervous system. The formalisms shown in the figure are: *AR* – Autoregresive models, *TF* – Transfer function, *BG* – Bond Graph, *CA* – Cellular Automata, *SDE* – Stochastic Differential equation, *MM* – Markov Models, *PN* – Petri nets, *ODE* – Ordinary Differential Equations and *PDE* – Partial differential equations.

We have completed this representation by projecting in this space, a number of different formalisms used in the literature for modelling the cardiovascular system. An analysis of this graph shows that there is a relationship between the formalisms used and the position of the models in this space. For example, the regulation of the cardiovascular activity by the ANS is considered on the "systems" level and is often modeled, based on experimental data, by a set of continuous transfer functions (TF).

As it will be shown in the presentation, the electrical activity of the heart can be modeled at different levels of detail, going from the cell to the whole organ, and is usually represented by means of continuous models (i.e. based on Ordinary Differential Equations) or by discrete models, (i.e. a set of coupled automata). However, both views still suffer from difficulties that reduce their clinical application: the former approach requires heavy computational resources while the latter is not able to reproduce certain pathologies defined at different scales.

To get round the practical limitations of existing multiscale models, multi-formalism modelling appears as a way to ease the integration of these different models together [21, 5]. In this context, one can easily think that a way to take advantage from the benefits of each approach into a model-based interpretation system would be to selectively define different regions of the modeled organ at different scale levels, depending on its physiological or pathological state. Such a multiresolution consideration is also legitimated by the practical clinical diagnosis performed by the physician, which aims at refining progressively the investigated region, going from a global consideration of healthy parts to a precise analysis of pathological sources. In this contribution, an original methodology allowing to combine different types of description formalisms will be presented [7, 5].

3 Adapting the model parameters to reproduce and interpret clinical observations

Once the model has been developed and validated, it can be used to interpret the state of the real system, by finding the set of the model parameters that reproduces at best the observed phenomena. This identification problem can be stated as follows: let X_O be a vector of observations from the studied system and $X_S = M(P)$ be the output of the developed model M, for parameter vector $P = [p_0, ..., p_{l-1}]$, Parameter identification can be seen as an optimization problem, consisting of finding the optimal parameter vector P^* that minimizes an error function

$$\epsilon(X_O, X_S) = \epsilon(X_O, M(P))$$

between the synthesized and the observed activities. The first problem to solve is thus to define an appropriate error function (or distance) between the observations and the output of the model. Different algorithms can then be used to identify the model parameters. The choice of a particular method depends strongly on the properties of the model, the error function and the observations. Particularly, some questions have to be addressed:

- Is the model linear or non-linear?
- How many parameters have to be adapted?
- Is the error function differentiable with respect to its parameters?
- Is there a unique bijective relation between the model parameters and a given model output (i.e. is this an ill-posed problem)?
- Is the set of observations (i.e. the system) stationary or non-stationary?

If we are considering a mathematical model for which the error function is differentiable with respect to each parameter (usually the case for observational models), information about the derivative can be exploited in order to find a solution. In the case of linear, static models, classical linear regression can be used [15]. For non-linear systems, the family of gradient-descent methods [14] can be used for parameter identification.

When the error function is not differentiable with respect to the model parameters, the search of solutions has to be based solely on the evaluation of the error function. In these cases, it becomes difficult to obtain information on the appropriate direction to follow, in the error space, in order to minimize the error (i.e. the information provided by the estimated gradient). Two types of optimization methods are adapted to this problem: exhaustive search and combinatory or stochastic search methods. The main drawback of exhaustive search is their computational cost. Stochastic search methods, like simulated annealing [13] and evolutionary algorithms (for which Genetic algorithms are a particular case) [11, 17], represent an interesting solution from the computational standpoint, providing, also, some interesting convergence properties towards the global-minima. These methods are robust and adapted to complex, non-differentiable models.

A model-based interpretation approach using evolutionary algorithms has been proposed in our laboratory and applied to the interpretation of different biological signals [9, 10, 22, 6]. Examples of the application of this method to the interpretation of cardiac beats, the analysis of the autonomic regulation of heart rate and the characterization of electroencephalographic signals will be presented.

References

[1] A. Bardou, P. Auger, P. Birkui, and J. Chassé. Modeling of cardiac electrophysiological mechanisms: from action potential genesis to its propagation in myocardium. *Crit Rev Biomed Eng*, 24(2-3):141–221, 1996.

[2] J. B. Bassingthwaighte, H. J. Chizeck, and L. E. A. ans Hong Qian. Multiscale modeling of cardiac cellular energetics. *Annals of the New York Academy of Sciences*, 1047:395–426, 2005.

[3] G. Beeler and H. Reuter. Reconstruction of the action potential of ventricular myocardial fibres. *J Physiol*, 268(1):177–210, Jun 1977.

[4] E. J. Crampin, N. P. Smith, and P. J. Hunter. Multi-scale modelling and the IUPS physiome project. *J Mol Histol*, 35(7):707–714, Sep 2004.

[5] A. Defontaine, A. Hernández, and G. Carrault. Multi-formalism modelling and simulation: Application to cardiac modelling. *Acta Biotheoretica*, 52(4):273–290, 2004.

[6] A. Defontaine, A. Hernández, and G. Carrault. Modèle de connaissances pour l'interprétation d'un électrogramme. *GRETSI 2005*, pages 377–380, 2005.

[7] A. Defontaine, A. Hernández, and G. Carrault. Multi-formalism modelling of cardiac tissue. *Lecture Notes in Computer Science - FIMH 2005*, 3504:394–403, 2005.

[8] C. S. Henriquez, J. V. Tranquillo, D. Weinstein, E. W. Hsu, and C. R. Johnson. *Cardiac electrophysiology: from cell to bedside – Edition 4*, chapter 30 – Three-dimensional propagation in mathematic models: integrative model of the mouse heart, pages 273–281. W B Saunders Company, 2004.

[9] A. Hernández, G. Carrault, F. Mora, and A. Bardou. Overview of carmem: A new dynamic quantitative cardiac model for ecg monitoring and its adaptation to observed signals. *Acta Biotheoretica*, 48(4):303–322–, 2000.

[10] A. I. Hernández, G. Carrault, F. Mora, and A. Bardou. Model-based interpretation of cardiac beats by evolutionary algorithms: signal and model interaction. *Artif Intell Med*, 26(3):211–35, Nov 2002.

[11] J. Holland. *Adaptation in Natural and Artificial Systems*. MIT Press, Boston, 1975.

[12] V. Iyer, R. Mazhari, and R. L. Winslow. A computational model of the human left-ventricular epicardial myocyte. *Biophys J*, 87(3):1507–1525, Sep 2004.

[13] S. Kirkpatrick, C. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680–, 1983.

[14] D. Luenberger. *Linear and non-linear programming*. Addison-Wesley Pub, Co. New York, 1984.

[15] V. Marmarelis. *The Biomedical Engineering Handbook*, chapter Methods and tools for identification of physiologic systems, pages 2432–2446. CRC Press, New Jersey., 1995.

[16] A. D. McCulloch and G. Huber. *'In silico' simulation of biological processes*, chapter Integrative biological modelling *in silico*, pages 4–19. Wiley, 2002.

[17] Z. Michalewicz. *Genetic Algotithms + Data Structures = Evolution Programs*. Springer-Verlag, New York, 1994.

[18] D. Noble. A modification of the Hodgkin–Huxley equations applicable to Purkinje fibre action and pace-maker potentials. *J Physiol*, 160:317–352, Feb 1962.

[19] N. P. Smith, D. P. Nickerson, E. J. Crampin, and P. J. Hunter. Multiscale computational modelling of the heart. *Acta Numerica*, pages 371–431, 2004.

[20] S. Uckun. Model-based reasoning in biomedicine. *Crit Rev Biomed Eng*, 19:261–292, 1992.

[21] H. Vangheluwe. *Multi-Formalism Modelling and Simulation*. PhD thesis, Universiteit Gent, 2001.

[22] F. Wendling, A. I. Hernández, J.-J. Bellanger, P. Chauvel, and F. Bartolomei. Interictal to ictal transition in human temporal lobe epilepsy: Insights from a computational model of intracerebral eeg. *Journal of Clinical Neurophysiology*, 22(5):343–356, October 2005.

[23] B. P. Zeigler, H. Praehofer, and T. G. Kim. *Theory of Modeling and Simulation Second Edition Integrating Discrete Event and Continuous Complex Dynamic Systems*. Academic Press, 2000.

Multi-scale vs function-dependent chromatin modelling

Annick Lesne[1], Christophe Lavelle[1] and Arndt Benecke[1]

[1] System Epigenomics Group, Institut des Hautes Études Scientifiques & Institut de Recherches Interdisciplinaires – CNRS 35 route de Chartres, F-91440, Bures sur Yvette, France
email: [lesne, lavelle, arndt]@ihes.fr

Abstract

Transcriptional regulation involves structural and dynamic events at many different time and space scales, the latter ranging from local DNA sequence to entire chromosomes. Multi-scale approaches inspired from theoretical physics have been developed in a bottom-up way to integrate parameters and mechanisms at a given scale into effective, and hopefully reduced, descriptions at higher scales. In chromosomes, however, the structures and dynamics at a given scale also exert a strong influence on smaller-scale ingredients, quantitatively controlling and even qualitatively modifying their properties. Self-consistent or iterative 'up-and-down' approaches are to be introduced to account for the strong interconnections between the levels and ensuing circular causal schemes, leading to an endless and prohibitive increase of complexity of the descriptions and models. Adopting another approach, we propose to devise models taking the biological function as a starting point and continued guideline; decomposition is achieved by dissecting its logic and implementation into basic processes. These elementary processes involve features at different scales and are already integrated in their formulation. More generally, such a decomposition results in 'self-scaled' functional modules, independent of the arbitrary description or observation scale. Self-scaling provides the relevant complexity reduction conjunct with the functional descriptive power in modelling complex living systems at multiple scales.

1 From DNA to chromosome: the central role of chromatin

It is now acknowledged that in eukaryotic cells, DNA is rarely present as a bare molecule but rather as a part of a nucleo-protein complex: the *chromatin* [34]. Its basic motif is made of the wrapping of 146 base pairs (bp) around an octamer of histone proteins, regularly spaced all along DNA. This motif is named a *nucleosome* and the DNA stretch connecting two adjacent nucleosomes, of length varying according to the species and cell type (roughly equal to a few tens of bp), is called a *linker*. This first level of genomic organization is observed in all eukaryote species; it is moreover highly conserved, hinting at a key role of nucleosomes in eukaryote mechanisms for implementing gene expression and cell differentiation. Other proteins are involved in chromatin, as well as ions. The assembly is further organized at different scales, up to the entire chromosome (see Figure 1). Among these higher levels of organization a central one, that might be observed in vivo and purified or reconstituted in vitro, is the *chromatin fiber*, also known as the *30 nm fiber* due to its roughly constant diameter of 30 nm. Our investigations on transcriptional regulation (e.g. [5], [19] [13], [6] [17]) and the discussion presented here about the relevant modelling approaches in this context, focus on this chromatin fiber. We argue that it is the central functional articulation between on the one hand, cell metabolism and signaling pathways and on the other hand, gene expression and nuclear regulatory networks [6] [19].

2 Multi-scale view of the chromatin fiber

2.1 A multi-scale organization

A major breakthrough has been provided by both experimental and theoretical investigations focusing on the multi-scale organization of the chromosomes and its mechanistic consequences, in particular for transcription and its control. As illustrated in Figure 1 or similar ones encountered in textbooks [34], structures at different scales superimpose in chromosomes: the DNA molecule, the nucleosomes and associated 10 nm beads-on-string fiber, the 30 nm chromatin fiber, chromatin loops and other (still debated [35]) higher-level structures, up to the entire chromosome.

The simplest multi-scale approach is to dissect the chromosome into *well-identified elementary structures* (DNA, nucleosomes, chromatin fiber) at *well-separated scales* (1 nm, 10 nm and 30 nm respectively) and to thoroughly investigate the properties of these elements in isolation [31]. But these elements at different scales are actually not isolated, and their properties reflect directly (though in an integrated way) in the higher-level features. In the spirit of a famous paper by P.W. Anderson entitled 'More is different' and describing the hierarchy of qualitatively different models (or even theories) that might be involved to investigate one and the same physical object [1], theoretical bottom-up approaches have been developed to derive effective coarse-grained models from more microscopic and more detailed ones. This point is illustrated below in § 2.2 and § 2.3 with studies we have conducted to determine mechanical and topological features of the chromatin fiber from the knowledge of DNA properties.

A first benefit of such a hierarchical bottom-up model is to provide a frame to interpret, exploit and integrate experimental or simulation data obtained at different scales. Let us quote as an illustrative sample some experimental facts and studies that will be discussed further in the present paper: epigenetic information about DNA methylation and histone-tail post-translational modifications [23], chromatin immuno-precipitation, nucleosome polymorphism [28], DNA [30] and chromatin fiber [12] [2] micro-manipulations, or numerical modelling [21].

But we shall emphasize in § 3 a biological specificity of this multi-scale organization, associated with its evolutionary history. In this regard, 'Life is different' and the analysis of Anderson cannot be straightforwardly transposed from physics to biology [24]. A main difference is the presence, in particular within chromosomes, of specific regulatory schemes, according which hyperstructures at the largest scales influences, controls the lowest scales.

2.2 Structural and kinematic effective models

As mentioned above, we here detail two successful implementations of a bottom-up multi-scale approach, namely the determination of the elastic properties and topological properties of the chromatin fiber, knowing its local architecture and assembly rules, as well as the elastic and topological properties of the underlying DNA stretch.

Elastic properties. Both on experimental (micro-manipulation of a single chromatin fiber [2]) and theoretical grounds [27] [19], the elastic properties of the chromatin fiber can be described within the formalism of *classical mechanics*. Namely, the fiber is seen simply as an *elastic rod* endowed with bend, twist and stretch elastic degrees of freedom. The relation between the stresses (force and torque) applied at the ends of the rod and the ensuing elastic strains in the rod is restricted to its dominant linear contribution, that defines four elastic coefficients [9]: the bend persistence length $\mathcal{A}$, the twist persistence length $\mathcal{C}$, the stretch modulus γ and the twist-stretch coupling factor g. Thermal fluctuations of each molecular degree of freedom (average energy $kT/2$) are accounted for in an effective and integrated way (the elastic coefficients will depend on the temperature T)

but the associated stochasticity is no longer described explicitly: at the fiber scale (30 nm), a deterministic model like those developed for macroscopic springs makes sense.

Also DNA can be described as an elastic rod and its elastic coefficients have been both measured experimentally by direct micro-manipulations of a single DNA molecule [30] and computed theoretically [11]. The stretch modulus value larger than 1000 pN shows that the stretch degree of freedom is in fact quenched in physiological conditions (the typical thermal energy kT corresponds to the work performed by a force of roughly 4 pN over a distance of 1 nm), hence it is enough to consider the bend and twist persistence lengths, respectively $A \approx 50$ nm and $C \approx 75$ nm. Note that such a description is in fact averaged over several base pairs and does not explicitly take into account sequence effects, which amounts to consider an homogeneous DNA molecule with no sequence variability.

Similarly, we assume an homogeneous local architecture for the fiber, characterized within a two-angle model [36] by the linker length l and the entry-exit angle ϕ (angle between the two linkers connecting a nucleosome to its neighbors along DNA). It is then possible to determine $\mathcal{A}$, $\mathcal{C}$, γ and g as a function of l, ϕ, A and C. The resulting expression evidences a *strong sensitivity of the fiber elasticity* with respect to these four constitutive parameters, that can be modified both in vivo and in vitro by changes in the ionic concentration, involvement of H1 or non histone proteins, nucleosome repositioning or remodelling, histone-tail post-translational modifications [8] [9]. The chromatin fiber thus appears as a *tunable spring* and its elastic properties cannot be ignored in its in vivo behavior and regulatory functions. Improvements of this minimal model can be considered, accounting for instance for the stacking interactions between the nucleosomes, without modifying the conclusion about the sensitivity of the fiber elastic properties.

Topological properties. The behavior of the fiber as an elastic rod is accompanied with topological properties. These properties are mainly encapsulated in the existence of a topological invariant when considering a fiber stretch with fixed ends (anchored on some substrate, e.g. a MAR region or boundaries, that has the same effect as connecting the ends and forming a closed loop) [20]. This topological invariant L_k^{fiber}, named the *linking number*, corresponds to the number of turns imposed on the fiber (starting from a relaxed state with free ends) before anchoring the ends or closing the loop by gluing these ends one to the other. A remarkable mathematical result relates this quantity to the *twist* T_w^{fiber} of the fiber (number of turns onto itself, around its own axis) and its *writhe* W_r^{fiber} (a quantity related to its path in the three-dimensional space): $L_k^{fiber} = T_w^{fiber} + W_r^{fiber}$. The balance between T_w^{fiber} and W_r^{fiber} varies with the three-dimensional conformation of the fiber, passing from $W_r^{fiber} = 0$ and $L_k^{fiber} = T_w^{fiber}$ in a straight fiber to a torsionally relaxed plectonemic conformation with $T_w^{fiber} = 0$ and $L_k^{fiber} = W_r^{fiber}$.

Such an inter-conversion presumably occurs during condensation and decondensation: it seems indeed possible to trigger decondensation at $L_k^{fiber} = \text{const.}$ by modifying the twist of the fiber through a change in the nucleosome shape [21]. Such a theoretical scenario is made quantitative by the analysis of the topological changes experienced by the underlying DNA molecule. It is indeed to be emphasized that topological properties and invariants can be defined both at the DNA and fiber levels, within the parallel elastic-rod modellings of DNA and the fiber encountered in describing their elastic properties. In particular, the same object is associated with two different notions of linking number: L_k^{fiber} and L_k^{DNA}, according to the considered scale, and one can show that $L_k^{fiber} = L_k^{DNA} + \text{const.}$ [3].

These two examples illustrate how the same issue for the same object can be tackled within different models according to the scale of description (also termed the *averaging scale* since it puts a limit between the details explicitly described and those, of smaller scale, accounted for in an effective average fashion). Considering one rather than the other model depends on the response properties to be analyzed and the nature of the stimuli exerted on the objects, i.e. whether the external stresses and boundary conditions are applied at the DNA or fiber level.

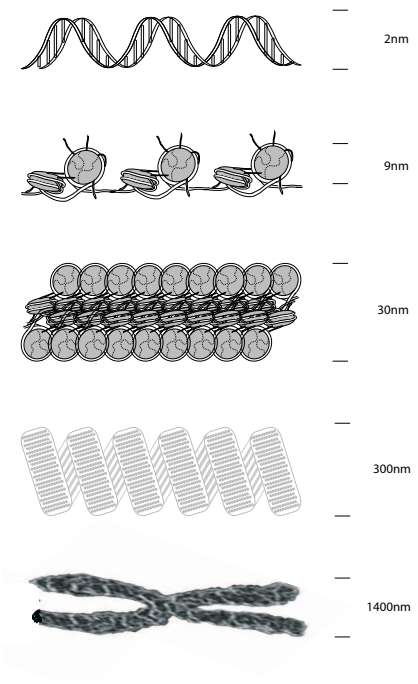


Figure 1: Hierarchical organization from DNA to chromosome, identifying well-defined structures at separated scales: DNA double-helix, nucleosomes and beads-on-string fiber, 30 nm chromatin fiber, chromatin loops, up to the entire chromosome

2.3 Chromatin dynamics and stochasticity

In the previous sub-section, we have described two examples of a non trivial multi-scale modelling, focusing on the connections between structural or kinematic models at different scales (respectively DNA and chromatin fiber scales) and working out the relation between their effective parameters. Similar bottom-up connections can be derived for chromatin dynamics. The cross-level couplings result in remarkable emergent properties (i.e. absent when the elements where they occur are considered in isolation). For instance, coordinated enzymatic histone-tail post-translational modifications generate alternating conformational changes at the chromatin fiber level [4] [5]. The associated '*chromatin breathing*' involves several nucleosomes and exhibits a far supra-molecular period of a few tens of minutes, evidencing the increase in the spatial and temporal scales of relevance for observing this emergent phenomenon, although the basic events occur at molecular scale.

Another bottom-up issue is to determine and describe the *level of stochasticity* present at the different scales and its functional consequences. The basic stochasticity originates in thermal fluctuations. In case of independent fluctuations, the resulting behavior at higher scales is dominantly deterministic, since accumulating independent fluctuations average out according to the law of large numbers. Observing or invoking stochastic events at large scales thus requires strong coupling between the molecular mechanisms, or strong correlations between the successive events, that generate an anomalous large-scale behavior with non negligible fluctuations [15].

2.4 A complex regulatory scheme

Multi-scale chromosome organization does not lead only to bottom-up relationships: the very complexity of living systems and biological functions, in particular the chromosomes and transcription, lies in the presence of *feedbacks from upper scales onto elements at smaller scales that have settled in the course of evolution.* As an illustration, let us cite the control exerted by the chromatin fiber superstructure onto the protein-DNA binding events in linker DNA. Indeed, when a linker is embedded in a condensed chromatin fiber, with stacked nucleosomes acting as fixed anchoring points, its linking number L_k^{linker} is conserved. This topological constraint has energetic consequences, since any strain (e.g. twist or bend) experienced by the linker, in particular

the strains generated upon protein binding (e.g. intercalation [33]) should be compensated so as to preserve $L_k^{linker} = \text{const.}$, at some elastic energy cost. The net effect is a noticeable modification of the protein-binding energy landscape. This landscape modification is moreover controlled by any means for tuning the nature of linker anchoring onto the nucleosomes and ensuing tolerance in the linking number constancy; possible means are changes in the chromatin fiber conformation and compaction state, histone-tail post-translational modifications, presence of ions or linker histones.

The strong and evolutionary adapted influence of larger scales onto elements at smaller scales within a chromosome led us to propose a notion of *generalized allostery* [33] [19], referring with this term to conformational transitions and associated change in activity and function, induced by an effector acting at a remote site and with no direct physico-chemical link with the activity. We here suggest that the involvement of this specific effector does not follow from an inescapable physico-chemical law but from a mutual adaptation settled in the course of evolution. In support and illustration of this generalized notion, DNA and chromatin can exhibit an allosteric behavior insofar as the hyperstructure in which they are embedded (respectively the chromatin fiber or a chromatin loop) and topological frustration it generates, can induce a bistable behavior in their conformation, controlled through modifications of the hyperstructure and the mechanical constraints it induces at lower levels. For instance, linker DNA can pass from a straight conformation to a buckled one with different binding affinity towards intercalating proteins [33]: the buckled conformation appears as the 'active' form of the allosteric linker; the transition to this active form is triggered by the first DNA-binding event and controlled through the tuning of DNA anchoring onto the nucleosomes. Our generalized notion includes as a special case the notion of *nested allostery* in which a cascade of allosteric behaviors in nested sub-systems is initiated by the effector-binding event [26]. It also embeds extensions of the original notion of allostery, e.g. the localized allostery observed in large assemblies and involving only a sub-system delineated by mechanical constraints [32].

This discussion in fact faces a very general property: the embedding in a superstructure can modify the very individual potentialities of low-level elements, hence *precluding a plain bottom-up strategy*. In a similar spirit, turning to more general interactions and couplings than mechanical or structural ones, it is a whole gene network that underlies transcriptional regulation, moreover exerting feedbacks on its own nodes (a node represents here at the same time a gene and the protein it codes): presumably, some nodes will have different individual properties when embedded in the whole network. For instance, interactions experienced by a protein within the gene network might modify its chemical or physical properties, and allow it to establish different interactions than those observed in vitro, in isolation.

3 A necessary shift of paradigm

3.1 Limits of a multi-scale description

We have seen in the previous section that in order to properly account for the autonomous regulation and behavior of a biological system, a multi-scale approach should tackle jointly all the scales, with no way to a priori ignore some microscopic detail [19]. Obviously, any such proper modelling would rapidly reach high level of complexity, the higher the more faithful and realistic the model is, hence *ultimately intractable* [13] [14]. To circumvent this difficulty, we claim that both integrated modelling and supervised data analysis *should parallel the biological functional logic*.

The hierarchy of spatial scales should enter the scene only because these nested scales correspond to our different but all subjective views on the system and to our various experimental accesses, confined in current practice to a given scale. Data analyses and model predictions have

to be ultimately bridged with this hierarchical categorization and the integrated model itself should embed, as special restricted situations, models developed at well-defined scales.

More conceptual objections also arise. In particular, scale-wise description is currently associated with a step-wise, sequential conception of the processes and their time course, e.g. chromatin fiber assembly [25] or transcription initiation. Traveling the hierarchy of scales is even logically correlated with temporal progression along a sequence of events where each event is causally rooted in the previous one. The above conclusions on biological complexity and associated causal loops suggest that this view is presumably too naive for biological functions. In the case of chromatin fiber assembly, it properly follows our ways of reconstituting a chromatin fiber but misses essential adaptive dynamics and self-organization occurring in vivo to achieve in one and the same step various delicate balances between the energetic, steric and topological constraints (reflecting e.g. in the nucleosome positioning and individual conformation), to control multi-stable pathways, or to manage with competing chemical reactions or concerted factor recruitment. The sequential view on chromatin dynamics have to be replaced with an *interaction network viewpoint*. It appears necessary to account of the joint dynamic behavior of numerous interconnected elements, for instance DNA sites and their chromatin surrounding, nucleosomes, or cofactors and coregulators [6] [18]. Hence, not only the spatial and structural models but also the dynamic descriptions should evolve into more integrated models.

Note finally that in biology, by contrast to physics, there is no ubiquitous relation between time and space scales, namely the events occurring at the smallest spatial scales are not necessarily the fastest ones (think for instance to ionic currents across membranes or molecular motor motion). This remark further evidences that a *splitting according the scales has no intrinsic relevance* for biological systems.

3.2 Need of a self-scaled, interconnected, function-dependent modelling

The above-discussed limitations and caveats about multi-scale descriptions lead us to propose a drastic change in the paradigm underlying chromatin modelling and more generally the modelling of biological systems, namely a shift from a scale-dependent focus to a *function-dependent* focus. A function-based analysis is required to account, through the notion of function, of the embedding of living systems or sub-systems *within an evolutionary history*. Speaking of the 'functions' of a biological system (by contrast to the 'properties' of a physical system) is a short-cut underlining the *cross-level consistency of all involved mechanisms, following from their co-evolution and resulting mutual adaptation.* Here, living systems meet the artificial ones, with the striking difference that living systems, following millions of years of trials and selective improvements, achieve in general an extraordinary efficiency. They possibly involve a different logic than the one at work in man-made functional and regulatory schemes.

Another motivation to change drastically the descriptive and modelling framework is related to its required *robustness*. Indeed, any modelling (and in fact any description) involves subjective restrictions, approximations, reference points ... and the only way to justify this arbitrary part of the modelling is to show the robustness of the predictions, namely that they do not depend on these subjective choices. In such a prospect, a model based on a multi-scale decomposition is not robust (except in some specific scale-invariant instances rarely encountered in biology): it explicitly involves the chosen scales, the details at each scales, the mechanisms considered at a given scale, and finally the coarse-graining procedures and closure approximations used in relating one scale to the other. On the contrary, a function-based viewpoint, relying on an objective biological fact, will be robust, at least qualitatively, with respect to the modelling choices. Quantitative agreement between predictions and observations might be then obtained with no change in the functioning principles nor in the logical/causal scheme, only requiring a tuning of the parameter values.

Figure 2: Within a function-based approach, all scales are to be considered jointly and the focus shifts to their functional interconnections: each level influences and is influenced by both lower and higher levels, according to causal loops settled in the course of evolution. This hallmark of complexity evidences the need of alternative approaches, replacing multi-scale models with a function-dependent, self-scaled decomposition in elementary processes.

This novel modelling approach, beyond being function-based, is also termed to be self-scaled and interconnected. By '*self-scaled*', we mean that the physical and experimental hierarchy of scales (Å atomic scale, nm base-pair scale, 10 nm bead-on-string, 30 nm chromatin fiber, 300 nm chromonema, up to the micron scale of the chromosomes and entire nucleus) is not an intrinsic and functional feature of the transcriptional regulation and more generally of any nuclear function. In consequence, it it does not provide a relevant nor operational decomposition. Rather, any biological function, in particular the chromatin function and still more specifically transcriptional regulation, has its own way to embed in real space and unfold its elementary processes *across these scales*, achieving *information circulation* in order to perform the processing of genetic and chromatin code and the concrete implementation of regulatory mechanisms, *each at the most convenient scale* (lowest energy cost, availability of co-regulators, adequate uptake mechanisms ensuring the stationarity of the fluxes ...). The biological function also demands an integrated implementation, what we summarize in the term '*interconnected*' (levels). In short, *functions and associated information processing pervade and exploit all scales jointly in their own evolutionary adapted way*, and the design of our modelling should account for and take the largest benefit of this point (Figure 2).

3.3 Basic elementary processes in transcription

Complexity reduction in living systems studies is presumed to rely on the identification of modules. Obviously, such a program makes sense *only if independent or weakly coupled modules can be delineated*. The precise requirement is that the behavior associated with a module is robust, qualitatively independent of the surroundings and inputs; it possibly affects quantitatively the model predictions but not the logical/causal scheme, kind of behavior and regulatory mechanisms. We have demonstrated in § 2 that defining modules as being well-defined structures each observed at a given scale does not fulfill this requirement. We thus propose another way of dissecting the biological system under consideration (here the chromosome) neither according to the experimentally isolable entities (nucleotides, nucleosomes, ...) nor according to the space and time scales, but according to the function [14].

The main idea is to consider basic processes articulating several physico-chemical mechanisms and unfolding across the scales, in a function-oriented scheme involving effective inputs and outputs, so as to get elementary 'responsive' or 'active' building blocks [14]. We term these multi-level and operationally dedicated mechanisms '*basic elementary processes*'. They intend to provide elementary links of an *effective network* achieving the same function and the same regulatory control. They are 'elementary' in the sense of being indecomposable, i.e. meaningful and describable only as a whole. We again underline that they fundamentally differ, in their spirit, of

modules introduced as building blocks of well-defined scale. Their description, modelling or experimental observation involve several scales and might extend in time, hence essentially crossing the standard levels of description and thinking. They describe elementary steps of information processing, transformation, circulation. Actually, we might consider a hierarchy by grouping the most elementary processes of our library into more complex ones, still non-autonomous (hence being only a part of a function) up to completion into a function (that can be defined formally, if necessary, as a pattern of information processing, involving mechanisms and parameters at different scales). A function is adaptive and robust, in the sense of being context-independent: it should be properly achieved in various external conditions and internal states. On the contrary, elementary processes will be context-dependent so as to maintain the function they participate in. Elementary processes (and even more the values of their parameters) are precisely highly sensitive to any external stress, changes in the surroundings or in the cell internal state, in order that the function is robust.

Let us implement more explicitly this program in the context of transcription. The base-pair, DNA and chromatin-fiber well-separated levels are replaced by the following processes:

— *DNA bending*, including its consequences on DNA affinity for various proteins [22];

— *chromatin tethering*, including as a consequence the invariance of the linking number L_k^{fiber} of the end-tethered chromatin loop, and also the invariance of the linking number L_k^{DNA} of the underlying DNA stretch; it follows that any imposed strain generates mechanical constraints in DNA and modifies its protein-binding energy landscape [33];

— *histone-tail post-translational modifications* with both chemical repercussions (in terms of recognition and recruitment of specific factors [29] [6]) and physical repercussions, either electrostatic (change of the local charge density) or mechanical (change of the local anchoring of DNA onto the nucleosome hosting modified histone(s) [33]);

— *hypercycles* involving the coordinated alternation of two reverse enzymatic reactions, e.g. acetylation and deacetylation; they prescribe a rhythm, i.e. a kind of internal clock inside the chromatin fiber, and provide the very first, non specific step of transcriptional initiation, termed chromatin breathing [4] [5].

— *DNA allostery*, where the chromatin fiber hyper-structure induces bistability properties at the DNA level (or more generally a switch potentiality following from the frustration induced by mechanical or topological constraints) controlled by the fiber compaction and triggered by DNA binding events or transactions [19] [17].

4 Tackling complexity in epigenomics

4.1 Several explanatory schemes

For any biological structure or process, several explanations and several levels of causality actually coexist, demanding to be articulated. Indeed, biological systems can be explained at the same time by a set of correlated mechanistic steps or by invoking an ecological (or even 'economical') balance of inputs and outputs, within an adaptive, evolutionary perspective. To take a simile, switching on a car into motion involves both the specific contact key and general thermodynamic principles of motor functioning. The relevant explanation will thus depend on the perspective, either evolutionary, mechanistic or therapeutic; in other words according to the causal representation to be provided, either in terms of the whole evolutionary consistency of the organism, of the outputs at a given point or response to a prescribed perturbation.

For instance, an allosteric reaction can be described at the molecular scale, as a succession (possibly highly complex, e.g. networked) of molecular interactions and modifications, following

physico-chemical laws. It might also be seen as a trick invented, stabilized and improved in the course of evolution. Our approach precisely *aims at reconciling mechanistic and evolutionary explanatory frame.* Its function-dependent focus refers to the inescapable natural selection, while the investigations of processes brings back to physic-chemical mechanisms, dissipative structures and self-organization. Additional schemes, for instance the information-theoretic notions of *genetic code* and *genetic program* have also to be considered to capture a comprehensive understanding of, say, cell differentiation in all its complexity [14]. As a way to bridge these complementary explanatory within a novel frame, we are developing a notion of *chromatin code*, reminiscent of the corpus of ciphers, codes, grammars, languages, computations and calculus developed in (theoretical, abstract) computer science, but with an architecture and logic of its own [6] [17].

4.2 Inverse renormalization-group

The strong interplay between the scales described in the previous section, where possibly some microscopic details directly influencing, or influenced by, the structure and dynamics at large scale, is reminiscent of situations encountered in physics, known as *critical phenomena*, where all scales are relevant and do not separate. In this context, a plain bottom-up representation of causality is questionable since the couplings and assembly of processes are strongly nonlinear and intricate: the assembly of two objects each described by a quantity X is not described by $2X$ but by $a(X)X$. A way out such difficulty is a *renormalization procedure* in which all the direct and indirect couplings and correlations are accounted for as a whole, through an effective, integrated and self-consistent contribution, here $a(X)X$ with $a(X) \neq 2$. More generally, renormalization-group offers a way to qualitatively understand and quantitatively predict anomalous behaviors associated with criticality [15]. For biological systems, criticality is even stronger and far more difficult to tackle since macroscopic features might exert some feedback on the very properties and behavior of the smallest scale elements, hence requiring some 'top-down' analysis. Plain top-down representation is confronted to an issue of information lack, since we try there to infer a detailed description from coarser ones with (far) less degrees of freedom.

To solve these difficulties, we are developing a novel mathematical framework, that we termed '*inverse renormalization-group*, intending in particular to provide a mathematical zoom to describe and understand biological functions and regulatory schemes [7]. Its design is parallel to the biological logic and implementation, insofar as it relies on the main theoretical principle of biology, namely evolution, natural selection and ensuing optimized fitness of biological systems. The information missing in top-down approaches is to be injected through a biological, evolutionary, adaptive, consistency argument, assessing that the microscopic level is so consistently related (co-evolved, mutually adapted and optimized) to the macroscopic level that the very existence of the macroscopic level brings information on the underlying ones. Thus, in adaptive systems (either living or man-made) evolution and selection allow a top-down inference that is essentially different from the inference scheme relevant in physics. Tackling complexity is precisely achieved here, in the reduction by ways of some optimization argument (here inter-level consistency and adaptive robustness) and educated microscopic implementation, of the whole system.

Additional features of biological systems motivate the development of this novel mathematical formalism [7]. We have yet underlined that in the set of biological functions or in the set of elementary processes from which they are built arise sub-categories, associated with various levels of emergence (for instance raw *vs* effective parameters, single-scale *vs* emergent elementary processes ...). A major conclusion follows: there is *no way out a continuous view* with fuzzy boundaries between the levels, the elements, and the elementary mechanisms or events. Moreover, biological systems do not exhibit a 'natural' underlying real space-time structure, but rather *several superimposed proper times* (that of cycles and hypercycles; that of evolution) and *several superimposed intrinsic topologies* (that of development –curved; that of regulatory networks –

infinite-dimensional; that of constraints – tensegrity) with competition (hence possibly frustration if beneficial, e.g. to generate multi-stability and switches) or cooperation (hence synergy).

5 Conclusion

Given a biological system (or even yet a physical or chemical one), the proper model is devised to answer a well-posed question. It cannot intend (and should not) to provide a universal representation of the system, nor to fully capture all its properties whatever their nature, scale and context. Otherwise it would resemble the scale-1 empire map described by Borges [10]. Modelling hence requires to focus on some elements and facts while ignoring a large amount of information of no direct relevance for the addressed issue. The merit of a good and efficient model is precisely to extract and enlighten those, and only those features of the system that are presumably relevant for solving the limited, well-delineated issue under consideration. Multi-scale viewpoint is at the same time required to design experiments or to analyze and interpret data, but irrelevant due to the essential cross-scale ('up-and-down') coupling following from evolution. Inspired by approaches currently followed in theoretical physics, this sorting of relevant information is usually done in the real time and space, according to the scales. We have proposed here to rather base this a priori sorting according to the biological functions. This leads to a drastic change in the paradigm underlying the design of experiments and choice of model systems, supervised data analyses and modelling, namely *a shift from a scale-dependent focus to a function-dependent focus*.

References

[1] Anderson, P.W. (1972) More is different: broken symmetry and the nature of the hierarchical structure of science, *Science* **177**, 393–396.

[2] Bancaud, A., Conde e Silva, N., Barbi, M., Wagner, G., Allemand, J.F., Mozziconacci, J., Lavelle, C., Croquette, V., Victor, J.M., Prunell, A., and Viovy, J.L. (2006) Structural plasticity of single chromatin fiber revealed by torsional manipulation, to appear in *Nature Struct. Mol. Biol.*

[3] Barbi, M., Mozziconacci, J., and Victor, J.M. (2005) How does the chromatin fiber deals with topological constraints, *Phys. Rev. E* **71**, 031910.

[4] Benecke, A (2003) Chromatin dynamics are a hyperstructure of nuclear organization, in *Proceedings Modelling and Simulation of Biological Processes in the Context of Genomics*, Amar P., Kepes F., Norris V., and Tracqui, P. (eds.) Platypus Press, pp. 31–40.

[5] Benecke, A. (2003) Genomic plasticity and information processing by transcriptional coregulators, *ComPlexUs* **1**, 65–76
— see also published comment: Buchanan, M. (2003) *ComPlexUs* **1**, 65–69.

[6] Benecke, A. (2006) Chromatin code, local non-equilibrium dynamics, and the emergence of transcription regulatory programs, to appear in *Eur. Phys. J. E*, Focus Point on "Structure and Dynamics of DNA".

[7] Benecke, A. and Lesne, A. (2006) Inverse renormalization-group, in preparation.

[8] Ben Haim, E., Lesne, A., and Victor, J.M. (2001) Chromatin: a tunable spring at work in our chromosomes, *Phys. Rev. E* **64**, 051921.

[9] Ben Haim, E., Lesne, A., and Victor, J.M. (2002) Adaptive elastic properties of chromatin fiber, *Physica A* **314**, 592–599.

[10] Borges, J.L. (1972) On the impossibility of drawing a map of the empire on a scale of 1, in *A universal history of infamy* [Historia universal de la infamia] (1935).

[11] Bouchiat, C. and Mezard, M. (1998) Elasticity model of a supercoiled DNA molecule, *Phys. Rev. Lett.* **80**, 1556–1559.

[12] Cui, Y. and Bustamante, C. (2000) Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure, *Proc. Natl. Acad. Sci. USA* **97**, 127–132.

[13] Lavelle, C., Benecke, A. (2006) Chromatin physics: replacing multiple, representation-centered descriptions at discrete scales by a continuous, function-dependent self-scaled model, to appear in *Eur. Phys. J. E*, Focus Point on "Structure and Dynamics of DNA".

[14] Lavelle, C. and Benecke, A., and Lesne, A. (2006) Function-dependent self-scaled interconnected chromatin models: integrated functions as a combination of basic elementary processes, in preparation.

[15] Lesne, A. (1998) *Renormalization methods*, Wiley, New York.

[16] Lesne, A. (2006) Multi-scale approaches, article MS101 in *Encyclopedia of Mathematical Physics*, J.P. Françoise, G. Naber, and T.S. Tsun (eds.) Elsevier.

[17] Lesne, A. (2006) Chromatin code: beyond an histone code, to appear in *Eur. Phys. J. E*, Focus Point on "Structure and Dynamics of DNA".

[18] Lesne, A. (2006) Complex networks: from graph theory to biology, submitted to *Letters in Mathematical Physics*.

[19] Lesne, A. and Victor, J.M. (2006) Chromatin fiber functional organization: some plausible models to appear in *Eur. Phys. J. E*, Focus Point on "Structure and Dynamics of DNA", — see also the commentary of this paper by H. Schiessel in the same issue.

[20] Mozziconacci, J., Lavelle, C., Barbi, M., Lesne, A., and Victor, J.M. (2006) A physical model for the condensation and decondensation of eukaryotic chromosomes, *FEBS Letters* **580**, 368–372.

[21] Mozziconacci, J. and Victor, J.M. (2003) Nucleosome gaping supports a functional structure for the 30 nm chromatin fiber, *J. Struct. Biol.* **143**, 72–76.

[22] Parvin, J.D., McCormick, R.J., Sharp, P.A., Fisher D.E. (1995) Pre-bending of a promoter sequence enchances affinity for the TATA-binding factor, *Nature* **373**, 724–727.

[23] Peterson, C.L. and Laniel, M.A. (2004) Histones and histone modifications *Curr. Biol.* **14** R546–R551.

[24] Polanyi, M. (1968) Life's irreducible structure, *Science* **160**, 1308–1312.

[25] Ridgway, P. and Almouzni, G. (2001) *J. Cell Sci.* **114**, 2711–2712.

[26] Robert, C.H., Decker, H., Richey, B., Gill, S.J., and Wyman, J. (1987) Nesting: Hierarchies of allosteric interactions, *Proc. Natl. Acad. Sci. USA* **84**, 1891–1895.

[27] Schiessel, H. (2003) The physics of chromatin, *Journal of Physics - Condensed Matter* **15**, R699–R774.

[28] Sivolob, A., Lavelle, C., and Prunell, A. (2003) Sequence-dependent nucleosome structural and dynamic polymorphism. Potential involvement of histone H2B N-terminal tail proximal domain, *J. Mol. Biol.* **326**, 49–63.

[29] Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications, *Nature* **403**, 41–45.

[30] Strick, T.R., Allemand, J.F., Bensimon, D., Bensimon, A., and Croquette, V. (1996) The elasticity of a single supercoiled DNA molecule, *Science* **271**, 1835–1837

[31] van Driel, R., Fransz, P.F., and Verschure, P.J. (2003) The eukaryotic genome: a system regulated at different hierarchical levels, *J. Cell Science* **116**, 4067–4075.

[32] Van Holde, K.E., Miller, K.I., and Van Olden, E. (2000) Allostery in very large molecular assemblies, *Biophysical Chemistry* **86**, 165–172.

[33] Victor, J.M., Ben Haim, E., and Lesne, A. (2002) Intercalation and buckling instability of DNA linker within locked chromatin fiber, *Phys. Rev. E* **66**, 060901.

[34] Wolffe, A.P. (1998) *Chromatin: structure and function*, Academic Press, New York.

[35] Woodcock, C.L. and Dimitrov, S. (2001) Higher-order structure of chromatin and chromosomes, *Curr. Opin. Genet. Dev.* **11**, 130-135.

[36] Woodcock, C.L., Grigoryev, S.A., Horowitz, R.A., and Whitaker, N.A. (1993) Chromatin folding model that incorporates linker variability generates fibers resembling the native structures, *Proc. Nat. Acad. Sci. USA* **90**, 9021–9025.

Static and dynamic methods for the analysis of biological networks

Ron Y. Pinter[1]

[1] Dept. of Computer Science Technion - Israel Institute of Technology

Abstract

Elucidating the behavior of biological networks requires the development and the application of computational tools and techniques. Naturally, different methods are suitable for different kinds of analyses, e.g. extracting qualitative vs. observing quantitative properties, employing discrete vs. continuous modeling, and utilizing static vs. dynamic analysis. Specifically, methods that are static in nature extract structural and semantic properties from the description of a network or a pathway; dynamic methods aim at understanding and predicting the functional characteristics of a network's or a pathway's behavior.
In this talk I will present a number of methods that we have devised recently to analyze both metabolic as well as regulatory pathways along with the biological consequences that they have yielded. Specifically, the talk will include a detailed description of:

1. A method for the alignment of metabolic pathways [Pinter et al., Bioinformatics 2005]: Here we present MetaPathwayHunter, a tool that - given a query pathway and a collection of pathways - finds and reports all approximate occurrences of the query in the collection, ranked by similarity and statistical significance. It is based on an efficient graph matching algorithm that extends the functionality of known techniques. The program also supports a visualization interface with which the alignment of two homologous pathways can be graphically displayed. We employed this tool to study the similarities and differences in the metabolic networks of the bacterium Escherichia coli and the yeast Saccharomyces cerevisiae, as represented in highly curated databases. We reaffirmed that most known metabolic pathways common to both species are conserved. Furthermore, we discovered a few intriguing relationships between pathways that provide insight into the evolution of metabolic pathways. We conclude with a description of biologically meaningful meta-queries, demonstrating the power and flexibility of our new tool in the analysis of metabolic pathways.

2. HFPN-based simulation of metabolic pathways [Assaraf et al., JTB 2006]: Here we devise a hybrid functional Petri nets (HFPN) modeling of folate metabolism under physiological and antifolate inhibitory conditions. In selecting the appropriate simulation method for this task we considered both the inherent computational properties of the underlying models as well as the features of the simulation systems that embody them. A careful comparative study showed that the GON system, based on the HFPN model, was the most suitable for modeling metabolic pathways as the one chosen. We performed a large number of virtual experiments on our comprehensive model of the pathway under study, measuring its reaction to four different anti-cancer drugs. The results were highly accurate when compared to experimentally obtained values and revealed many interesting phenomena; the effectiveness of the system allowed large-scale experimentation leading to filtering of some of the more intriguing results for in vitro validation. Furthermore, this HFPN-based simulation offers an inexpensive, user-friendly, rapid and reliable means of preclinical evaluation of the inhibitory profiles of antifolates.

3. Analysis of transient behavior in developmental pathways [Rubinstein et al., under review]: Here we present a computational model that allows for qualitative analysis of regulatory pathways, enabling the examination of characteristics such as transient behavior, robustness, and sensitivity

to initial conditions, in an effective manner. We have extended the Boolean network model, which has limited modeling power, to a richer albeit discrete network model, while maintaining computational efficiency. Moreover, we have borrowed a simple technique for the representation of functions, namely Karnaugh maps, to elucidate and visualize the behavior of the pathways under study. We have applied our method to analyze the transience and robustness of a representative developmental pathway, namely early meiosis in budding yeast. Some of our analytic observations, such as the pathway's response to premature expression of a key regulator, were validated in the lab and were found to be in agreement with experimental data. Furthermore, our analysis predicts new modes of regulation by which negative feedback loops accomplish their roles.

I will conclude with some suggestions concerning the potential integration of the methods into a unified framework so as to fit the end-users' needs.

Rule-based modelling and reasoning for biochemical applications

O. Andrei[1]

[1] INRIA-LORIA, 615 rue du Jardin Botanique, Villers les Nancy, France

Abstract

I present the context and the motivation of my PhD work, the modelling of biochemical systems and the analysis of their behavior by means of rewriting systems. I end with some preliminary results obtained from modelling a chemical reactor and the possibilities of analyzing some of its properties.

1 Introduction

Providing a formal description for the structure and function of biological systems, as well as formal tools for reasoning about their behavior has been identified as a scientific *grand challenge*. In the context of computer science applied to biology and chemistry, an important amount of work is currently devoted to automatic model generation. The goals are to better understand what happens in a cell or a solution of molecules, and to predict its behavior. Applications are various in the health domain (for example, drug design) or in chemical industry (quality of fuels,...).

The descriptions of biomolecules or chemical reactants belong to specific classes of graphs. Moreover the interactions between the reactants involve rules applied on these classes of graphs and controlled by numerical data or specific filters. In this context of biochemical systems, typical problems are the exhaustive generation of all possible states of the system, the detection of specific states, or the prediction of producing specific states. There is a real need of concepts, methods, and tools to model interactions in these systems and address the previous problems.

Previous work and expertise have already been developed in the Protheo team as presented in [4, 5]. The authors of these papers explore the field of automated generation of chemical reaction mechanisms. The choice of modelling these particular mechanisms by means of rewriting systems is motivated by the fact that the chemical reactions are naturally expressed as conditional rewrite rules and that the control of the chemical reactions chaining is easy to describe using a strategy language. As a formal basis for the modelling process, the authors proposed a class of graphs, called molecular graphs, and a graph rewriting relation where vertices are preserved and only edges are changed.

Among many other approaches, we found interesting connections with [6], where a method for generating a biochemical reaction network from a description of the interactions of components of biomolecules, is specified in the form of reaction rules. Another related work is the Biochemical Abstract Machine (BIOCHAM) [8] that provides a formal modelling environment for network biology, in particular for representing and analyzing protein-protein and protein-DNA interaction networks, consisting in formal tools for modelling, querying, validating, and completing biomolecular interactional networks.

2 Main goals

The subject of my PhD thesis is to study the concepts, methods, and tools to model interactions in biochemical systems and address the problems of behavior prediction, exhaustive generation of all possible states in case of termination, detection of specific states in possibly infinite search spaces.

The goal is to develop the foundations for an adequate calculus based on graph rewriting, taking into account numerical information and constraints induced by physical environment and knowledge (for example, kinetics information for chemical reactions).

This framework should offer a language for describing the molecules and the reactions, together with reasoning and predicting capabilities that will be studied and compared. Such modelling is usually faced to an inherent combinatorial complexity and reasonably efficient solutions. Moreover, the developed models have to be discussed and validated by biochemists.

3 Methodology

The first step consists in the identification of concepts necessary to model the interactions in biochemical systems based on a bibliographic study.

The next two steps consist of defining a language for describing the initial molecules in the system and the reactions to be applied, and providing a formal model for reasoning and predicting the behavior of the systems. A starting point should be to extend in an appropriate way the rewriting calculus [3]. After a first stage that might be mostly syntactic, I will take into account numerical constraints to achieve more realistic models, leading to a concept of hybrid rewriting. For this context, I will adapt the former reasoning and predicting tools.

To support the ideas and discussions with biochemists, a research prototype will be designed with the TOM environment in the Protheo team (**`http://tom.loria.fr/`**).

This work will take benefit of many interactions with other members of the Protheo team working on graph rewriting, constraint rewriting, and probabilistic strategies.

4 Preliminary results

During my internship within the Protheo team, I studied the capabilities of TOM for modelling the class of molecular graphs and its associated graph rewriting relation by means of term rewriting [1].

In [2] we give a new prototype for the oxidizing pyrolysis using TOM. We present the model as a particular case of an artificial chemistry with the molecules/molecular graphs as algebraic terms, the chemical reactions as rewriting rules on terms, the reactor dynamics as a composition of elementary strategies (the reactions) and strategy operators. The formal background of strategic rewriting is quite relevant for the considered problem:

1. chemical reactions are naturally expressed by chemists themselves using conditional rules;
2. matching power associated with rewriting allows retrieving patterns in chemical species,
3. defining the control on rules is essential for designing automated mechanisms generators in a flexible way and controlling combinatorial explosion. This gives the possibility to the chemist for activating and deactivating reactions patterns, and for tuning their application during each stage.

The first prototype for the oxidizing pyrolisis, GasEl [4], was implemented in ELAN, a system developed in the Protheo team for specifying and prototyping deduction systems in a language based on rewrite rules controlled by strategies ((**`http://elan.loria.fr/`**). The main technical difficulty with GasEl, consisted in the encoding of reaction patterns on GasEl terms that correctly

simulates the corresponding transformation on molecular graphs. The TOM implementation provides another approach to this problem, while keeping the same molecular graph rewriting relation, and preserving the same chemical principles and hypotheses as in GasEl.

Our second concern in [2] was to explore the formal island concept and methodology on a significant example, a chemical reactor. The objective of the formal island approach to extend the expressivity of the host language with higher-level constructs at design time is well-illustrated in this example. From this point of view, the TOM implementation appeared to be quite convenient to implement chemical rules with conditions and actions expressed in the Java host language. The control is expressed with a high-level language of strategies which makes now possible to reason about formal properties, especially the termination property of each phase of the reactor [1]. This illustrates the idea to perform formal proof on the formal island constructions.

It may be worth noticing that the rule-based approach on graph structures has also been studied in the modelling of signal transduction networks [7] and metabolic pathways [9] in the domain of biological systems and protein interactions. Our model of chemical reactor seems to be easily adaptable to these domains.

References

[1] Oana Andrei. *Term graph and chemical rewriting*. Internship report, LORIA, Protheo Team, Nancy, France, September 2005.

[2] Oana Andrei, Liliana Ibanescu, Hélène Kirchner. *Non-intrusive formal methods and strategic rewriting for a chemical application*. In Festschrift in Honor of Prof. Joseph Goguen, 2006. Accepted.

[3] Horatiu Cirstea, Claude Kirchner. *The rewriting calculus - Part I and II*. In *Logic Journal of the Interest Group in Pure and Applied Logics*, 9(3):427-498, May 2001.

[4] Liliana Ibanescu. *Programmation par régles et stratégies pour la génération automatique de mécanismes de combustion d'hydrocarbures polycycliques*. Thèse de Doctorat d'Université, Institut National Polytechnique de Lorraine, Nancy, France, June 2004. (**`http://www.loria.fr/~ibanescu/these_en.html`**).

[5] Olivier Bournez, Liliana Ibanescu, Hélène Kirchner. *From Chemical Rules to Term Rewriting*. In Proceedings of RULE'05, ENTCS, RDP Conference, Nara, Japan, 2005.

[6] James R. Faeder, Michael L. Blinov, Byron Goldstein, William S. Hlavacek. *Rule-Based Modeling of Biochemical Networks* In *Complexity*, Volume 10, Issue 4, 2005. Special Issue: Understanding Complex Systems: Part II, 22-41.

[7] James R. Faeder, Michael L. Blinov, William S. Hlavacek. *Graphical rule-based representation of signal-transduction networks*. In *ACM Symposium on Applied Computing*, pages 133-140, 2005.

[8] F. Fages, S. Soliman and N. Chabrier-Rivier *Modelling and querying interaction networks in the biochemical abstract machine BIOCHAM*. In *Journal of Biological Physics and Chemistry*, 2004, Volume 4, Number 2, p. 64-73

[9] F. Roselló, G. Valiente. *Analysis of metabolic pathways by graph transformation*. In H. E. et al., editor, *2nd International Conference on Graph Transformation - ICGT'04*, Roma, Italy, volume 3256 of *Lecture Notes in Computer Science*, pages 70–82, Springer, 2004.

A generic formal framework for dynamic modelling of biological regulatory networks

Vincent Bassano[1]

[1] Université de Paris3-Sorbonne Nouvelle. vincent.bassano@univ-paris3.fr

Abstract

Functional genomic is a topic studying the functional properties of the macromolecules encoded by the genes. Its goal is to understand exhaustively a biological system. For this, a great number of heterogeneous data has to be put into relations. Paradigms from mathematics and computer science can be used to model these data.
Studying and modelling biological regulatory networks is part of functional genomic. A family of modelling (called modelling of René Thomas[20, 18, 19, 5, 4, 7, 6, 11]) is especially attached to the study of the temporal properties of small sized systems([16, 17, 12, 9, 8]). This family uses qualitative data (global behaviour in big ranges of values) instead of quantitative data (numerical values).
Our goal in this paper is to propose a generic formal framework unifying the R. Thomas modelling.

Introduction

One of the important issues of post-genomic is to understand the criss-crossed interactions between genes inside the cell. R. Thomas works has given birth to a set of "on demand" modelling methods (Boolean or multi-valued values, with or without Snoussi hypothesis(cf [14, 13, 15]), synchronous or asynchronous dynamics, taking or not the functionality into account) to study qualitatively the dynamics evolution of biological regulatory networks.
All these methods have been inherited from a common basis and are adapted to the kind of system they model.
In this paper we will present a generic framework unifying the existing modelling and allowing to define new one.
We specially attract the reader attention on the definition of the family of *firing semantics* allowing defining on demand dynamical behaviour for the studied system and expanding the lone synchronous and asynchronous transitions of the R. Thomas model.

1 Context

1.1 Biological regulatory networks

A biological regulatory network is a complex interactions network, dynamically regulating the produce of genes expression. We will treat here of the modelling of networks (and not of their inference). We have a precise goal in mind for that: the study of their dynamical behaviour.
We will be able to treat a family of networks:

- Small sized (at most ten variables).
- All interactions between variables are known.
- The knowledge we have on the variable is **qualitative** and not quantitative.
- Time will not be considered as continuous. Instead we will take into account a succession of state $\{t_1, t_2, ...\}$.

1.2 R. Thomas modelling

One of the oldest way to model biological regulatory networks is quantitative and based on differential equations ([10]). It requires a fine definition of the parameter and function involved. this is one of its main drawback: we often are limited to a kind of qualitative knowledge.
To address this issue, René Thomas developed a **qualitative** approach (see Thomas and d'Ari book [21]).
The models based on this approach have, in a first time, used Boolean data (the so called "logical approach of R. Thomas") before being extended to treat any kind of discrete data ("generalized approach of R. Thomas" or "multi-valued model").

1.3 Discrete variables seen as switchs

From the beginning we have been speaking of discrete variables. But how are these variables made discrete?
All the modelling derived from the family of R. Thomas use a strong biological hypothesis: Physically, all the interactions between variables are sigmoid shaped (meaning two level separated by a threshold). This hypothesis is often biologically true.
It means that

- Variables do not act linearly on an other variables. They act more like a **swith** activated at a given level.
- This activation is linked to the threshold notion

see [2] for more details.

1.4 Data model in our generic framework

The goal of our generic framework is to study the temporal succession of the state of a system. For this we have defined the data manipulated by all the modelling of the R. Thomas family.

As an output of all modelling method, we want to obtain a **state graph**. A state graph is an oriented graph where

- the set of vertices is the set of all the possible state
- an arrow between two state indicate that one is the temporal successor of the other.

To follow a path in this graph is effectively to do a simulation.

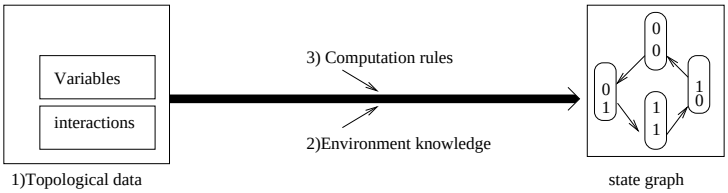


Figure 1: Data in and out : The three kind of knowledge needed in our generic framework and the state graph we obtain.

To compute this graph, biologists may have different kind of knowledge at their disposal. We, in our generic framework, have differentiated three kind of input knowledge

- the **topological knowledge** describing what are the variables of the system and their interaction
- the **parametrical knowledge** describing the background of the system. Topological knowledge is not enough to understand the behaviour of systems. We have to indicate that, in the studied system, some interactions are stronger than other. Actual behaviour of the system depend on value of some key parameters.
  The value of the parameters of a real system can be deduced from the observation of the stable and steady states of this system.

- the **set of rules** used to compute the dynamical behaviour of a system knowing its topological and parametrical knowledge.

Interestingly enough, only the whole topological knowledge of a system is necessary

- If parametrical knowledge is absent or partial, we can always study the behaviour of the system for **all** the value that may be taken by the non-defined.
- if we don't know how to "tune" the computation of the dynamical behaviour, we can always use standard set of rules such as the synchronous and asynchronous semantic of R. Thomas modelling

1.5 Many approach : The need for a generic framework

Due to the characteristic of the studied biological system, the first R. Thomas approach has been adapted into a great number of different modelling. Our study of those approach had led us to believe that one of the thing they have in common (beside dealing with discrete variable) is a generic structure more or less following the three kind of knowledge described in the preceding section (section 1.4).
It has naturally led us to define ([1],[2]) a formal generic framework for the discrete modelling of the biological regulatory networks.
Our generic approach is structured in three different layers, linked with the three kind of knowl-

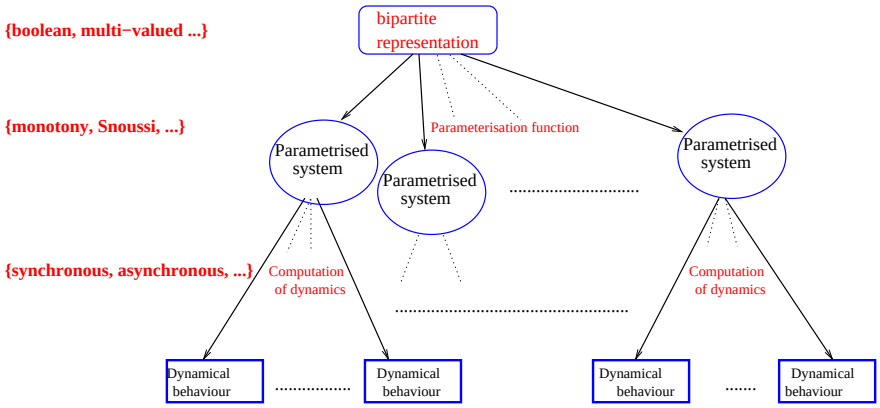


Figure 2: *Structure of our generic approach*

edge we have already described. Of course, as our approach is **generic**, each of these part has to encompass the many options encountered in the other modelling. We have

- a part dealing with **the description of the topology of the interactions**
- a part treating the issue of **the instantiation of the parameters**
- a part effectively computing **the dynamic**

In the rest of this paper, we will describe these three parts, as shown in figure 2, and explain why our framework is generic.
We can nevertheless already mention that for one system described, we may have many parameters value and that for each couple system-parameters we can compute as much possible dynamical behaviour than we have semantics.

2 Our generic formal framework

2.1 Topological layer : bipartite description

The classical modelling of biological regulatory networks use a description based on simple oriented graph where the vertices are the variables and where arrows represent interactions between two variables. This description ignores one essential knowledge: the way than (at least) two variables may combine to produce a more complex interaction with an other one.

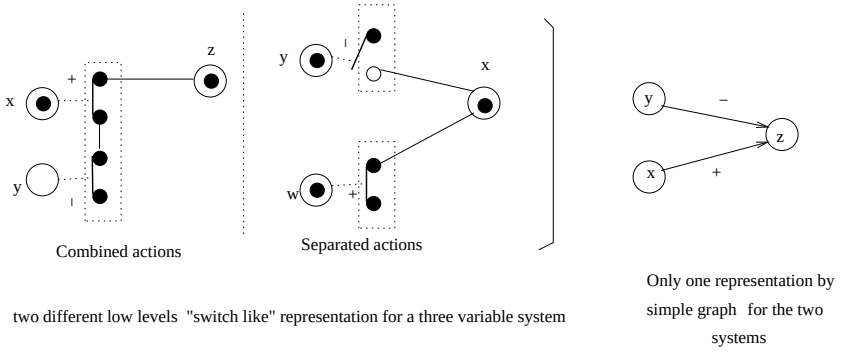


Figure 3: *Single graph representation is not enough to differentiate two different structure of networks*

We can see by the example described in figure 3 that the classical oriented graph representation do not make a difference between separated and combined action of two variable on another one. If we pursue the low level switch analogy seen in the precedent section (1.3) the simple graph representation only show that x and y act on z. It does not differentiate if these two actions are in "parallel" or in "series"

In fact, we deduce the need of a new object regrouping all the interaction acting combiningly. This object is used as a **transition** between the activating and the activated variables. These activating variables can (if conditions are correct) block or not the transition acting on the production of the activated variables.
These conditions are a kind of **ward**.

We can thus adopt for our generic framework a representation by **bipartite graphs** where we manipulate two set variables and transitions. A formal definition is called a resources graph

Définition - Resource graph :
A *resource graph* is a labelled oriented bipartite graph $\mathcal{G} = (V, T, Out, In)$ where

- V is a set of nodes called variables
- T is the other set of nodes called transitions
- Out is the set of arc from V to T, such as each arc $(u \rightarrow \tau)$ is labelled $(seuil(u,\tau), sign(u,\tau))$
- In is the set of arc from T to V

$(seuil(u,\tau), sign(u,\tau))$ is **a ward** and describe the conditions that block the transition when false

- if $sign(u,\tau)$ is $+$ then value of u must be greater or equal than $seuil(u,\tau)$
- if $sign(u,\tau)$ is $-$ then value of u must be strictly inferior than $seuil(u,\tau)$

This description is generic. The other approach (Boolean, multi-valued, by single graph...) can always be easily encompassed by this one. For example, it is trivial to pass from a single graph representation to this bipartite one: you just have to associate a transition to each arc.

2.2 Parametrical layers : transition oriented parameters

We now have to define the parameters influing the behaviour of the system.
Many possibilities exist in other method of the R. Thomas family. Basically it can be described as

the effect on a variable of the action of its neighbouring variables.
We have seen on the previous subsection that action of a group of variables can be grouped by transition (with ward). We now have to define, for each variable the effect of transition (or set of transition) acting on it.

Thus we define a **parameterisation function** $\mathcal{K}(x, \mu)$oriented by transition (opposed to other, less generic function oriented by variable; see [1] for a more complete review).
If a set of transition μ have a positive action on a variable x, then the value x tend toward $\mathcal{K}(x, \mu)$. Its define the **image state** (or attractor) of a variable.
For example, if we define a function $\mathcal{K}$ as:

$$\mathcal{K}(x, \{\})=0,\ \mathcal{K}(x, \{t1\})=1,\ \mathcal{K}(x, \{t2\})=1,\ \mathcal{K}(x, \{t1, t2\}) = 2$$

This means that if x has no influence it tend toward 0. If it is helped by one transition, it tends toward 1. If it is helped by the two transitions, it tends toward 2.
If we choose a different value for the instantiation function (let say $\mathcal{K}(x, \{\})=2$) we have a different behaviour of our system

2.3 Dynamic layer: firing semantic

In this layer, our approach to compute dynamic is very different from the one used in R. Thomas classical models. In the classical modelling there is no generic way to compute the dynamic, there is only a finite number of algorithm. Two of this way to compute dynamics are common in every models of the family: synchronous and asynchronous.

Let us precise that the original approach does not have the concept of multiple ways to compute dynamic. The two dynamics are just seen as one method of computation with more or less precision.

Unfortunately, we know that having only one way to compute dynamic is not enough to understand the complex behaviour of some systems: We are not able to use some important information such as the kinetic of the occurring reaction, the proximity of some expression level, ...

To apply the classical methods to a studied system, one has often to bend the computation of dynamics to its own need.

We propose here the first generic approach allowing to define a **family** of dynamics. This approach is based on the transition notion. If, at a given time, all the conditions (the wards) on a transition are true, we can say that the transition is **enabled**.
Moreover, we can define the **marking** of a variable at a given time t as a set of transition such as

- if the variable receive a positive influence from a transition τ then τ must be in the marking.
- if the variable do not receive a positive influence from a transition τ then τ must not be in the marking.

Thus, we can define the **firing** of a transition τ. It is an update of the marking of all the variable after the transition. If the transition is enabled τ will be present, if the transition is not enabled τ will be absent.
If we couple this **marking** notion (a set of transitions) with the **parameterisation function** (which need also a set of transition) we understand that firing transition is, in fact, **updating the image** of the variable after the transition.
Knowing the values of a variable v and its image at a given time t, we can easily compute the value of v at ne next update $t + 1$ by applying the formula

$$val(v, t + 1) = f(val(v, t), image(v, t))$$

By doing this, we have defined a whole new family of dynamics at a given time t

1. We choose a set T' of transition to fire.
   Let us call θ the choice function: $T' = \theta(T)$.

2. We fire the transition of T'. We update the marking.
   We use this marking and the parameterisation function to compute $image(v,t)$ of each variable v after the T' transitions.
   We compute and and update the value of these variables by using the rule

$$val(v,t+1) = f(val(v,t), image(v,t))$$

In fact, there are as much dynamics in this family than there are possibility in defining θ and f

θ \ f	Identity	restricted	incremental	...
Identity	Identity			
Synchronous		Synchronous restricted	Synchronous incremental	*ad lib.*
Asynchronous		Asynchronous restricted	Asynchronous incremental	*ad lib.*
Asynchronous of Thomas		As. of Th. restricted	As. of Th. incremental	*ad lib.*
...		*ad lib.*	*ad lib.*	*ad lib.*

Figure 4: *A whole new family.*

2.3.1 θ : choice of the fired transitions

This choice of transition introduces non-determinism in the computation of dynamics. It can be coupled with a probabilistic approach (not treated here) in a context of simulation. We can, for example choose

- To not fire any transitions : the system do not evolve.
- To fire all the transition together. This is a synchronous semantic. For one state you have one and only one successor.
- To fire only one transitions. This is what is called "asynchronous" in the field of dynamics system
- To fire all the transition located just after a given variable v. This is the way the so-called "asynchronous" semantic works in the R. Thomas modelling. We have moved the asynchronism from the transition to the variable. We still have a choice to make: the variable v

- To group together set of transitions (based on known biological fact) and apply different rules to the different set. For example, we can separate the fast and the slow transitions and act asynchronously on the fast (you can't distinguish them so they apply together) and synchronously on the slow

2.3.2 The f function

We still have to define the f function used to calculate the new value of a variable based on the old value and the image. we can for example say that

- f ignores the image $val(v,t+1) = val(v,t)$: the system do not evolve.
- f is restricted : if the value of a variable tends toward an image then it will reach this value at the next time step. We have $val(v,t+1) = image(v,t)$:
- f is incremental : the value of a variable v evolve in direction of the image by the value of a given step ε : $val(v,t+1) = val(v,t) + / - \varepsilon$. Particularly useful when we have a step of 1

We can of course also (as in the definition of Θ) develop our own f function. More complex study are available in [1]

2.3.3 A whole new family of dynamics

As we can see in figure 4, choosing a f and θ give us a new way to compute a dynamic adapted to the studied system. As a remark we can add that the two classical dynamic of R. Thomas (the so-called synchronous and asynchronous) still exit in our model.
"Synchronous" is now Θ synchronous and f restricted. "Asynchronous" is θ asynchronous of R. Thomas (variable oriented) and f incremental.
We also see that other possible way to compute dynamic appears : we can imagin to use for example θ synchronous and f incremental. New, previously ignored, semantic of the dynamic appears.

Conclusion

We have proposed a way to model biological regulatory networks using the notion of transition.
It has allowed us to divide our architecture in three separate layers

- Description of interaction. This layer has been made generic by the use of bipartite graph. Our model may encompass all the other existing modelling: single graph of bipartite, Boolean or multi-valued, ... as long as the interaction are not linear.
- Definition of the parameters. This layer use a transition oriented parameterisation function. Such a modelling is efficient and can englobe all the other way to define parameters based on variables.

  We can also use simple rules on the functionality of string and loops in a system to easily define this parameters. Classical hypothesis (such as Snoussi hypothesis) can also be applied.

- Computation of the dynamic . This layer allows us to define a method based on the firing of transition to compute dynamic. We are able to define a whole new family of dynamic

This architecture is robust and modular. It can be easily extended and adapted to all the existing modelling of the R. Thomas family.
It defines a formal framework which can be use to model biological regulatory networks.

We plan to extend this framework in two directions

- to take into account some linear transitions
- to model not one system but a population of systems

We also want to use this formal framework to automatise reasonning on biological regulatory networks. For this, we use this framework to translate regulatory graphs into systems of universally quantified propositional formulae. This translation can be fruitfully used to modify the regulatory graph structure without changing its external behaviour. This lead to formal manipulation of biological regulatory networks (see[3]).

References

[1] V Bassano. *Modélisation formel des réseaux de régulations biologiques*. PhD thesis, Université d'Évry, 2004.

[2] V Bassano and G Bernot. Marked regulatory graphs: a formal framework to simulate biological regulatory networks with simple automata. In *14th IEEE International Workshop on Rapid System Prototyping (RSP'03)*, pages 93–100, 2003.

[3] V. Bassano, P. Le Gall, and G. Bernot. A propositional logic approach to manipulate topology of biological regulatory networks. In *CMSB'05, Edimbourg*, April 2005.

[4] H de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *J Comp Biol*, 9:69–105, 2001.

[5] H de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T Sari, and H. Geiselmann. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *J. Math. Biol*, 66(2):301–340, 2004.

[6] H deJong, J Geiselmann, and D Thieffry. *On Growth, Form, and Computers*, chapter Qualitative modeling and simulation of developmental regulatory networks, pages 109–134. Academic Press,, 2003.

[7] H deJong, M Page, C Hernandez, and J Geiselmann. Qualitative simulation of genetic regulatory networks: Method and application. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI-01*,, pages 67–73, 2001.

[8] J Guespin-Michel, B Polack, and A Mérieau. Bacterial adaptation and epigenesis. *Recent research and Developments in Microbiology.*, 7:289–305, 2003.

[9] L Louis, Mand Holm, L Sánchez, and M Kaufman. A theoretical model for the regulation of sex-lethal, a gene that controls sex determination and dosage compensation in drosophila melanogaster. *Genetics*, November 2003.

[10] Perko. *Différential equations and dynamical systems*. Springer, 1991.

[11] S Pérès and J-P Comet. Contribution of computational tree logic to biological regulatory networks: example from pseudomonas aeruginosa,. In *Proceedings CMSB 2003, LNCS 2602*, pages 47–56,, 2003.

[12] L Sanchez and D vanHelden, J andThieffry. Establishment of the dorso-ventral pattern during embryonic development of drosophila melanogaster: a logical analysis. *J. Theor. Biol*, 189:337–389, 1997.

[13] Snoussi. Necessary conditions for multistationarity and stable periodicity. *J. Biol. Syst*, 1998.

[14] E.H Snoussi. Qualitative dynamics of a piecewise-linear differential equations : a discrete mapping approach. *Dyn. Stability Syst.*, 4(185), 1989.

[15] R. Snoussi, E.H.and Thomas. Logical identification of all steady states: the concept of feedback loop characteristic states. *Bull. Math. Biol.*, 1993.

[16] L Sánchez and D Thieffry. A logical analysis of the drosophila gap gene system. *J. theor. Biol.*, 211:115–141, 2001.

[17] D Thieffry and D Romero. The modularity of biological regulatory network. *Biosystem*, 50:49–59, 1999.

[18] D. Thieffry, E.H. Snoussi, J. Richelle, and R. Thomas. Positive loops and differentiation. *J. Biol. Systems*, 3:457–466, 1995.

[19] Thomas. Feedback loops : The wheels of regulatory networks. In Collado-Vides, and Magasanik and Smith, editors, *Integrative Approaches to Molecular Biology*. MIT Press, Cambridge, Mass, 1996.

[20] J. Thomas, R.and Richelle. Positive feedback loops and multistationarity. *Discrete Applied Mathematics*, 19, 1988.

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

Games Network Modelling Toward a Simplified Game Design based on Abstract Interpretation Application to PAs system

C. Chettaoui[1], M. Manceny[1], M. Malo[1] and F. Delaplace[1]

[1] IBISC FRE 2873, CNRS - Evry Val d'Essonne University, France

Abstract

In this paper, we present a games network based modelling applied to Plasminogen Activation system (PAs). Games network theory extends game theory by including the locality of interactions. Each game in a games network represents a set of local interactions between some biological agents. We first introduce main features of the model, then we describe the modelling of the Plasminogen Activation system which is a signal transduction pathway involved in cancer cell migration. The PAs system is implicated in cytoskeleton modifications via actin and microtubules regulations, which in turn favours cell migration. Finally, we briefly present an extension of this work which consists on designing a simplified model (game) from an initial one. The method is based on abstract interpretation and aims at reducing the capabilities of each agent (strategies) while preserving its behaviour.

1 Game Theory and Theory of Games Networks

The theory of games networks is an extension of game theory. Section 1.1 presents the main results about game theory, and section 1.2 presents the extension.

1.1 Game Theory

Game theory allows us to model a set of interacting agents. Each agent chooses the strategy that maximizes its payoff. The payoff depends on the strategies played by other agents. The reader can refer to the books [6, 8, 9, 10] for a complete overview of game theory and its applications.

1.1.1 Strategic games

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

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

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

1.1.2 Nash Equilibrium

Nash Equilibrium is a central concept of Game Theory ([9]). This notion captures the steady states of the play for a strategic game in which each agent holds the rational expectation about the other players behavior. A *Pure Nash Equilibrium* (**pne**) corresponds to a *strategic profile* $c = (c_i)_{i \in A}$ (or vector) where c_i is the strategy "chosen" by the player i.

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

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

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

1.2 Theory of Games Networks

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

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

1.2.1 Definition

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

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

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

1.2.2 Graphical representation

A games network offers a synthetic representation to define the different interplays between several players.

A games network is represented by a bipartite graph. Graphically, agents are represented by circles, and game nodes by rectangles (See figure 1 for an illustration).

1.2.3 Global Equilibria

In biological systems, steady states are considered as the characteristic observable states of the system dynamics. In game theory, these steady states are obtained by computation of Nash equilibria. A similar notion has been developed in games networks: *Pure Games Network Equilibria* (**PGne**). Pure Games Network Equilibrium corresponds to a compatible association of local equilibria.

2 Application to the Plasminogen Activation system (PAs)

The theory of games networks has been used to model biological systems where agents correspond to genes or proteins for example. We model here a part of the Plasminogen Activation system (*PAs*), which is involved, among others, in the migration of cancer cells. *PAs* is composed of 3 molecules (*uPAR*, *uPA* and *PAI-1*). Understanding the regulation of the *uPAR/uPA/PAI-1* complex formation appears to be central to analyze the migration.

2.1 Cell migration and the PA system (PAs)

Cellular migration is a complex process which can be described like a succession of stages: *adhesion*, *contraction*, *de-adhesion* [7]. We are interested in the *PAs* system which participates to the establishment of a molecular bridge between the cell and the extra-cellular matrix. This bridge leads to the migration of the cell [1] . *PAs* system is composed of a receptor *uPAR* (Receptor of urokinase), a protease *uPA* (urokinase Plasminogen Activator), and a specific inhibitor *PAI-1* (Plasminogen Activator Inhibitor-1)[11]. The sequence of interactions implied in the promigratory process is as follows: *PAI-1* can blind to *VN* (the Vitronectin, a protein of the extra cellular matrix) which stabilizes *PAI-1* in its activation form. Once *PAI-1* is activated, it clings to a complex formed by *uPAR* and *uPA*. The complex is internalized by a receptor $\alpha 2$ *M-LRP* (Low-density lipoprotein receptor-related protein) in the cell. Then *uPAR* is recycled at the front of the cell. The signaling molecule *PAI-1* induces modifications of cell morphology including changes in cytoskeleton of actin, necessary to the migration. These modifications imply the regulation of the activation of GTP-ases of the *Rho* family [12].

2.2 Games network modelling

Figure 1 shows the games network modelling of PAs. Circles represent the different agents involved in PAs (c_i agents are intermediary complexes), rectangles are games which define the rules of interactions between agents.

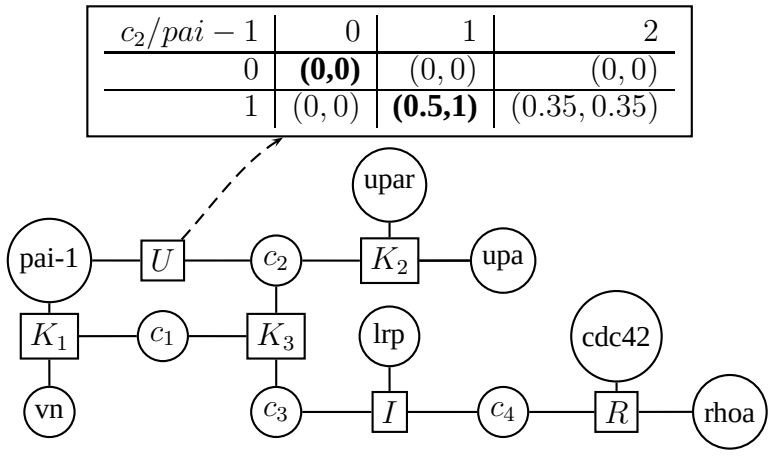


Figure 1: Games network modelling of PAs

Once we have modeled the structure of the PAs network, we have to define the payoffs of each game. Theory of games networks allows us to determine that the K_i games (which represent complexation) are not central in the determination of **PGne**. The main game is U, whose payoffs are shown in figure 1. It is to be noted that for biological reasons, not explained here, the complex

c_2 has 2 strategies, and *PAI-1* has 3 strategies ([2]). In the table, the first number represents the payoff of c_2, and the second number the payoff of *PAI-1*.

With the payoffs defined, we are able to compute **PGne**, i.e. is steady states of the *PAs*. Some tools of theory of games networks, not described here, enable us to affirm that **PGne** are exactly composed of Nash equilibria from game U. These equilibria are written in bold in figure 1.

According to experiments, both Nash equilibria correspond to characteristic biological states. The first one $(0, 0)$ corresponds to a non migratory state and the last one $(1, 1)$ corresponds to a promigratory state.

3 Game abstraction

In this section we present the evolution of games network modelling. More precisely, we present a formal approach to abstract a game based model from observations.

Given observations made on a set of agents interacting together, we build an initial game (agents, strategies and payoffs). This game has a certain number of Nash equilibrium which are assimilated to stationary states of the dynamics of interactions and which describe the behaviour of the system (or a part). We try to define automatically an abstraction preserving the richness of the behavior of the initial game. This abstraction is a game which has a reduced number of strategies and the property that its Nash equilibrium corresponds to the abstraction of Nash equilibrium of the initial game.

It is summarized by the following equation:

$$Abstract(Nash(Initial - game)) = Nash(Abstract(Initial - game)) \quad (1)$$

The equation can be represented by the following diagram:

$$\begin{array}{ccc} Initial - game & \overset{Abstraction}{\longrightarrow} & Abstract - game \\ \downarrow \text{Nash} & & \downarrow \text{Nash} \\ Nash - equilibrium & \overset{Abstraction}{\longrightarrow} & Nash - equilibrium \end{array}$$

We establish a sufficient condition to obtain the equation 1 (These conditions are not explained here due to the lack of space but we plan to do it in details during the talk).

This approach can be expressed, in abstract interpretation [4, 5], as the search for a *relevant abstraction compared to observed dynamics*. For that we introduce the concept of *Galois connection* which defines two monotonous functions $\langle \alpha, \gamma \rangle$. The first is related to abstraction; it allows to abstract the initial game. The second is related to concretization; it allows to have an approximation of the initial game. We currently work on an algorithm to automate the building of the abstract game given an initial one.

We show within the framework of the study of the *PAs* system presented previously, that the game *PAI-1* is an abstraction of the initial game.

4 Conclusion and perspectives

In this paper we have presented the extension of our modelling using games network. Theory of games networks has been used to model a part of the Plasminogen Activation system. The modelling confirms the existence of two characteristic states which correspond to physiological configurations (i.e. non and pro migratory states). We have also defined the notion of abstraction of a game. This notion allows a better comprehension of biological functions. In fact, the abstract game is a simplification of the initial one. So it is easier to extract the different interactions between

agents and to understand their interplays. The perspectives of this work is to find the conditions that the algorithm must satisfy in the aim to obtain a *valid abstraction*.

References

[1] B. Chazaud, R. Ricoux, C. Christov, A. Plonquet, RK. Gherardi, and G. Barlovatz. Promigratory Effect of Plasminogen Activator Inhibitor-1 on Invasive Breast Cancer Cell Populations. *American Journal of Pathology*, 160(1):237–246, January 2002.

[2] F. Delaplace, C. Chettaoui, M. Manceny, and M. Malo. Games Networks & Application to PAs System. In *IPCAT*, 2005.

[3] F. Delaplace and M. Manceny. Games network. Technical Report 101-2004, Laboratoire des Méthodes Informatiques (LaMI), CNRS-UMR 8042, Université d'évry-Val d'Essonne, 2004.

[4] P. Cousot Interprétation abstraite. Technique et Science Informatique, Vol. 19, Nb 1-2-3. Hermès, pp. 155-164. Janvier 2000.

[5] P. Cousot Abstract Interpretation. In ACM Computing Surveys, 28(2):324 328, June 1996.

[6] R. Gibbons. *Game Theory for Applied Economists*. Princeton University Press, 1992.

[7] DA. Lauffenburger and AF Horwitz. Cell migration : a physically integratedmolecular process. *Cell*, 84:359–369, 1996.

[8] R. B. Myerson. *Game Theory : Analysis of Conflict*. Harvard University Press, 1991.

[9] J. F. Nash. *Essays on Game Theory*. Edward Elgar Publishing Limited, 1996.

[10] M. J. Osborne. *An introduction to game theory*. Oxford University Press, 2003.

[11] KM. Providence and PJ. Higgins. PAI-1 Expression Is Required For Epithelial Cell Migration In Two Distinct Phases Of In Vitro Wound Repair. *Journal of cellular physiology*, pages 200:297–308, 2004.

[12] AJ. Ridley. Rho GTPases and cell migration. *Journal of cell science*, 114:2713–2722, August 2001.

Encouraging change. The role of foundations in European research

Eric Fourmentin[1], Damien Larivière[1]

[1]Fourmentin-Guilbert Scientific Foundation
2 Av du Pavé Neuf, F 93160, Noisy le Grand, France (fondation@fourmentinguilbert.org),
website: http://www.fourmentinguilbert.org

Abstract

By their nature and approach, foundations play an important role to increase the impact of research. This short paper aims to explain how the research community and the society at large can benefit from foundations.

What is a foundation?

Foundations are private initiatives, either from an individual or from legal entities, to support causes in line with the goals of the foundation. There are characterized by:

- A public interest goal. The purpose of a foundation must be broad enough to be useful to the society at large.
- A financial endowment. Through asset management, the initial donation provides enough revenues to finance its actions.
- Independence of its governance bodies. A nominated board and a scientific counsel are ultimately responsible for ensuring adherence to the original mission statement.
- Long term view beyond its funders.

In many European countries, tax incentives have been designed to foster individual and companies giving money to foundations. Through a matching donation scheme, half the amount is coming from the donor, and half from tax exemptions. The end result for research is twice the amount which would have been collected by taxation only. Therefore, transparency and accountability of foundations are crucial given the trust they received from the donors and the states.

The European foundation landscape

European foundations are a very heterogeneous pool of institutions whose defining characteristics often depend on local factors and the regulatory environment. Some countries like Germany, United Kingdom and Nordic countries house several foundations with significant assets (more than a billion of Euros). In comparison to the US, foundations in Europe have played a less prominent role until now.

In recent years the importance of foundations has significantly grown. According to the latest comparative statistics in Italy and Germany, around 50 percent of registered foundations

have emerged since 1990, while other countries such as Belgium, Finland, France and Sweden report between 19 and 29 per cent increases in the number of foundations.

What foundations are doing for research?

Besides grant attribution, foundations are acting for research. They are :

- Asking difficult questions. Several new research field started from the impulsion of a foundation. In France, the IPSEN foundation initiated research on Alzheimer disease.
- Leverage. Through partnering, foundations can increase the available budget.
- Take a long term view. For more than 120 years, the Institut Pasteur is contributing to the prevention and treatment of disease.
- Apolitical advocates.

A good illustration of these principles is the involvement of the Welcome Trust. One of its major achievements has been the support of the Human Genome project and its role in ensuring free global access to sequence data. Currently, foundations are also advocating for open access making it a requirement for grants.

Current challenges for researchers

The society is more and more asking researchers to go beyond their traditional role. It is no longer sufficient to provide scientific facts and results and to do knowledge transfer. Scientists are now asked to actively participate in interactive communications with the public to foster its involvement in science and technology.

Conclusion: The role of foundations in facilitating change

Given the billions of Euros spent by public authorities and enterprises it is indeed not the overall amount of money spent, but rather the approach taken by foundations that makes the difference.

Research foundations are helping research efforts in a variety of ways:

- Stimulating private means and initiatives for the benefit of the public at large.
- Identifying relevant topics or infrastructural demands for priority-setting.
- Assisting in implementing topical or structural innovation framework.
- Contributing to the creation of a research-friendly society.

Private foundations have several unique advantages to promote change:

They can act much more freely and quickly, putting objectives on top of rules and regulations, without having to wait for political consensus.

They can act autonomously in supporting the first experiments in new areas and in taking risks.

Foundations have the flexibility to quickly respond to the needs of the research community, to pilot projects, and to trigger spending on research by bigger funders.

PART III POSTERS

Learning Genetic regulatory networks with an evolutionary algorithm

Cédric Auliac[1] & Vincent Frouin[1] and Florence d'Alché–Buc[2,3]

[1] CEA Service de Génomique Fonctionnelle, 2 rue G. Crémieux, 91057 Evry
[2]IBISC 2873 FRE CNRS - Université d'Evry
[3] Programme Epigénomique,Genopole, 523 pl. des terrasses de l'Agora, 91057 Evry

Abstract

Context: Probabilistic graphical models offers an attractive framework for the representation and the learning of biological complex systems. As generative approaches, they allow to model the joint probability law of variables at hand. As graphical models, they represent observed and hidden variables as nodes of a graph whose structure governs the conditional dependencies between variables. In the case of gene regulatory networks, different Bayesian networks (BN) have been proposed from static models to linear (or non parametric) dynamical Bayesian networks. The identification or the completion of these models from transcriptomic data and prior knowledge raise several questions: the model complexity, the minimal number of data needed to find a suitable model and the issue of structure learning. In this work, we focus on learning structure in models for which the structure is not encapsulated in parameters but has to be determined explicitly. As the problem of structure learning is known to be NP-hard, the approaches such as evolutionary algorithms that favor a large exploration of the solutions space and avoid to be stacked in local minima seem especially adapted. We propose here to enhance an optimization method based on an evolutionary algorithm and focus on static Bayesian networks, essentially for data availability.

Methods: BN structure manipulation (which has to be a directed acyclic graph (DAG)) is a very difficult task due to the acyclicity constraint on this class of models [1] . Instead of evolving binary strings, we apply dedicated recombination and mutation operators that deal with the building blocks of the DAGs : edges. They rely on an efficient method that finds and repairs cycles appearing in a child during the reproduction step. Finally, we introduced in the selection step an adaptation of the deterministic Crowding method [2] which only involves a "low cost" similarity measure between DAGs, to avoid premature convergence.

Results: We test and validate our approaches on known artificial data. We sampled an artificial networks with 12 nodes (discrete random variables) with a parsimonious topology and generated a 500 samples data set. We performed 10 learnings per parameter configuration and study the performances of the optimization strategy depending of the reproduction methods used and on the Crowding method. As expected, the usage of the Crowding method improves the optimization process : obtained BN graphs are more accurate and reproducible. Conversely, the various reproduction methods yield similar results. Future work will include additional studies on the ability of our evolutionary approach to deal with samples of small size in order to cope with bio-realistic data. A similar approach will be also implemented on dynamic Bayesian networks.

References

[1] C. Cotta and J. Muruz´abal. Towards a more efficient evolutionary induction of bayesian networks. In *Parallel Problem Solving From Nature VII*, volume 2439 of LNCS, pages 730-739. Springer-Verlag, 2002.

[2] Samir W. Mahfoud. Niching methods for genetic algorithms, April 21 1995.

Processus de Markov à sauts et deterministes par morceaux appliqués en génétique moléculaire

O. Radulescu[1] and A. Crudu[1]

[1]IRMAR, UMR CNRS 6625, Université de Rennes1, Campus de Beaulieu, 35042 Rennes, France

Résumé

Les processus de Markov à sauts permettent la modélisation des phénomènes stochastiques en biologie moléculaire. Néanmoins, il y a peu de résultats mathématiques sur la dynamique de ces processus. Aussi, leur simulation sur ordinateur rencontre des difficultés reliées au temps d'exécution. Nous présentons des résultats permettant de réduire la complexité de la dynamique stochastique. Ces méthodes utilisent des théorèmes limites probabilistes.

1 Introduction

Plusieurs observations nous conduisent à introduire une approche stochastique dans la dynamique des réseaux de gènes en biologie moléculaire. Parmi ces observations une des plus importante est le phénomène de faible dosage: certains gènes en nombre insuffisant peuvent être exprimés par certains cellules et ne pas l'être par d'autres cellules du même tissu. C'est le phénomène de haploinsuffisance (Cook at al., 1998). Dans les mesures par des méthodes de fluorescence des niveaux d'expression du même gène dans des cellules du même type, on constate des distributions statistique des niveaux d'expression, dépendantes des conditions externes (ceci étant connu sous le nom de variabilité épigénétique). Une troisième catégorie importante est celle des phénomènes de bistabilité du phage lambda (liés au niveau du gène lytique cro).

Les trajectoires des processus de Markov à sauts convergent vers des trajectoires déterministes, solutions d' équations différentielles, lorsque le nombre de chaque type de réaction par unité de temps est important (Kurtz, 1970, Kurtz, 1971). Dans cette limite, appelée limite fluide par analogie avec le même concept en recherche opérationnelle, les phénomènes stochastiques disparaissent. Il s'agit de la loi des grands nombres pour les processus de Markov. Dans les mêmes conditions, il existe aussi un théorème central limite qui permet l'approximation des processus de Markov par des diffusions (Kurtz, 1971, Allain, 1976, Gillespie, 2000).

Des modèles déterministes par morceaux, même très simples, couvrent une large gamme des comportements. Nous avons classé ces comportements selon la loi invariante du processus et selon le temps de retour.

La présentation sera concentrée sur trois exemples:
(A) bistabilité de phage lambda (modèle de Hasty);
(B) le phénomène de haploinsuffisance (modèle de Cook)
(C) le transport du bruit dans les modules de signalisation (cas du NfkB).

Measles Virus infection: a mathematical approach

J Druelle[1], E Grenier[2] & TF Wild[1]

[1] "Immunité et Vaccination" INSERM U404 IFR128 Biosciences Lyon Gerland 69007 Lyon, France

[2] "Unité de Mathématiques Pures et Appliquées" École normale supérieure de Lyon, CNRS UMR 5669 69007 Lyon, France

Six structural proteins (N, P, M, H, F, L), one RNA, a few lipids compose Measles Virus (MV). From this simple recipe, appears a complex replication mechanism far from being completely understood. Understanding the basis of viral infection and predicting mathematically the progression of viral cycle presents a great scientific challenge. In this study we want to analyze *in silico* MV replication. We have therefore established a partial differential equations system following step by step what we know presently about MV infection.

The replication cycle of a single stranded negative non segmented RNA virus is composed of several steps more or less well characterized.

The viral particle (***pfu***) fixes to one of the 10^5 ***receptors*** at the cell surface with a certain "probability of attachment". The virus "entry" depends on the cell. After the fusion, one ***genomic RNA (gRNA)*** is introduced. It is associated to 3000 ***N protein***. ***Polymerases*** composed of ***P*** and ***L proteins*** are also released into the cell. Those ***polymerases*** begin the "transcription" of the ***gRNA*** in different ***messenger RNA (mRNA)*** coding for each of the 8 ***proteins*** (*N, P, C, V, M, H, F, L*). It is admitted that when the ***N protein*** concentration exceeds a "certain amount", the ***polymerase*** shifts its activity in "replication". It produces ***anti genomic RNA (agRNA)*** equally associated to the ***N protein***. This ***agRNA*** is also "replicated" into ***gRNA***. All ***the metabolic resources, the nucleotides and the amino acids*** are provided by the cell that becomes exhausted as the infection goes. When the concentration in ***M protein*** exceeds a "certain amount", "encapsidation" (*i.e.* the assembly of ***gRNA*** and the ***proteins***) occurs. The ***free virus (= pfu)*** that has been produced will be able to infect other cells if they still have a "sufficient" number of ***receptors***. Actually, the contact between the ***H protein*** and the ***receptor*** induces the "downregulation" of the ***receptor***. But, if the cell expresses a "certain amount" of ***H protein*** at its surface and if the adjacent cells still express a "sufficient" number of ***receptors***, they will fuse and form a syncytium.

At the end, we have also to take into account that every molecular species degrade themselves at a "certain rate". Moreover, even if they are ill characterized, it is admitted that in a viral suspension, there are ***non infectious defective particles*** that enter in competition with the ***efficient virus (= pfu)***. It is supposed that its "shorter genome" does not allow it to produce every necessary ***mRNA*** but permit it to be "replicated faster".

In the context of this actual understanding of MV infection, our model follows the evolution of ***every molecular species*** time after time. Some parameters are well known whereas many are still "unknown", "supposed" or "evaluated". Numerical simulation of our system of PDE gives encouraging results. However, we lack sure data about numerous parameters. This project will allow us to improve the importance of parameters like the existence of several receptors for the virus depending on the strain or the precise role of V and C non structural proteins.

Computer simulation of a Darwinian cell differentiation model

Jérôme Glisse[1,2]

[1] Université Pierre et Marie Curie-Paris 6, CNRS/IN2P3, UMR 7585, Laboratoire de Physique Nucléaire et des Hautes Énergie, Paris, F-75005
[2] École Normale Supérieure, Centre Cavaillès, INSERM, 29 rue d'Ulm, F-75005 Paris

Abstract

The stochastic nature of gene expression and the stochastic variability in cell differentiation challenge the deterministic view of the genetic program. Other models have been proposed, such as a Darwinian theory of cellular differentiation [1]. A description of computer simulations of this model is given. Kinetic analysis of the simulations exhibits the advantage of restricted stochasticity over deterministic version of this model. It also shows that cell apoptosis improve the performance of the model. Last part deals with the simulation of gene expression.

Darwinian theory of cellular differentiation

Computer based simulations

The computer model is described in detail in [2]. The cell population consists of two cell types A and B, each corresponding to the activity of one gene either a or b. These two genes code a or b molecules. They diffuse, subsequently each cell is situated in a micro-environment characterized by their concentrations. The probability P of a cell being either A or B is a decreasing function F of the concentration in its environment of a or b molecules. The simulation of this model generates a bi-layer cell structure of finite growth.

Kinetic analysis

The time step at which the bi-layered cell structure appears are recorded for 1000 simulation runs. Different values of the parameter governing the stochasticity of the model are used to produce either an extensively stochastic, a restricted stochastic or a deterministic version of the model. The results show that the restricted stochastic version of the model improves the reproducibility of the bi-layer formation.

Cell apoptosis

We have suppressed cell death from the model. The simulations show that the cell bilayer formation is impaired. Thus cell death improves tissue organization in the context of the Darwinian model. This provides a strong explanation for its evolutionary origin.

Gene expression and chromatin simulation

This model is closely linked to the expression of gene. Expression of gene is a stochastic process. We are currently designing such simulation using a multi-agent system approach, with agent such as bio-molecule or chromatin.

References

[1] Kupiec JJ. A Darwinian theory for the origin of cellular differentiation. *Mol Gen Gene*, 255:201-208, 1997

[2] Laforge B., Guez D., Martinez M., Kupiec JJ. Modeling embryogenesis and cancer[...]. *Progress Biophys Mol Biol*, 89:93-120, 2005

3D simulation of biological membrane: mitochondrial application

C. Lales[1]

[1] LaBRI CNRS UMR5800, Université Bordeaux 1, 350 avenue de la libération, F-33405 Bordeaux

Abstract

"In silico" simulations of biological processes must take into account several kinds of complex molecular behaviors. Simulation of membranes, due to interactions between phospholipid bilayers and enzymes embedded in them, are typically an example of such complexity. A way of modeling theses membranes and the enzymes embedded is to use ordinary differential equations. Such models describe the global evolution of the system with the use of averaged variables but many biological aspects are still difficult to describe: too few molecules reacting (no average behavior), influence of the conformation of the molecules, etc.. On the contrary, multi-agent models are powerful tools to decompose complex interactions between agents (molecules) in simple local behavior animating each agent. Interactions between agents (phospholipids) allow to observe the emergence of membranes with various properties: self-assembly, self-repair, semi-permeability and insertions of membrane proteins. Such models recently published (L. Edwards Artificial Life, (4), 1998) have already shown emergence of self-assembly in 2D space. In this work we have focused our work on a 3D space simulation of the mitochondrial membrane.

In a multi-agents system, the agents are considered to be autonomous entities. Their behaviors are the consequence of their local perception, their knowledge, their goals and the interactions with the neighbors in an environment. Some models of membrane have been already developed with multi-agents formalism, from the more realistic one (molecular dynamics) to the more abstract one (graph of interconnected nodes).

The model we propose can take into account several types of reactants agents (phospholipids, enzymes, metabolites,...). The time is discretized with a constant time step and a scheduler is defined for each type of agent. The 3D space is continuous but a discretized grid optimizes the neighbors research. The granularity of a model is given by the precision of the abstraction done by the agents (molecules). On the one hand an abstraction of a molecule by a single point is not enough to take into account a dynamic 3D structure and its spatial orientation. On other hand molecular dynamics at the atom level is not possible for a whole organelle due to the large number of entities to compute and the huge resource consumption increase. Thus, we propose to reduce the granularity at the atom set level, abstracting each agent (molecule) to its gravity center and a unrestricted list of interacting points moving around it. An interacting point is a portion of molecule, an atom set, that exhibits common properties and could be affected by external forces: hydrophobia, electrostatic charges, etc. Interactions between agents (molecules) are reduced to a set of forces, linear and rotation movements resulting from an physic engine. This design keeps information about the 3D structure of the molecule and its space orientation but also its internal dynamics. It gives a compromise between the "unreachable" complexity of molecular modelling and the abusive abstraction of molecule by points (spheres).

In this context we have studied the specific case of the respiratory chain, a pool of five enzymatic complexes embedded into the inner mitochondrial membrane. We would be able to understand the emergence of inner membrane complex structures and their impacts on enzymatic chain reactions. We think that the multi-agent paradigm, thanks to the features of our model, allows to study the link between the cristae of the inner membrane and the raft of the enzymes of the respiratory chain.

Modelling Q cycle using Multi-Agent System

N. Parisey[1], M. Beurton-Aimar[2] and JP. Mazat[1]

[1] Physiopathologie mitochondriale U688, Université Bordeaux 2, 146 rue Leo-Saignat, F-33076 Bordeaux
[2] LaBRI CNRS UMR 6800, Université Bordeaux 1, 350 avenue de la libération, F-33405 Bordeaux

Abstract

Metabolic processes involve numerous enzymatic reaction pathways and constitute complex interaction networks. Most often each enzymatic reaction implies at least two metabolites (molecules) and the level of interactions depends upon each participant current state. Even if participant exhibit a small number of different possible states, this can lead to a combinatorial explosion of the number of reachable states of the whole system. Using a multi-agent design it is possible to reduce this complexity. In such design, agents are enzymes or metabolites and each of them stores its own state. The complexity is then recovered during the computer simulation of the process. We have studied a specific case of metabolic process, called the Q cycle, which involves one enzyme and two types of metabolites. The Q cycle is the central part of the respiratory chain in mitochondria which mainly produces cell energy. Each metabolite can have two or three possible oxydo-reduction states depending on their type. The enzyme has six reactive sites and each site has two possible states : free or occupied by a metabolite. Taking into account all the possible states of the system leads to model several hundred system states. A traditional way in biology is to code the system by a set of differential equations. We show that a multi-agent system can model this kind of process more easily than the traditional one. The obtained tool allows to manage and to modify conditions of simulations for testing different hypotheses on normal or pathological situations in an easy way.

Metabolic pathways classification

Sabine Pérès[1], Marie Beurton-Aimar[2] and Jean-Pierre Mazat[1]

[1] Physiopathologie mitochondriale U688, Université Bordeaux 2, 146 rue Leo-Saignat, F-33076 Bordeaux
[2] LaBRI CNRS UMR 6800, Université Bordeaux 1, 350 avenue de la libération, F-33405 Bordeaux

Abstract

Metabolic pathway analysis is essential to study metabolic network behaviour. Theorical approach as elementary flux mode enables to study the network properties. Their determination lead to combinatorial explosion of their number when the network is complex. We have applied this formalism to three metabolic networks : mitochondrial energetic metabolism of muscle, liver and yeast. We have elaborate classification method of elementary modes to analyze the obtained re- sults. This method based on an agglomeration of commun patterns allows us to interpret sets of elementary modes, to find their biological meaning and to express links between reactions.

Artificial Ant-Based Systems for the Resolution of Optimization Problems

Yoann Pigné[1]

[1] Laboratoire d'Informatique Traitement de l'Information et Systèmes, Université du Havre, France

Problem

The class of problems we address belong for the most to the field of optimization problems. These problems are modeled with graphs and bound to uncertainty. This uncertainty can be related to :

- information about this problems that are imprecise or unreliable,
- dynamics in the structure of the problem that evolve qualitatively/structurally in time.

We focus here on dynamic environments.

The evaluation of a solution in this kind of problem seems difficult since the problem is constantly changing and a solution generated at time t can become irrelevant at time $t + 1$ since the problem has changed. This is typically the case for some classical optimization methods that construct solutions iteratively. When the solution arrives it is already outdated because it refers to an environment that do not exist anymore. That is why the definition of a relevant evaluation function seems problematic.

Approach

Rather than focusing on the formulation of such global function, our interest goes to the nature of a solution. In a graph, even bounded by dynamic, for some problems, solutions can be qualified as parts of the original graph. These parts are identified as structures that can be subsets of vertices and/or edges, subgraphs, it can be paths or sets of paths. The shape of a structure mainly depends on the way the problem is modeled. In this way any structure/solution observed in the graph corresponds exactly to the actual state of the graph since it is part of it.

Ant-Based Systems

In concrete terms, the system that can construct this kind of structure needs to be able to evolve in a dynamic environment and to modify the environment so as to raise structures from it. Ant colonies have a decentralized, multiagent system based approach that allow dynamic environments and their stigmergic way of communication give them the ability to produce observable structures.

Moreover, this system presents more abilities. Firstly, indirect communication allow robustness. Indeed, robustness is said to be the resilience of the system, the ant colony is still able to run when some parts of the colony are out of order because no synchronization, no meet point is reached. Secondly, interactions between entities gives the system some flexibility. This means that changes in the environment don't drastically change the behavior of the system. The system adapts when the graph looses or gets some edges/nodes.

Related Bibliography

E. Bonabeau, M. Dorigo, G. Theraulaz. *Swarm Intelligence: from natural to artificial systems*, Oxford University Press, 1999

F. Guinand and Y. Pigné. *Problem Solving and Complex Systems*, to appear in Emergent Properties in Natural and Artificial Dynamical Systems, Springer ed., 2006

Mathematical modelling of apoptosis

N. Reymond[1] , E. Grenier[1] and J-P Boissel[2]

[1] Unité de Mathématiques Pures et Appliquées, UMR CNRS 5669, ENS Lyon, France
[2] Service de Pharmacologie Clinique, EA 643, Faculté de médecine Laënnec, Université Claude Bernard, Lyon, France

Context

The intracellular environment is the place of a large variety of enzymatic reactions, transport and diffusion of molecules. These networks of thousands of reactions are differently activated or repressed, according to the type and the state of the cell (Atlan, 2002). Modelling appears to be a relevant approach to gather all the parts of this puzzle, not only to predict the behaviours of the system, but also to structure and organize the knowledge. The development of such models implies a multi-field and collective approach, associating biological knowledge, mathematical formalism, and simulations.

In this recent context, the aim of this work is to develop a global model of apoptosis, integrating the presently distributed and heterogeneous knowledge, about this signalling pathway. Programmed cell death or apoptosis is a highly conserved pathway, that is functional in all higher organisms. This mechanism eliminates defect cells without damaging neighbouring tissue cells and is therefore recommended for tissue maintenance (Reed, 2002). However, dysregulation of apoptotic signalling can play a role in various diseases, with insufficient apoptosis leading to cancer (cell acumulation), whereas excessive apoptosis contributes to ischaemia (stroke).

Most of the molecular mechanisms of the apoptosis have been elucidated during the last years. However, in spite of an always increasing number of works on the subject, there are not any experimental approach to know all the modifications of the molecular concentrations, occurring in the apoptotic process. Then, mathematical modelling and simulations become essential tools to understand and study the global behaviour of this signalling pathway (Bentele et al., 2004).

1 Modelling

The first stage of work was to develop a discursive model of the apoptose, which integrates the knowledge that is actually available. This model is a rigorous framework to synthesize information of various origins (Fussenegger et al., 2000).

The second stage was to implement the mathematical model. The global discursive model of the apoptosis had been divided into two sub-models. The first one describes the extrinsic pathway (with an extracellular signal on the death receptors). The second one describes the mitochondrion pathways (intracellular signal by DNA damage, treatment with cytotoxic drugs or irradiation). In both cases, the cascade of the kinetic reactions involves a large number of enzymes. For this reason, the mathematical formalism chosen is based on a system of ordinary differential equations. The two sub-models are implemented with the software *Matlab*. The first runs of the simulation are based on data found in the literature. These simulations have been used to determine the missing parameters of the model and to test its robustness.

2 Simulation

Currently, the simulations are used to find the most sensitive signalling molecules and to predict the systemic behaviour of apoptotic signalling, with different molecular concentrations or with interaction of chemotherapeutics. Our main objective is to use this model to generate experimentally testable hypothesis, regarding new therapeutic approaches. At this stage, the project will rely on the expertise of teams of biologists and clinicians, in order to integrate data provided by *in vivo* and *in vitro* studies. The new experiments, raised from these simulations, will provide valuable feedbacks to improve the model.

Finally, the ultimate objective of this project is to integrate the cellular model of death, in models developed to study cancer (ModCan project, Ribba et al., 2005) and stroke (AVCisi project, Dronne et al., 2004). The aim is to obtain a global model of these complex pathologies, in order to carry out simulations to guide therapeutic innovations.

References

- Kitano, H. Computational systems biology. *Nature*. 2002, vol. 420, pp. 206-210.
- Atlan, H. La cellule vivante : un paradigme des systèmes naturels complexes. *Médecine/Sciences*. 2002, vol. 18, pp. 764-766.
- Reed, J. C. Apoptosis-based therapies. *Nat Rev Drug Discov*. 2002, vol. 1, n°2, pp. 111-121.
- Bentele, M., Lavrick, I., Ulrich, M., Stosser, S., Heermann, D. W., Kalthoff, H., Krammer, P. H. et Eils, R. Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis. *J Cell Biol*. 2004, vol. 166, n°6, pp. 839-851.
- Fussengger, M., Bailey, J. E. et Varner, J. A mathematical model of caspase function in apoptosis. *Nat Biotechnol*. 2000, vol. 18, n°7, pp. 768-774.
- Ribba, B., Marron, K., Agur, Z., Alarcon, T. et Maini, P. A mathematical model of Doxorubicin treatment efficacy for non-Hodgkin's. *Bull Math Biol*. 2005, vol. 67, n°1, pp. 79-99.
- Dronne, M., Boissel, J., Grenier, E., Gilquin, H., Cucherat, M., Hommel, M., Barbier, E. et Bricca, G. Mathematical modelling of an ischemic stroke: an integrative approach. *Acta Biotheor*. 2004, vol. 52, n°4, pp. 255-272.

A non linear dynamical bayesian network for modelling and identifying gene regulatory networks from experimental data

Minh Quach[1] & Pierre Geurts[1,2]& Florence d'Alché–Buc[1]

[1]IBISC FRE CNRS 2873 & Epigenomics project, Genopole, 523, place des terrasses 91 Evry, France
[2]Department of Electrical Engineering and Computer Science, University of Liège, Sart Tilman B28, 4000 Liège, Belgium

Abstract

Understanding and identifying biological complex systems at work in the cell requires to develop models able to capture the stochastic nature of biological processes as well as their dynamics. Focusing on genetic regulatory networks, we propose a new quantitative model in the form of a bayesian dynamical network that allows to represent both genes and proteins in the same network. We started from the non-linear differential equations of Michaelis-Menten which are the gold-standard to represent biochemical interactions and develop a discrete-time and probabilistic model from these equations. Our work can be seen as an generalization of the model proposed recently by Nachman et al [1] in which the regulation function of transcription factors on their target genes takes the non-linear Michaelis-Menten form. In our approach, we introduce a higher detailed model where the dependency between the regulatory proteins and the genes that code for them as well as post-transcriptional events (e.g., protein-protein interactions, protein degradations) can be taken into account. In the resulting continuous nonlinear dynamical system, the proteins are considered as hidden variables while genes are observed variables through transcriptomic measurements. In order to learn the parameters of this new dynamical system, we have developed an approach based on the Unscented Kalman Filter which is able to estimate simultaneously the kinetic parameters and unobserved protein activities given the structure of the network. The generality of the learning method opens the door to various adaptations of the model if required by the biology. However this Unscented Kalman Filter like algorithm does not solve the structure learning problem whose nature remains NP-complete. While forging new strategies for this problem, we present in this work a classical greedy-like approach for the search in the structure space.

Numerical results on parameter estimation in several small networks with synthetic data are presented and show the relevance of the model.

References

[1] I. Nachman, A. Regev, and N. Friedman. Inferring quantitative models of regulatory networks from expression data. *Bioinformatics*, Vol. 20 Suppl. 1:i248–i256, 2004.

Statistical Analysis of Couples of Networks Applied to Biological Systems

Serge Smidtas[1]

[1]PhD Student, Université Evry. sergi@sergi5.com

Abstract

Network analysis and modeling consist of study of individuals that are linked together into complex networks. Networks refer to artificial and natural systems like random networks and biological networks. They constitute a very active area of research in a variety of scientific disciplines, including Physics, Biology and Social sciences. Graph theory and techniques recently developed for the analysis of networks, provide a substantial background for studying complex networks structures and dynamics. One of the key features of natural networks is their ability to adapt to their environment. Graphs are not isolated. In this paper the environment is modeled by an other network that interacts with the first one. Such adaptive capacity can be found in a whole range of natural networks like gene-protein interaction networks within individual cells involving both genetic regulation and protein-protein interactions. In order to understand the structure and dynamic network characteristics that underlie interactivity of the network, one need to understand how couples of graphs interact together and correlate.

We studied the Saccharomyces cerevisiae data set. The complete network contains 4487 proteins, 9971 protein-protein interactions or complexes (PPI) and 7455 transcriptional regulations interactions (TRI).

Analysis of this heterogeneous network showed that protein interactions are organized in strong layers upstream of the genetic regulatory. This topological structure also strongly correlated with localization data. This allows to annotate proteins as Co-Regulators and Co-coregulators and to understand how crosstalk occurs and are hardware coded upstream of regulation.

Mathematical model on the dynamics of cyclical organization of simple protein regulatory networks and its application to the budding yeast cell cycle

K. Sriram[1], Gilles Bernot[1,2] and François Képès[1]

[1] Epigenomics Project, Genopole® , Evry F-91000
[2] LaMI Laboratoire de Méthodes Informatiques, UMR 8042 CNRS/Université d'Evry, Boulevard François Mitterand, F-91026 Evry cedex

Abstract

In this paper, the dynamics of the cyclical organization of simple protein networks and its application to the budding yeast cell cycle are studied by constructing nonlinear models. The protein network consists of two small cyclical loops, where each loop in the absence of interaction with the other exhibit different dynamical behavior. Bistability is exhibited by one loop in which the proteins are positively regulated by the preceding one and in turn regulates positively the subsequent one in a cyclic clockwise fashion. Limit cycle oscillations are exhibited by the second loop in which the proteins are negatively regulated by preceding one and in turn negatively regulates the subsequent one in a cyclic anticlockwise fashion. Coupling of both the cyclical loops by positive feedback loop displays complex behavior such as multi-stability and coexistence of limit cycle and multiple steady states. The coupling of two cyclical loops by the positive feedback loop brings in the notion of checkpoint in the model. The model also exhibits dominoe like behavior, where limit cycle oscillations takes place in a stepwise fashion. As an application, the events that govern the cell cycle of budding yeast is considered for the present study. In budding yeast, the feedback interactions among the important transcription factors, cyclins and its inhibitors in G1, S-G2 and M phases are considered for the construction of the biological circuit diagram.

Surprisingly, the sequential activation of the transcription factors, cyclins and its inhibitors forms two independent cyclical loops, with transcription factors involves in the cyclic positive regulation in clockwise direction, while the cyclins and its inhibitors involves in the negative regulation in anticlockwise direction. The coupling of the transcription factors and the cyclin and its inhibitors by positive feedback loops generates rich bifurcation diagram that can be related to the different events in the G1, S-G2 and M phases in terms of dynamical system theory. The different checkpoints in the cell cycle is accounted for by appropriately silencing the positive feedback loops that couples the transcription factors and the cyclin and its inhibitors.

A general approach based on automated inference for regulatory pathways retrieval in the case of pertubed systems: application to the response of the yeast to irradiation

Nizar Touleimat[1,2,3], Marie Dutreix[3] and Florence d'Alché–Buc[1,2]

[1]IBISC 2873 FRE CNRS - Université d'Evry

[2]Programme Epigénomique,Génopole, 523 pl. des terrasses de l'Agora, 91057 Evry

[3]Recombinaison et Instabilité des Génomes, Institut Curie - CNRS UMR2027, Centre Universitaire - Bâtiment 110-112, 91405 Orsay cedex - FRANCE

Abstract

In this work, we propose a general methodology based on automated deduction and statistical inference that helps the biologist to mine large datasets, extract and filter information, confront sources of data and finally gather various pieces of evidence for raising hypotheses about the regulatory processes. Devoted to the analysis of perturbed data, this approach is directly inspired from the work of the biologist who defines a set of experiments to analyse logically the results using his/her background knowledge as well as logical reasoning. We used this approach to identify new transcriptional regulatory pathways in the case of the response of the yeast (*Saccharomyces cerevisiae*) to irradiations. The variables of interest (genes of the yeast) are described by seven-time points series of expression taken during the period time of cell response to irradiation, and measured for yeast strains differing by their genomic status (gene knock-out, ploidy and mating type variations). While our methodology is driven by these gene expression data, it makes use of other sources of data that provide descriptions of the same genes but according complementary criteria: functional annotations, existence of known or potential regulators, position on chromosomes. Four steps are required: **1)** we identify groups of genes that remain co-expressed across all the tested strains. This search is performed with spectral clustering, a recent kernel-based clustering method, using on a linear combination of similarities measured for each considered yeast strain. Clustering provides a codebook of clusters representants (average expression profiles). **2)** we use these mean expression profiles to determine clusters of interest. A cluster is considered to be of interest if its average kinetics profile reveal expression variations after irradiation and present a singular behaviour for one or more of the genetic alterations. **3)** we consider each cluster of interest and test how much the set of genes in a cluster of interest is homogeneous according the secondary sources of data. At this stage, we are able on one hand, to identify which of the co-expressed genes participate to same physiological function and on the other hand, to detect if the co-expressed genes share some biological features that could reveal and explain their co-regulation. **4)** the last step is a deduction process which can be formalized using first order logic: we formalize biological knowledge relevant for the studied genes into (logical) theorems and use them in combination of informations induced at steps 2 and 3. Using logical deduction process such as modus ponens, assumptions about the nature of regulatory mechanisms are produced. By means of this four step strategy we characterized a set of gamma-rays yeast responses. First, we confirmed the relevance of our methodology by retrieving a response of a large set of genes that has previously been identified in [1] and second, we provided evidence for the hypothesis that this set of genes could be in fact divided into two clusters with two distinct transcriptional regulatory mechanisms (one of them involving a potential dependency on chromatin structure modification). Targetted laboratory experiments are now programmed to validate this hypothesis.

References

[1] N. Berthault, G. Mercier, N. Touleimat, F. Kepes, G. Fourel, E. Gilson, and M. Dutreix. A haploïd-specific transcriptional response to irradiation in *S. cerevisiae*. Nucleic Acids Research, 33(20):6635-6643, 2005.

Computation of bases of elementary flux modes and of a lower bound for their number

A. Zemirline[1], S. Pérès[2], M. Beurton-Aimar[2,3], J.-P. Mazat[2] & P. Ballet[1]

[1] Lysic, EA3883, U.B.O., 20 avenue Le Gorgeu C.S. 93837 BP 809, 29238 Brest Cedex 3, France.
[2]Lab. Physiologie Mitochondriale, univ. Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France.
[3]LaBRI, univ. Bordeaux 1, 351 cours de la Libération 33405 Talence Cedex, France.0

Abstract

The concepts of stoichiometric matrix, elementary flux modes, extreme pathways and metabolic pathways provide insights into the analysis of metabolic networks for understanding the relationships between structure, function and regulation of the metabolic systems reconstructed from genomic data. In recent years, several studies were dedicated to these concepts which yielded a rigourous formalism to describe and assess the metabolic processes. Specially, various algorithms where devoted to the computation of elementary flux modes. The most prominent algorithms are the first one developed by S. Schuster, D. Fell. and T. Dandekar, then the one developed by C. Wagner and the more recent one elaborated by J. Gagneur, S. Klamt in their unifying framework and new binary approach where it is shown that the both former algorithms can be embedded in a more general algorithmic framework stemming from computational geometry.

In this work, we describe a new algorithm based on the duality between two concepts stemming from network flows theory. This algorithm elaborates a basis of elementary flux modes that is also a basis of extreme pathways. Moreover, we also present how one can get a lower bound for the number of bases of elementary flux modes in metabolic networks.

LIST OF ATTENDEES

AIGUIER	Marc	aiguier@lami.univ-evry.fr
AMAR	Patrick	pa@lri.fr
ANDREI	Oana	Oana.Andrei@loria.fr
AULIAC	Cédric	cedric.auliac@cea.fr
BALLET	Pascal	pascal.ballet@univ-brest.fr
BARLOVATZ-MEIMON	Georgia	gbm@lami.univ-evry.fr
BASSANO	Vincent	vincent.bassano@univ-paris3.fr
BEN JACOB	Eshel	eshel@tamar.tau.ac.il
BERNOT	Gilles	bernot@epigenomique.org
BESLON	Guillaume	guillaume.beslon@gmail.com
BEURTON-AIMAR	Marie	aimar@labri.fr
BORNBERG-BAUER	Erich	ebb@uni-muenster.de
BOURHIS	Mikaël	bourhis@enib.fr
BRETON	Annick	annick.breton@ibgc.u-bordeaux2.fr
BRIERE	Christian	briere@scsv.ups-tlse.fr
CABOCHE	Ségolène	caboche.segolene@laposte.net
CAMARASA	Carole	camarasa@ensam.inra.fr
CASCANTE	Marta	martacascante@ub.edu
CHAOUIYA	Claudine	chaouiya@esil.univ-mrs.fr
CHETTAOUI	Chafika	chetta@lami.univ-evry.fr
COLOMBIÉ	Sophie	scolombi@bordeaux.inra.fr
COMMENGES	Daniel	daniel.commenges@isped.u-bordeaux2.fr
COTTRET	Ludovic	cottret@biomserv.univ-lyon1.fr
CRUDU	Alina	alina.crudu@univ-rennes1.fr
D'ALCHÉ-BUC	Florence	florence.dalche@ibisc.univ-evry.fr
DRUELLE	Johan	druelle@cervi-lyon.inserm.fr
DURRENS	Pascal	pascal.durrens@labri.fr
FAGES	François	francois.fages@inria.fr
FERRARO	Pascal	pascal.ferraro@free.fr
FRIGERIO	Jean-Marc	Frigerio@pierroton.inra.fr
GALLINA	Sophie	s.gallina@good.ibl.fr
GEURTS	Pierre	pgeurts@lami.univ-evry.fr
GIAVITTO	Jean-Louis	giavitto@ibisc.fr
GLISSE	Jérôme	glisse@lpnhep.in2p3.fr
GODIN	Christophe	godin.christophe@cirad.fr
GSCHWIND	Régine	regine.gschwind@pu-pm.univ-fcomte.fr
GUESPIN	Janine	janine.guespin@univ-rouen.fr
HEINIG	Matthias	mheinig@genoscope.cns.fr
HEINRICH	Reinhart	reinhart.heinrich@rz.hu-berlin.de
HENRI	Raphaël	raphaelhenri@wanadoo.fr
HERNANDEZ	Alfredo	Alfredo.Hernandez@univ-rennes1.fr
HUTZLER	Guillaume	hutzler@univ-evry.fr
JOURDAN	Fabien	Fabien.Jourdan@toulouse.inra.fr
KÉPÈS	François	Francois.Kepes@genopole.cnrs.fr
KLAUDEL	Hanna	hanna.klaudel@lami.univ-evry.fr
LALES	Charles	lales@labri.fr

LARIVIERE	Damien	damien.lariviere@fourmentinguilbert.org
LEFÈVRE	Thomas	thomas.lefevre@enst-bretagne.fr
LEGALL	Pascale	legall@lami.univ-evry.fr
LESNE	Annick	lesne@ihes.fr
MALO	Michel	mmalo@lami.univ-evry.fr
MANCENY	Matthieu	mmanceny@lami.univ-evry.fr
MARTIN	Pascal	Pascal.Martin@toulouse.inra.fr
MATTE-TAILLIEZ	Oriane	oriane.tailliez@cpbs.univ-montp1.fr
MAZAT	Jean-Pierre	jpm@u-bordeaux2.fr
MELANÇON	Guy	Guy.Melancon@lirmm.fr
MESTIVIER	Denis	mestivier@ijm.jussieu.fr
MOLINA	Franck	franck.molina@cpbs.univ-montp1.fr
MOLINO	François	francois.molino@igf.cnrs.fr
MORANGE	Michel	michel.morange@ens.fr
MORMEDE	Pierre	pierre.mormede@bordeaux.inra.fr
NAZARET	Christine	nazaret@sm.u-bordeaux2.fr
NORRIS	Vic	vjn@univ-rouen.fr
OUANGRAOUA	Aida	aida.ouangraoua@labri.fr
PAKLEZA	Christophe	pakleza@epigenomique.org
PARISEY	Nicolas	parisey@labri.fr
PIGNE	Yoann	yoann.pigne@univ-lehavre.fr
PINTER	Ron	pinter@cs.technion.ac.il
PÉRÈS	Sabine	sabine.peres@u-bordeaux2.fr
QUACH	Minh	mquach@lami.univ-evry.fr
QUERREC	Gabriel	gabriel.querrec@enib.fr
REDER	Christine	Christine.Reder@math.u-bordeaux1.fr
REYMOND	Nancie	nancie.reymond@umpa.ens-lyon.fr
RICHARD	Adrien	arichard@lami.univ-evry.fr
RODOLPHE	François	Francois.Rodolphe@jouy.inra.fr
SABARLY	Victor	victor.sabarly@m4x.org
SADO	Pierre-E	sado@nantes.inra.fr
SANCHOU	Laurent	laurent.sanchou@insa-toulouse.fr
SCHACHTER	Vincent	vs@genoscope.cns.fr
SINOQUET	Christine	christine.sinoquet@univ-nantes.fr
SMIDTAS	Serge	sergi@sergi5.com
SOKOL	Serguei	sokol@insa-toulouse.fr
SPICHER	Antoine	aspicher@lami.univ-evry.fr
SRIRAM	Krishnamachari	sriram@epigenomique.org
STANO	Pasquale	stano@uniroma3.it
TAHI	Fariza	tahi@lami.univ-evry.fr
TEICHMANN	Sarah	sat@mrc-lmb.cam.ac.uk
TERRASSE	Gaétan	gaetan.terrasse@wanadoo.fr
THIERRY	Alain	thierrya@univ-montp2.fr
TOULEIMAT	Nizar	touleima@lami.univ-evry.fr
TRONCALE	Sylvie	troncale@lami.univ-evry.fr
TROXLER	Laurent	L.Troxler@ibmc.u-strasbg.fr
ZEMIRLINE	Abdallah	zemirlin@univ-brest.fr
ZOLLINGER	Raphaël	raphael.zollinger@curie.net
ZRIOUAL	Saloua	saloua.zrioual@chu-lyon.fr