Proceedings of
The Lille Spring School
on

Modelling Complex Biological Systems in the Context of Genomics

April 4th - 11th, 2008

Edited by

Patrick Amar, François Képès, Vic Norris, Bernard Vandenbunder

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

Strohman RC (1977) Nature Biotech 15:199

FOREWORD

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

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

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

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

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

ACKNOWLEDGEMENTS

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

Of course the organisation team would like to express gratitude to all the staff of the Hotel *Ascotel* and the restaurant *le Pariselle* for the very good conditions we have found during the conference.

Special thanks to the Epigenomics project for their assistance in preparing this book for publication. The cover photography shows *the tulips* by Yayoi Kusama.

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
- Région Nord-Pas de Calais:
  http://www.nordpasdecalais.fr
- Centre National de la Recherche Scientifique (CNRS):
  http://www.cnrs.fr
- Réseau national des systèmes complexes (RNSC):
  http://rnsc.csregistry.org/tiki-index.php
- GDR CNRS 3003 Bioinformatique Moléculaire:
  http://www.gdr-bim.u-psud.fr
- Université des Sciences et Technologies de Lille (USTL):
  http://ustl1.univ-lille1.fr/projetUstl/
- Fondation Scientifique Fourmentin-Guilbert:
  http://www.fourmentinguilbert.org

THE EDITORS

INVITED SPEAKERS

MICHEL BORNENS	Institut Curie, Paris (F)
DENNIS BRAY	University of Cambridge (UK)
VINCENT DANOS	Université Denis Diderot, Paris (F)
MARC DUMAS	ENS, Lyon (F)
JEREMY GUNAWARDENA	Harvard University (USA)
DON INGBER	Harvard Medical School (USA)
DANIEL KAHN	Université Lyon 1 (F)
ANDREW LANE	Louisville University (USA)
JÖRG LANGOWSKI	DKFZ Heidelberg (D)
JIM MCNALLY	NIH Bethesda (USA)
HOWARD PETTY	Ann Harbor University (USA)
EYTAN RUPPIN	Tel Aviv University (Israël)
RENÉ THOMAS	ULB (Belgium)
MIN ZHAO	University of California Davis (USA)

CONTENTS

PART I INVITED TALKS

ANALYSE LOGIQUE DES CIRCUITS DE RÉTROACTION

René Thomas et Marcelle Kaufman

Université Libre de Bruxelles, Faculté des Sciences
Campus Plaine CP 231, B-1050 Bruxelles, Belgique

1. INTRODUCTION

La régulation peut être définie comme l'ensemble des processus qui ajustent le taux de production et de dégradation des éléments d'un système à l'état de ce système et à celui de variables environnementales appropriées. Les opérateurs de ces ajustements sont les circuits de rétroaction (en anglais, feedback circuits), dont le rôle biologique est absolument crucial. Pour une analyse détaillée, voir [42, 43, 46].

Classiquement, les systèmes biologiques sont traités de l'une ou l'autre de deux manières diamétralement opposées: la description purement verbale ou la description quantitative en termes d'équations différentielles. Cette dernière est extrêmement puissante, et nous l'utilisons abondamment. Cependant, en raison du caractère non-linéaire d'une large part des interactions, ces systèmes différentiels ne peuvent être résolus analytiquement. Ils peuvent être résolus numériquement avec toute la précision demandée, mais cette précision même peut être illusoire, car les valeurs des paramètres et la forme exacte des interactions doivent souvent être inventés.

On peut se demander s'il est possible de saisir les caractères qualitatifs essentiels de la dynamique d'une autre manière. C'est dans cet esprit qu'il a été proposé de manière répétée de recourir à une description logique: voir, par exemple, [7, 12, 26, 27, 35, 38, 49]. Les descriptions logiques, ou discrètes, utilisent des variables qui ne peuvent prendre qu'un nombre limité de valeurs, le plus souvent, deux seulement: 0 et 1 (dans ce cas, on parle de description binaire ou Booléenne).

A première vue, cette description, qui, dans les cas les plus simples, considère une substance comme "présente" ou "absente" et un gène comme "allumé" ou "éteint", peut apparaître comme une caricature bien grossière. En fait, on se rend compte à l'usage que cette impression est inutilement pessimiste. L'une des raisons est la suivante. En biologie et, pour autant que nous puissions juger, également dans bien d'autres domaines, un régulateur est le plus souvent inefficace en dessous d'une concentration "seuil" et son effet plafonne rapidement aux concentrations supérieures.

En d'autres termes, la plupart des interactions régulatrices ont une forme sigmoïde, caractérisée par une valeur "seuil" de leur concentration et par un plafonnement de leur effet. (Figure 1a). Cette non-linéarité est responsable de la difficulté de traiter les équations différentielles analytiquement, mais elle est en fait absolument essentielle à l'émergence de dynamiques complexes. On s'est évidemment demandé s'il était possible d'idéaliser la forme de ces interactions de manière à faciliter leur traitement. L'idée d'une idéalisation linéaire (Figure 1b) est tentante par sa simplicité. Cependant, on se rend compte rapidement qu'elle n'est acceptable qu'au voisinage immédiat d'un état stationnaire. L'idéalisation logique (Figure 1c) est en quelque sorte diamétralement opposée à la précédente en ce sens qu'elle a un caractère "infiniment non-linéaire": on raisonne comme si l'effet du régulateur était nul tant que sa concentration est inférieure au seuil, et plafonnait au-delà de cette concentration. Il est apparu que ce type d'approximation, contrairement à l'approximation linéaire, conserve pleinement le comportement qualitatif essentiel des systèmes: voir à ce sujet les beaux articles de Glass & Kauffman [8, 9], et aussi Kaufman et al. [13, 16, 17]. Notons d'ailleurs que les biologistes les plus allergiques à tout traitement formel ne manquent aucune occasion de dire, et même d'écrire: "en présence de l'immunité, ceci, en absence de l'immunité, cela", ou encore: "à basse température, ceci, à haute température, cela". Voilà pourtant une attitude bien Booléenne!

Figure 1: Une courbe sigmoïde (a), sa caricature linéaire (b) et sa caricature logique (c). Les courbes sigmoïdes peuvent être décrites analytiquement par des fonctions du type tangente hyperbolique ou par des fonctions de Hill: $F^+(x) = \frac{x^n}{s^n + x^n}$ s'il s'agit d'une sigmoïde croissante, $F^-(x) = \frac{s^n}{s^n + x^n}$ s'il s'agit d'un sigmoïde décroissante. Pour n = 1, la courbe croissante est une simple branche d'hyperbole (courbe de Michaelis-Menten). Pour toute valeur de n > 1, la courbe a un point d'inflexion et est une sigmoïde. Plus n est élevé, plus la sigmoïde est raide. Quand $n \to \infty$, la sigmoïde tend vers une fonction seuil.

Dans ce texte, nous aborderons successivement dans les sections:

2. notre description logique sous sa forme la plus simple, que nous qualifions ici de logique "naïve", quoiqu'elle se distingue déjà de la description classique par son caractère *asynchrone*

3. l'emploi, lorsque la nécessité s'en fait sentir, de *variables à plus de deux niveaux* (Van Ham, 1979, Richelle, 1986; Snoussi, 1989, Thomas & D'Ari, 1990, Thomas, 1991, [30, 33, 42, 43, 48])

4. l'introduction, cruciale, de *paramètres logiques* (Snoussi, 1989 [33]), qui donnent un poids à chacun des termes d'une expression logique

5. l'introduction de *valeurs logiques associées aux seuils* (Thomas & D'Ari, 1990,Thomas, 1991, Snoussi & Thomas, 1993, [34, 42, 43])

6. l'analyse logique généralisée d'un système simple

7. une *"logique inverse"* qui consiste, plutôt que de procéder du modèle vers ses implications, à utiliser une démarche inductive, synthétique, des faits expérimentaux vers le modèle

8. l'emploi d'une analyse logique "au second degré", portant sur les inégalités entre valeurs (ou somme de valeurs) de délais.

2. Description logique "naïve" mais asynchrone

Sous sa forme la plus simple, la description logique associe une variable logique x à chaque élément jugé pertinent du système. Cette variable prend la valeur 1 ("présent") quand la valeur *réelle x* excède *s*, le niveau du seuil, 0 ("absent") sinon. Insistons sur le fait que x = 0 ne signifie nullement que la concentration *x* soit nulle, mais simplement qu'elle est inférieure au seuil *s.*

L'état du système peut donc être décrit par un vecteur logique x y z …, qui donne les valeurs logiques des variables dans un ordre convenu; par exemple, le vecteur d'état 010 décrit un état où l'élément x est "absent", l'élément y, "présent" et l'élément z "absent".

Classiquement, un système est décrit par des équations logiques donnant pour chaque variable, sa valeur au temps t+1 en fonction de l'état du système au temps t:

$$(x, y, z, \ldots)_{t+1} = f(x, y, z, \ldots)_t$$

Le temps est donc introduit par l'attribution, à chaque valeur $(x\ y\ z, \ldots)_t$ du vecteur d'état au temps t, d'un "état suivant" $(x\ y\ z, \ldots)_{t+1}$. Cette attitude "synchrone" pose de sérieux problèmes pour l'application aux systèmes biologiques, (a) parce que dans cette description chaque état du système a un, et un seul état suivant possible, ce qui empêche toute possibilité de choix, (b) parce qu'elle exige dans bien des cas la commutation exactement simultanée (de 0 à 1 ou de 1 à 0) des valeurs de deux ou plusieurs variables.

Dans le but d'éviter ces problèmes, nous avons proposé une autre manière d'introduire le temps dans la description logique [38, 39, 43].

Alors que dans la description classique $(x, y, z, \ldots)_{t+1}$ est *l'état suivant* de $(x, y, z, \ldots)_t$, nous écrivons:

$$(X, Y, Z, \ldots) = f(x, y, z, \ldots)$$

où (X, Y, Z, …) est, à tout moment, l'*image* du vecteur d'état (x, y, z, …) par la transformation f.

Pour bien montrer la différence entre les deux attitudes, prenons l'exemple d'un gène dont le produit est x. Quand le produit est présent, nous écrivons, x = 1, quand il est absent, x = 0. Quand le gène est allumé, le produit x est destiné à apparaître et nous écrivons X = 1. Si au contraire le gène est éteint, le produit du gène, qui est périssable, est destiné à disparaître, et nous écrivons X = 0.

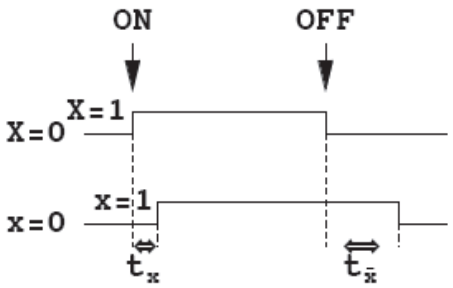


Figure 2: Ligne supérieure: le système synthétique, initialement inactif (X = 0), est allumé (X = 1) par un signal (flèche des gauche), puis éteint (X = 0) par un second signal (flèche de droite). Ligne inférieure: le produit du gène "apparaît" (= atteint sa concentration seuil) un certain temps (délai: t_X) après le signal "on" et "disparaît" (= tombe sous sa concentration seuil) un certain temps (délai: $t_{\bar{X}}$) après le signal "off".

Supposons que le gène soit initialement éteint, qu'un signal (par exemple l'apparition d'un activateur ou la disparition d'un répresseur) l'allume puis qu'après un certains temps, un second signal l'éteigne. Nous aurons successivement X = 0, X = 1, X = 0 (Figure 2, haut). Qu'en est-il de la présence du produit (Figure 2, bas)? Initialement, si le gène est éteint depuis un temps suffisant, son produit qui est périssable est absent (x = 0). Dès que le premier signal a été donné, le gène X entre en fonction et le produit commence à être synthétisé; il va apparaître, mais pas immédiatement. Durant cette période intermédiaire, nous avons X = 1 (gène allumé) mais x = 0 (produit pas encore présent): la valeur de la variable apparaît comme une mémorisation de la situation qui prévalait avant que la commande (X = 1) ne soit donnée Le signal a fonctionné comme une commande dont la réalisation effective n'aura lieu qu'après un délai t_X. Formellement, la variable x a gardé la valeur 0 mais il y a un ordre d'aligner cette valeur à la

valeur "1" de la fonction X; ceci a lieu après un délai t_x, à moins qu'un contre-ordre (extinction du gène) ait eu lieu avant l'expiration du délai.

Quand le produit atteint sa concentration efficace, nous sommes dans la situation X = 1 et x = 1, et nous nous maintenons dans cette situation de régime tant que le gène est allumé. Dès que le second signal (extinction du gène) a eu lieu, nous avons X = 0, mais, initialement du moins, le produit reste présent, en sorte que x garde momentanément la valeur 1. Comme le produit a cessé d'être synthétisé et qu'il est périssables, il y a une commande de disparition du produit. En d'autres termes, la variable x a un ordre de s'aligner sur la valeur de la fonction X. Cet ordre n'est exécuté qu'après un délai $t_{\bar{x}}$, à la suite duquel nous nous retrouvons dans la situation X = 0, x = 0. La situation est dépeinte par la Figure 2.

Il n'y a aucune raison pour que les délais t_x et $t_{\bar{x}}$ soient égaux. Dans la réalité, ils sont le plus souvent très inégaux. Ainsi, dans le cas du fonctionnement d'un gène, t_x dépend principalement du taux de synthèse du produit du gène, $t_{\bar{x}}$ surtout de la stabilité du produit. En pratique, la durée des t_x est typiquement de quelques minutes, alors que celle des $t_{\bar{x}}$ peut varier entre quelques minutes et de longues heures selon la stabilité du produit.

Ce qui a été dit des délais d'enclenchement et de déclenchement d'une même variable est évidemment aussi vrai des délais associés à des variables différentes: il n'y a aucune raison que les divers délais d'enclenchement (ou de déclenchement) soient égaux entre eux. Ceci nous conduit à une description pleinement *asynchrone*, où tous les délais sont différents à moins d'une égalité fortuite. De fait, nos systèmes sont traités comme des automates asynchrones. En retour, notre formalisme (pas seulement le formalisme "naïf" décrit ici, mais aussi ses généralisations décrites ci-dessous) constitue un outil efficace pour la description et le traitement d'automates asynchrones en général.

La différence entre les descriptions synchrone et asynchrone peut encore être illustrée par l'exemple très simple de deux gènes qui s'inhibent mutuellement. La formulation synchrone est:

$$\begin{aligned} x_{n+1} &= \overline{y_n} \\ y_{n+1} &= \overline{x_n} \end{aligned} \qquad (1a)$$

et la table des états correspondante est donnée dans la Table 1a.

La formulation asynchrone est:

$$\begin{aligned} X &= \bar{y} \\ Y &= \bar{x} \end{aligned} \qquad (1b)$$

et donne la table des états de la Table 1b. A première vue, le contenu des deux tables paraît identique. Leur signification est cependant, profondément différente. Dans les deux cas, la colonne de gauche donne simplement, dans un ordre arbitraire, la liste de 4 états possibles d'un système binaire à deux variables.

Table 1a

$(x\ y)_t$	$(x\ y)_{t+1}$
0 0	1 1
0 1	0 1
1 0	1 0
1 1	0 0

Table 1b

x y	X Y
$\overset{+\ +}{0\ 0}$	1 1
0 1	0 1
1 0	1 0
$\overset{-\ -}{1\ 1}$	0 0

Table 1: (1a) Table des états du système logique synchrone. (1b) Table des états du système logique asynchrone. On peut remarquer que les données sont identiques mais que dans le premier cas on donne l'état suivant $(x,y)_{t+1}$ en fonction de $(x,y)_t$ alors que dans le second cas on donne l'image (X,Y) de l'état (x,y) à un temps quelconque.

Dans la table 1a, la colonne de droite donne, pour chacun de ces états, l'état suivant. Dans cette description, chaque état a donc un, et un seul état suivant, sans possibilité de choix. En outre, le fait que l'état suivant 00 soit 11 (et que l'état suivant 11 soit 00) implique que dans ces situations les produits des deux gènes apparaissent ou disparaissent systématiquement de manière exactement simultanée, ce qui est totalement irréaliste.

Dans la seconde table, la colonne de droite donne, pour chaque vecteur, non plus son état suivant, mais son *image* par la transformation f. Or, si l'image du vecteur 00 est 11, cela ne signifie pas que l'état suivant 00 soit 11, mais que les variables ont toutes deux un ordre de commutation de 0 à 1. Selon que l'un ou l'autre de ces deux ordres sera obéi le premier, le système passera de 00 à 10 ou de 00 à 01. La double commutation n'est pas exclue, mais considérée comme marginale, alors qu'elle est impérative dans la description synchrone,

Notons qu'une variable fait l'objet d'un ordre de commutation chaque fois que sa valeur diffère de celle de son image. Il est commode de symboliser ces situations par un suscrit "+" ou "-" selon que la variable est l'objet d'un ordre de commutation de 0 à 1 ou de 1 à 0. Ainsi, l'état 00, dont l'image est 11, peut se symboliser 00/11 ou, de manière plus compacte par $\overset{+\ +}{0\ 0}$, et de

même, l'état 11, dont l'image est 00, par 11/00 ou par $\bar{1}\,\bar{1}$. On retrouvera cette formulation dans la table 1b, où elle est redondante, mais commode.

Dans la table 1a, l'état 01 a pour état suivant 01, et l'état 10 a pour état suivant 10. De même, dans la table 1b, l'état 01 a pour image 01, et l'état 10, l'image 10. Les deux formalismes s'accordent donc pour dire que le système en question (un circuit positif) donne lieu à *deux états stables.*

Cependant, la description synchrone (Table 1a) prédit un troisième attracteur, périodique $00 \rightleftarrows 11$, qui ne correspond à aucune réalité concrète et est en contradiction avec la description différentielle.

Par contre, la description asynchrone (Table 1b) prévoit que, partant, soit de l'état $\overset{+}{0}\,\overset{+}{0}$, soit de l'état $\bar{1}\,\bar{1}$, le système puisse atteindre l'un ou l'autre des deux états stables, 01 ou 10 selon l'ordre des commutation, en d'autres termes, selon les valeurs des délais. En outre, la forme raffinée de la description asynchrone (voir section 6) prévoit un troisième état stationnaire, instable, de type col, localisé sur la séparatrice entre les bassins d'attraction des deux états stables, en parfaite conformité avec la description différentielle.

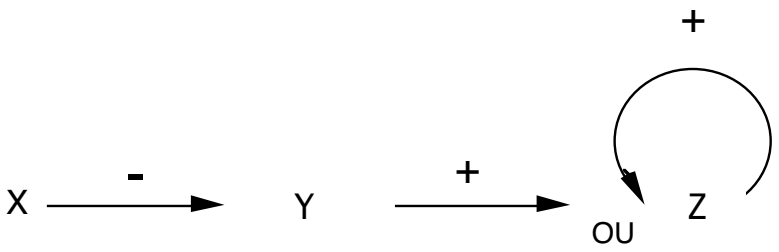
L'exemple, choisi très simple à dessein, n'a d'autre motivation que de montrer que les prévisions des descriptions synchrone et asynchrone sont différentes et que la première peut faire prévoir des attracteurs artefactuels. On trouvera dans Thomas & D'Ari ([43], appendice 4) et dans [41] des exemples plus complexes, où le graphe des séquences d'états est infiniment plus riche dans la description asynchrone.

Un exemple de la description naïve (mais asynchrone), inspiré du cas concret du choix de la réponse lytique vs lysogène chez les bactéries infectées par un bactériophage tempéré

Lorsqu'un bactériophage tempéré infecte une population bactérienne, une partie de cette population, pourtant génétiquement homogène, se lyse en libérant de nombreuses particules de phage, alors que les autres bactéries de la culture établissent avec le virus une symbiose stable et survivent; on les qualifie de bactéries lysogènes. Dans ce cas, le génome viral, inséré dans la continuité du chromosome bactérien, et devenu dès lors partie intégrante du génome bactérien, exprime un gène (cI) qui réprime l'expression de tous les autres gènes du virus, ainsi rendu inoffensif. La bactérie lysogène est *immune* non seulement vis-à-vis du virus qu'elle véhicule désormais héréditairement, mais aussi de l'infection par tout virus extérieur de même spécificité.

Voici une description très simplifiée du mécanisme responsable du choix entre réponse lytique et réponse lysogène.

Soient trois gènes, X, Y et Z, soumis aux contrôles suivants:



Dans ce modèle simplifié, le gène X est "constitutif", c'est-à-dire qu'il s'exprime en toutes circonstances. Le gène Y fonctionne à condition que x, le produit du gène X, soit absent. Le gène Z fonctionne si l'une au moins des conditions: y présent, z présent, est réalisée. Les équations logiques sont donc:

$$\begin{aligned} X &= 1 \\ Y &= \overline{x} \\ Z &= y + z \end{aligned} \qquad (2)$$

où "+" est le symbole logique du OU inclusif. La Table 2 est la table des états correspondante.

Table 2

x y z	X Y Z
$\overset{+}{0}\,\overset{+}{0}\,0$	1 1 0
$\overset{+}{0}\,\overset{+}{0}\,1$	1 1 1
$\overset{+}{0}\,1\,\overset{+}{0}$	1 1 1
$\overset{+}{0}\,1\,1$	1 1 1
[1 0 0]	1 0 0
[1 0 1]	1 0 1
$1\,\overset{-}{1}\,\overset{+}{0}$	1 0 1
$1\,\overset{-}{1}\,1$	1 0 1

Pour construire le graphe des séquences d'états à partir de la table des états, nous choisirons comme état initial 000, puisqu'au moment de l'infection l'ADN du virus est nu et qu'aucune protéine virale n'est présente à l'intérieur de la bactérie. Le graphe des séquences d'états est donné à la Figure 3.

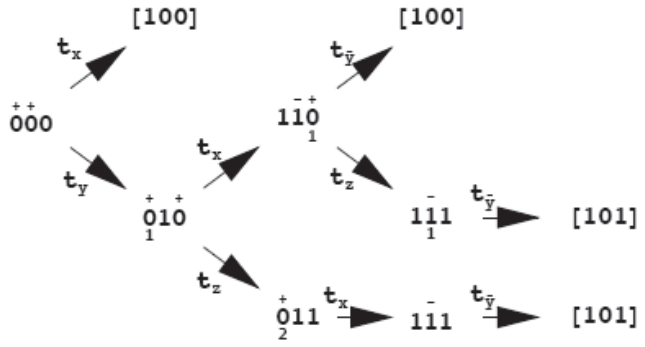


Figure 3: Graphe de transition (= des séquences d'états) du système simplifié représentant le choix entre la réponse lytique et la réponse lysogène après infection d'une bactérie par le bactériophage lambda.

Pour construire le graphe des séquences d'états à partir de la table des états, nous choisirons comme état initial 000, puisqu'au moment de l'infection l'ADN du virus est nu et qu'aucune protéine virale n'est présente à l'intérieur de la bactérie. Le graphe des séquences d'états est donné à la Figure 3.

Partant de l'état $\overset{+}{0}\,\overset{+}{0}\,0$, le système aboutit en fin de compte soit à l'état stable [100], soit à l'état stable [101], par l'un ou l'autre de quatre chemins possibles. Notons au passage que cet exemple, comme le précédent, conduit à de la multistationnarité. Celle-ci est liée à la présence d'un circuit positif - en l'occurrence, l'autorégulation positive du gène Z. Du point de vue biologique, ce mécanisme simple suffit à rendre compte de ce qu'un gène peut être *durablement éteint* ou *durablement allumé* selon le trajet suivi.

A ce niveau de l'analyse, on ne peut manquer de se poser la question: de quoi dépend le chemin suivi? Ce problème est abordé en termes de délais. Au niveau de l'état initial $\overset{+}{0}\,\overset{+}{0}\,0$, deux des trois gènes sont allumés. Soient t_x et t_y les délais requis entre l'allumage de ces gènes et l'exécution de l'ordre, c'est à dire, le moment où les produits de ces gènes atteignent leur seuil d'efficacité. Selon que t_x ou t_y aura été le plus court, l'état suivant sera [100] ou $\overset{+}{0}\,1\,\overset{+}{0}$. Dans le premier cas, on atteint l'état [100], qui est stable, et les ordres d'expression des gènes Y et Z sont annulés. Si le délai t_y a été le plus court, on atteint l'état $\overset{+}{0}\,1\,\overset{+}{0}$; ici, un nouveau choix s'impose, symbolisé par la présence de deux suscrits. Ce cas est plus intéressant: en

effet, si on consulte le graphe des séquences d'états, on peut remarquer que l'ordre de synthétiser z vient d'être donné, alors que l'ordre de synthétiser x a déjà été donné à l'étape précédente. Un simple schéma permet de voir sans difficulté qu'ici la décision de passer de $\overset{+}{0}1\overset{+}{0}$ à $\overset{-}{1}\overset{+}{1}0$ ou à $\overset{+}{0}11$ dépend des valeurs relatives de t_x et de $t_y + t_z$. Si le système choisit $\overset{-}{1}\overset{+}{1}0$, un nouveau choix s'impose, qui aboutit soit à l'état stable [100], soit (via $\overset{-}{1}\overset{-}{1}1$) à l'état stable [101]. Par contre, si on aboutit en $\overset{+}{0}11$, il n'y a plus de choix et on débouche (en deux étapes) sur le second état stable. Ce type d'analyse peut être rationalisé sans problème et automatisé si on le désire.

Nous associons à chacune des inégalités qui se présente une variable logique, m, n, p… . Dans le cas présent, nous posons:

$m \equiv t_x < t_y$

$n \equiv t_x < t_y + t_z$

$p \equiv t_x + t_{\bar{y}} < t_y + t_z$

Chaque chemin (ou ensemble de chemins menant, par exemple, à un état stable donné) obéit à des contraintes bien définies sur ces inégalités entre délais ou sommes de délais, donc sur les valeurs des variables logiques m, n et p. Simplifications faites, les conditions se réduisent, dans le cas présent à ceci: le système évoluera vers l'état stable [100] (pas d'immunité) si et seulement si m + p, et vers l'état stable [101] (immunité) si et seulement si $\bar{m}.\bar{p}$. On peut trouver dans [39] et [43] les méthodes de simplification appropriées et leur application au cas qui vient d'être décrit et à des cas nettement plus complexes.

En termes concrets, le gène X est allumé d'emblée et tôt ou tard son produit sera présent de manière stable. Dès que cette situation est réalisée, le gène Y s'éteint, et tôt ou tard son produit sera absent de manière stable. Quant au gène Z, il s'allumera (et, dans ce cas, restera allumé) si et seulement si le gène Y arrive à le faire s'exprimer avant que son propre produit y ait disparu sous l'effet de la répression par le produit de X. Ceux qui connaissent la régulation de l'immunité chez les bactériophages tempérés auront reconnu en X, Y et Z les gènes cro, cII et cI du bactériophage lambda. Ils auront remarqué aussi le caractère provisoirement simplifié des la description: en réalité, le gène Z (cI) exerce un contrôle négatif sur l'expression de tous les autres gènes, et en particulier de X (cro) et de Y (cII), et de plus, le gène cro exerce son effet négatif non seulement sur le gène cII mais aussi sur lui-même (les deux gènes sont localisés dans le même opéron). Les équations logiques deviennent:

$$X = \bar{x} \cdot \bar{z}$$
$$Y = \bar{x} \cdot \bar{z}$$
$$Z = y + z$$

Cependant, malgré les apparences, passer, fût-ce à titre provisoire, par un schéma plus simple, se justifie par le fait que ce dernier suffit à rendre compte de l'essence du mécanisme de choix entre expression et non-expression du gène cI. En effet, les conditions sur les délais sont pratiquement identiques, que l'on analyse le schéma simplifié ou ses dérivés. Remarquons par ailleurs que tant que la sophistication croissante du modèle n'implique pas l'utilisation de variables supplémentaires, le traitement en logique naïve de ces variantes plus raffinées est à peine plus complexe: dans un cas comme dans l'autre on se trouve en présence d'un système d'équations logiques dont la traduction en table des états est immédiate. Il faut cependant être attentif au fait que, dans la version élaborée, les gènes agissent tous deux en plus d'un point (X agit négativement sur lui-même et sur Y, Z agit positivement sur lui-même et négativement sur Y). Ceci nous mène tout naturellement à la section suivante. Comme nous le verrons, dans de telles conditions il faut envisager l'emploi de variables logiques à plus de deux niveaux (0, 1, 2,…) au lieu de simplement 0 et 1.

3. Variables logiques à plus de deux valeurs

Soit une variable réelle x, à laquelle nous sommes amenés à attribuer deux seuils, $s^{(1)}$ et $s^{(2)}$. La variable logique "multivaluée" x associée à la variable réelle x peut prendre trois valeurs: 0, 1, ou 2:

$x = 0$ pour $x < s^{(1)}$

$x = 1$ pour $s^{(1)} < x < s^{(2)}$

$x = 2$ pour $s^{(2)} < x$

Cependant, lorsque nous considérons un processus particulier, ce qui nous intéresse, c'est de savoir si la valeur réelle de la variable excède ou non le seuil associé à ce processus. C'est pourquoi, outre la variable multivaluée x, nous introduisons des variables binaires auxiliaires $x^{(1)}$, $x^{(2)}$ [48] définies comme suit [1]:

$x^{(1)} = 0$ pour $x < s^{(1)}$

$x^{(1)} = 1$ pour $x > s^{(1)}$

et

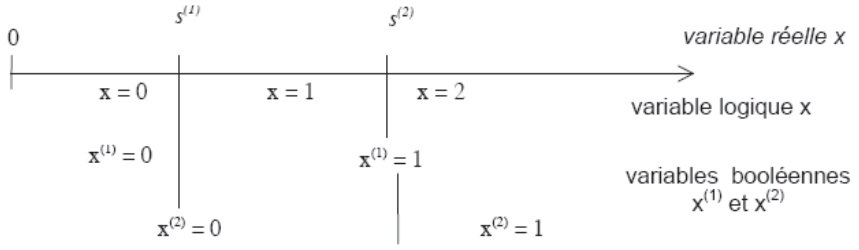
[1] On notera que dans cette description, comme dans la description "naïve", le cas marginal où la variable réelle a précisément la valeur d'un seuil n'est pas considéré. Nous reviendrons sur ce point à la section 5.

$x^{(2)} = 0$ pour $x < s^{(2)}$

$x^{(2)} = 1$ pour $x > s^{(2)}$

Le schéma de la Figure 4a visualise la variable réelle, la variable multivariée et les variables booléennes auxiliaires. Les deux notations logiques ont chacune leurs avantages, et nous utilisons l'une ou l'autre selon les besoins.

(a)



(b)

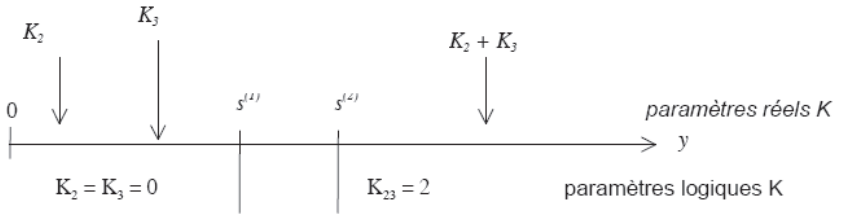


Figure 4: (a) L'axe représente la variable réelle *x*.. On peut voir sur cet axe les valeurs réelles ($s^{(1)}$, $s^{(2)}$) des deux seuils associés à la variable *x* . La première ligne sous l'axe donne les valeurs (0, 1 ou 2) de la variable multivaluée x. La deuxième et la troisième lignes sous l'axe donnent les valeurs des deux variables booléennes auxiliaires, $x^{(1)}$, $x^{(2)}$. (b) Dans le cas présent, les paramètres réels K_2 et K_3 sont tous deux inférieurs au seuil $s^{(1)}$, mais leur somme K_2+K_3 est supérieure au seuil $s^{(2)}$. Les valeurs des paramètres logiques sont donc $K_2 = 0$ et $K_3 = 0$, mais $K_{23} = 2$.

Pour rendre plus concret ce premier raffinement de la description logique, considérons le système dont la formalisation "naïve" est :

$$\begin{aligned} X &= \overline{y} \\ Y &= x + y \end{aligned} \qquad (3)$$

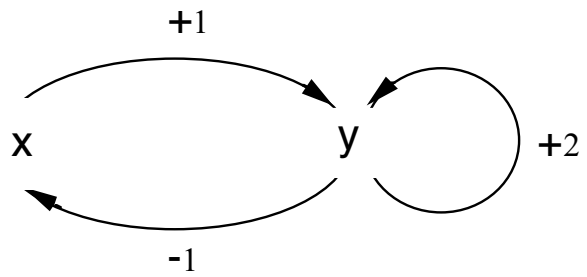
Table 3

x y	X Y
$\overset{+}{0}\,0$	1 0
[0 1]	0 1
$1\,\overset{+}{0}$	1 1
$1\,\overset{-}{1}$	0 1

La Table 3 donne la table des états, qui conduit au graphe des séquences d'états:

$$\overset{+}{0}\,0 \rightarrow 1\,\overset{+}{0} \rightarrow 1\,\overset{-}{1} \rightarrow [01]$$

Cependant, le fait que y exerce à la fois un contrôle à la fois sur l'expression de X et sur sa propre expression nous amène à considérer deux seuils, $s^{(1)}$ et $s^{(2)}$, correspondant à des valeurs croissantes de y. Supposons que dans le cas considéré y exerce son contrôle négatif sur X dès que sa concentration excède le seuil inférieur ($s^{(1)}$), mais n'exerce son effet positif sur sa propre expression que si sa concentration excède le seuil supérieur ($s^{(2)}$). Le graphe des interactions devient:



Les équations logiques sont à présent:

$$\begin{aligned} X &= \overline{y^{(1)}} \\ Y &= x^{(1)} + y^{(2)} \end{aligned} \qquad (4)$$

où $x^{(1)}$, $y^{(1)}$ et $y^{(2)}$ sont des variables booléennes et "+" est, comme précédemment, le OU inclusif. Dans cette description purement Booléenne, les images X et Y ne peuvent donc prendre que les valeurs 0 ou 1 (voir, par exemple, [30]). La table des états (Table 4a) est obtenue comme suit:

- dans le cas considéré, y doit être traité comme une variable ternaire, mais x reste binaire

- X = 1 si $\overline{y^{(1)}} = 1$, c'est-à-dire, si y = 0. Sinon X = 0
- Y = 1 si $x^{(1)} = 1$ ou (inclusif) $y^{(2)} = 1$, c'est-à-dire, x = 1 ou y = 2. Sinon Y = 0

Table 4a

x y	X Y
$\overset{+}{0}\ 0$	1 0
$0\ \overline{1}$	0 0
$0\ \overline{2}$	0 1
$1\ \overset{+}{0}$	1 1
$\overline{1}\ 1$	0 1
$\overline{1}\ \overline{2}$	0 1

Table 4a: Table des états du système (4) sous sa forme courante.

Table 4b

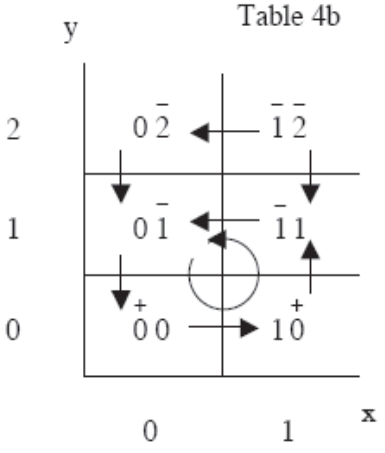


Table 4b: Table des états du système (4) avec une disposition des variables qui permet une vision dans l'espace des phases.

La situation est représentée de manière plus vivante dans la Table 4b, où l'on voit que la nouvelle description prévoit un comportement plus complexe que celui de la description naïve: comportement périodique où x et y prennent chacune alternativement les valeurs 0 et 1. Nous verrons dans la section suivante qu'une sophistication supplémentaire est opportune.

4. Paramètres logiques

Un gène soumis à une régulation complexe peut être exprimé de manière significative, mais néanmoins à des degrés très différents selon les conditions. D'une manière générale, il peut être souhaitable d'attribuer un *poids* distinct aux différents éléments d'un contrôle. Ce progrès crucial a pu être réalisé grâce à l'introduction des paramètres logiques par Snoussi [33, 34].

Plutôt que de donner une description générale mais abstraite, nous allons poursuivre l'exemple précédent. Il peut s'écrire en termes d'équations différentielles linéaires par morceaux [10]:

$$\begin{aligned} \dot{x} &= k_1 \overline{y^{(1)}} - \lambda_1 x \\ \dot{y} &= k_2\, x^{(1)} + k_3 y^{(2)} - \lambda_2 y \end{aligned} \qquad (5)$$

où les fonctions à seuil son symbolisées en termes des variables booléennes $x^{(1)}$, $y^{(1)}$ et $y^{(2)}$. Les caractères italiques représentent les variables et constantes réelles.

SNOUSSI [33] montre (dans le cas général) que la dynamique qualitative d'un tel système peut être déduite des équations logiques généralisées:

$$\begin{aligned} X &= d_x(K_1 \overline{y^{(1)}}) \\ Y &= d_y(K_2 x^{(1)} + K_3 y^{(2)}) \end{aligned} \qquad (6)$$

où $x^{(1)}$, $y^{(1)}$ et $y^{(2)}$ sont les variables booléennes déjà décrites, qui, rappelons-le, ne peuvent prendre que les valeurs 1 ou 0. Par contre, les K sont des nombres *réels* qui correspondent aux coefficients des termes des équations différentielles homologues (5): $K_1 = k_1 \tilde{l}_1$, $K_2 = k_2/l_2$ et $K_3 = k_3 \tilde{l}_2$**.** Le "+" est ici la somme *algébrique* (et non le OU logique) et d_x, d_y sont des opérateurs qui discrétisent les expressions entre parenthèses selon l'échelle de la variable multivaluée x ou y, respectivement.

Notons que dans l'expression de X, vu le caractère booléen de la variable $y^{(1)}$, l'argument de l'opérateur d_x ne peut prendre que les deux valeurs réelles suivantes:

0 si $y^{(1)} = 1$, c'est à dire y = 1 ou 2

K_1 si $y^{(1)} = 0$, c'est à dire y = 0

De même, dans Y, vu le caractère booléen des variables $x^{(1)}$ et $y^{(2)}$, l'argument de l'opérateur d_y ne peut prendre que l'une des 4 valeurs réelles suivantes:

0	si $x^{(1)} = 0$ et $y^{(2)} = 0$
K_2	si $x^{(1)} = 1$ mais $y^{(2)} = 0$
K_3	si $x^{(1)} = 0$ mais $y^{(2)} = 1$
$K_2 + K_3$	si $x^{(1)} = 1$ et $y^{(2)} = 1$

L'évolution du système est déduite de la comparaison entre chaque état logique (x, y) et son image (X, Y) définie par les équations (6). Quand X = x, la variable logique est stationnaire, quand X > x, il y a un ordre d'accroître sa valeur, et quand X < x, un ordre de décroître cette valeur.

Dans l'exemple choisi, tout nombre réel soumis à l'opérateur de discrétisation d_x a 2 valeurs logiques possibles, 0 ou 1, et tout nombre réel soumis à l'opérateur de discrétisation d_y a trois valeurs logiques possibles, 0, 1 ou 2. En particulier, si on pose:

$$d_x(K_1) = K_1,\ d_y(K_2) = K_2,\ d_y(K_3) = K_3 \text{ et } d_y(K_2 + K_3) = K_{23}$$

le paramètre logique K_1 ne peut avoir que les valeurs 0 ou1 (comme la variable x) et les paramètre K_2, K_3 et K_{23} ne peuvent avoir que les valeurs 0, 1 ou 2 (comme la variable y).

Il faut insister sur le fait que les paramètres logiques K résultent de la discrétisation des réels *K*, qui sont chacun le rapport des constantes cinétiques de synthèse et de dégradation du produit considéré. Il en résulte que, même s'ils ne sont pas affichés de manière explicite sous la forme de termes diagonaux négatifs dans la matrice des interactions, les taux de dégradation des produits sont pris en compte dans notre approche logique par l'intermédiaire des paramètres logiques, qui déterminent le poids relatif des différentes combinaisons d'interactions, et par les délais de déclenchement associés à chaque variable, qui déterminent la vitesse à laquelle les produits se dégradent. Il en est ainsi non seulement pour la description logique généralisée qui vient d'être évoquée mais tout aussi bien pour notre description naïve.

Trois remarques importantes:

1) comme l'expression $K_2 + K_3$ est la somme de deux réels positifs, sa valeur ne peut être inférieure ni à celle de K_2, ni à celle de K_3. Il en est de même pour les résultats de la discrétisation: K_{23} ne peut être inférieur ni à K_2 ni à K_3.

2) en général, $d_y(K_2) + d_y(K_3) \neq d_y(K_2 + K_3)$, et donc $K_{23} \neq K_2 + K_3$. Pour s'en convaincre, le schéma présenté à la Figure 4b est plus éloquent qu'une longue description. On peut constater, en effet, que dans l'exemple choisi $K_2 = 0$ et $K_3 = 0$, mais $K_{23} = 2$.

3) une généralisation importante consiste à introduire chaque fois que c'est nécessaire des termes indépendants dans la description logique. K_0 représente le niveau d'expression de base d'un gène, c'est-à-dire le niveau présent dans les conditions les plus défavorables (répresseur présent, activateur absent).

En pratique, comment construire la table des états à partir des équations logiques (6)?

La table tient compte de ce que, dans notre cas, la variable multivaluée x est binaire, alors que la variable y est ternaire. La liste des vecteurs d'état, qui constitue la colonne de gauche de la table, sera donc:

00, 01, 02, 10, 11, 12

Pour remplir la colonne de droite de la table (vecteurs image), il est commode d'écrire *successivement* que:

$X = K_1$ chaque fois que $\overline{y^{(1)}} = 1$ (c'est à dire, y = 0); sinon X = 0.

De même, nous écrirons que:

$Y = K_2$ si seule la condition $x^{(1)} = 1$ (c'est à dire, x = 1) est réalisée

$Y = K_3$ si seule la condition $y^{(2)} = 1$ (c'est à dire, y = 2) est réalisée

$Y = K_{23}$ si les deux conditions sont réalisées

Y = 0 si aucune des deux conditions n'est réalisée.

On obtient ainsi la table des états (Table 5). Notons que ces tables peuvent être obtenues sans difficulté de manière automatisé [11, 28, 36]. Cette possibilité devient réellement utile dès que le nombre de variables ou de niveaux logiques s'élève.

Si l'on compare la Table 5 à la Table 4a, on constate qu'elle est beaucoup plus générale en ce sens que les valeurs "1" ou "2" sont remplacées par les paramètres K_1 (qui peut prendre les valeurs 0 ou1), K_2 ou K_3 (qui peuvent prendre les valeurs 0, 1 ou 2), ou K_{23} (qui peut prendre la valeur 0, 1 ou 2 pourvu que cette valeur ne soit inférieure ni à celle de K_2 ni à celle de K_3). Notons que donner à l'un des paramètres la valeur logique 0 revient à admettre que par elle-même l'interaction considérée est sans effet. Cependant, comme nous l'avons vu plus haut par un exemple, deux ou plusieurs interactions peuvent être inefficaces chacune par elle-même mais fonctionnelles par la concertation de leurs effets. La table 6 montre trois situations qualitatives distinctes résultant de trois choix de valeurs des paramètres.

Table 5

x y	X Y
0 0	K_1 0
0 1	0 0
0 2	0 K_3
1 0	K_1K_2
1 1	0 K_2
1 2	0 K_{23}

Table 5: Table des états du système (4), généralisée par l'introduction des paramètres logiques.

Comme il est décrit en détail dans Thomas & D'Ari ([43], chapitre 8) on peut repasser de cette description logique généralisée à une description différentielle utilisant comme interactions des fonctions à seuil ou sigmoïdes. La relation entre nos paramètres logiques et les constantes cinétiques de la description différentielle est simple. K = n dans la description logique signifie que dans la description différentielle linéaire par morceaux $k/l > s^{(n)}$. Nous verrons d'autre part plus loin (section 6) comment utiliser la table des états généralisée de manière plus rationnelle, en évitant de devoir procéder de manière quelque peu arbitraire au choix des valeurs de paramètres logiques.

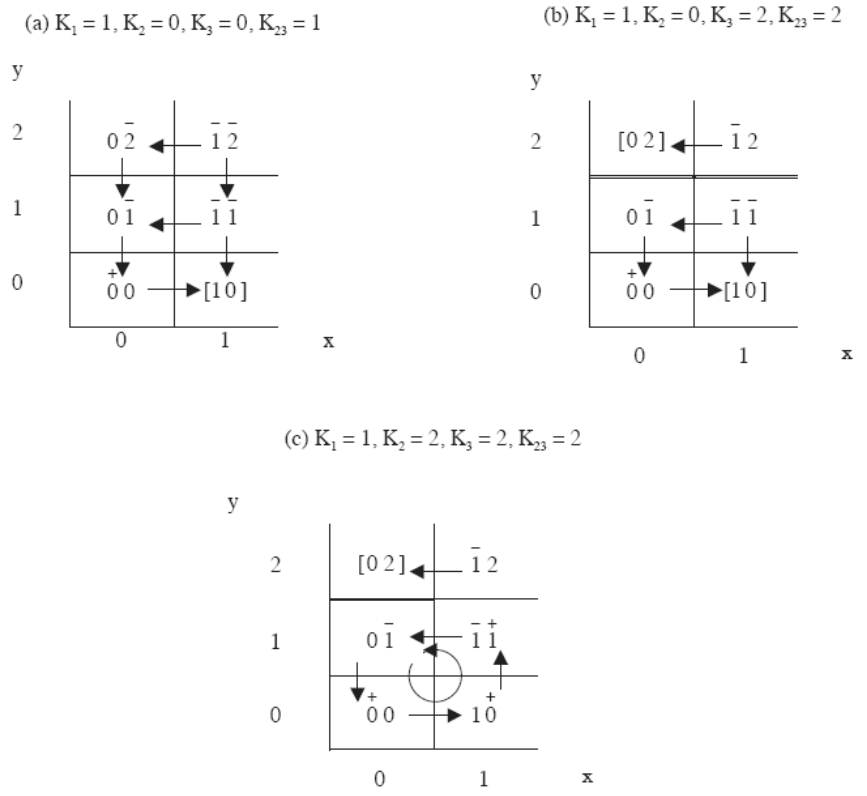


Table 6: Tables des états du système (4): trois choix de valeurs des paramètres.

5. Attribution d'une valeur logique aux seuils. Identification en termes logiques de tous les états stationnaires. Le concept d'état caractéristique d'un circuit

Jusqu'ici, aussi bien dans la description logique "généralisée" que dans la description "naïve", nous avons défini comme états logiques stables ceux dont le vecteur d'état (x y z …) et le vecteur image (X Y Z …) sont égaux. En effet, dans ces cas, aucune des variables n'a d'ordre de changer sa valeur (pas de suscrit "+" ou "-").

Il a été réalisé rapidement (puis démontré par Snoussi [33]) que ces états logiques stables correspondent à des nœuds stables, tout au moins dans la description différentielle de systèmes à seuils. Les autres types d'états stationnaires de la description différentielle, et en particulier les états

stationnaires instables, ne sont pas identifiés par la description logique courante. La raison en est simple: un état stationnaire peut être localisé au niveau d'un ou plusieurs seuils; or, dans la description logique utilisée jusqu'ici, nous avions considéré les situations réelles $x < s$ (décrite en termes logiques par x = 0) et $x > s$ (décrite en termes logiques par x = 1), mais pas la situation marginale $x = s$. Ainsi, nous ne pouvions identifier comme états logiques ceux des états localisés sur un (ou plusieurs) seuils. Cette difficulté peut être résolue si l'on inclut les valeurs de seuil dans l'échelle des valeurs logiques. Nous écrivons donc:

valeurs logiques		valeurs réelles
$x = 0$	si	$x < s$
$x = s^{(1)}$	si	$x = s^{(1)}$
$x = 1$	si	$s^{(1)} < x < s^{(2)}$
$x = s^{(2)}$	si	$x = s^{(2)}$
$x = 2$	si	$s^{(2)} < x < s^{(3)}$, etc.…

Notons que les valeurs *réelles* du n$^{\text{ième}}$ seuil de x et de y sont symbolisés $s_x^{(n)}$, $s_y^{(n)}$, etc., parce que ces valeurs sont en général différentes. Mais dès qu'il s'agit de valeurs *logiques*, $s^{(n)}$ est la valeur logique correspondant au n$^{\text{ième}}$ seuil, quelle que soit sa valeur réelle, et ce même symbole convient pour désigner le niveau logique du n$^{\text{ième}}$ seuil des variables x et y, même si ces seuils correspondent à des valeurs réelles différentes.

Dès le moment où les seuils sont inclus dans l'échelle des valeurs logiques, il devient nécessaire de distinguer des états *réguliers*, correspondant aux états logiques classiques et des états *singuliers*, où la valeur d'une (ou plusieurs) variable est une valeur de seuil. Alors qu'un état logique régulier est un "pavé" dont la dimension est celle du nombre de variables du système, un état singulier localisé sur un seuil occupe la jonction entre deux pavés et, plus généralement, un état singulier localisé sur n seuils est à l'intersection entre les 2^n états réguliers adjacents à l'état singulier. Pour fixer les idées, dans un système à trois variables, un état régulier occupe un pavé 3D et des états singuliers localisés sur 1, 2, 3 seuils occupent, respectivement, une surface (intersection de deux pavés), une ligne (intersection de 4 pavés) et un point (intersection de 8 pavés).

Il peut être utile d'introduire ici le terme d'*état partiel*, pour désigner un état dont les niveaux logiques ne sont définis que pour une partie des variables, en d'autres termes, un état défini seulement dans le sous-espace de ces variables.

Tant que l'échelle des valeurs logiques ne comportait que des valeurs entières, on pouvait se contenter de définir un état comme "stable" ou non selon que son image lui est on non identique. Dès que l'échelle est étendue

aux valeurs de seuil, et qu'en conséquence les états stationnaires instables sont pris en compte dans la description logique, il devient indispensable d'inclure dans cette description une définition de la stationnarité en termes logiques. Nous avons proposé (THOMAS, [42]) d'appeler "équations logiques d'état stationnaire " les relations:

$$x = X, y = Y, z = Z, \ldots \text{ etc},$$

et de définir comme *état logique stationnaire* tout état logique, régulier ou singulier dont l'image soit *compatible* avec les équations d'état stationnaire. Cette définition est cohérente avec la description différentielle, puisque, comme nous l'avons vu plus haut la dérivée par rapport au temps est, à un facteur près, égale à $X - x$, et par conséquent, s'annule quand $X = x$. Nous disons donc qu'une variable est stationnaire quand son image est compatible avec l'équation d'état stationnaire correspondante (par exemple, l'état 101/011 ou $\bar{1}\dot{0}1$ n'est pas stationnaire, mais la *variable* z est stationnaire car $z = Z = 1$).

Pour voir si un état est stationnaire, il faut comparer cet état avec son image. Mais comment calculer l'image d'un état singulier, ou même plus simplement d'une variable singulière?

Soit un système qui comporte l'équation logique:

$$X = d_x(k\, x^{(2)}) \qquad (7)$$

La table des états *réguliers*, réduite pour simplifier à ce qui concerne la variable x, est donnée dans la Table 7. Quelles sont les images des valeurs *singulières* de x?

Table 7

x	X
0	0
1	0
2	K

Table 7: Table des états du système (5).

Cette table montre que $X = 0$, que x ait la valeur 0 ou la valeur 1. Pour des raisons de continuité, nous écrivons que $X = 0$ également pour la valeur intermédiaire $x = s^{(1)}$. *Plus généralement, quand les images d'une variable dans les états réguliers adjacents à un état singulier sont identiques, nous les prenons pour image de la variable dans l'état singulier considéré.*

Par contre, pour $x = s^{(2)}$ les deux états adjacents peuvent avoir des images différentes. En effet, pour x = 1, X = 0 et pour x =2, X = K. Dans ce cas, ce que nous pouvons dire à ce stade, c'est que pour $x = s^{(2)}$, X est compris dans l'*intervalle*]0, K[.Ceci conduit à une table élargie (Table 8, gauche).

Table 8

x	X
0	0
$s^{(1)}$	0
1	0
$s^{(2)}$	]0,K[
2	K

K = 0

x	X
0	0
$s^{(1)}$	0
1	0
$s^{(2)}$	0
2	0

K = 1

x	X
0	0
$s^{(1)}$	0
1	0
$s^{(2)}$	]0,1[
2	1

K = 2

x	X
0	0
$s^{(1)}$	0
1	0
$s^{(2)}$	$s^{(2)}$
2	2

Table 8: Tables des états du système (5), généralisées par l'attribution d'une valeur logique aux seuils. A gauche, table globale. A droite, les tables correspondant aux trois valeurs possibles du paramètre logique K.

Ce qui est nouveau ici, c'est que si la valeur logique d'une variable est une valeur de seuil, son image peut être un *intervalle*. Dans ce cas, notre définition de la stationnarité implique que la variable soit stationnaire si et seulement si la valeur de seuil considérée est incluse dans l'intervalle. Plus concrètement, dans le cas qui nous occupe, $x = s^{(2)}$ est stationnaire si et seulement si la valeur de K est telle que $s^{(2)}$ soit inclus dans]0,K[. Il faut pour cela que K vaille au moins 2. Pour les trois valeurs possibles de K, on a trois situations représentées par la Table 8 (droite).

Lesquels de ces états sont stationnaires? Les valeurs de x autres que $s^{(2)}$ ne posent pas problème: dans tous ces cas, on voit immédiatement si l'image est ou non égale à l'état. La situation est plus intéressante pour la valeur $s^{(2)}$ de la variable: on voit que pour $x = s^{(2)}$:

si K = 0 X = 0

si K = 1 $X \in]0,1[$

si K = 2 $X \in]0,2[$

Dans le troisième cas, (mais non dans les deux autres), $s^{(2)}$ est inclus dans l'intervalle considéré, et, par conséquent, dans ce cas seulement les valeurs de x et de X sont compatibles avec l'équation d'état stationnaire (x = X). Selon notre définition des états logiques stationnaires, $x = s^{(2)}$ est donc stationnaire pourvu que K = 2, et comme à l'état stationnaire x =X, on peut écrire que $X = s^{(2)}$. La table des états peut donc être complétée (Table 8). On voit que pour K = 2 le système admet trois états stationnaires, correspondant à x = 0, $x = s^{(2)}$ et x = 2.

Lorsqu'on cherche l'image d'un état singulier localisé sur plus d'un seuil, la situation se complique à première vue, car un état dont n variables sont situées sur un seuil est coincé entre 2^n "pavés", entre 2^n états réguliers adjacents. Pour pouvoir traiter cette situation générale, il est nécessaire préalablement de rappeler les propriétés fondamentales des circuits et d'introduire le concept d'état caractéristique d'un circuit.

Bref rappel de la définition et des propriétés des circuits

Nous parlons de circuit de rétroaction (feedback), ou, plus simplement, de circuit, chaque fois que nous nous trouvons en présence d'une chaîne d'interactions fermée, en d'autres termes, chaque fois que le niveau d'une variable agit, directement ou indirectement, sur sa propre évolution[2]. Un n-circuit est un circuit de n éléments (pour plus de détails, voir [42, 43]).

Tout élément d'un circuit exerce un effet direct sur l'élément suivant mais un effet indirect sur tous les éléments, y compris lui-même. Il existe deux types de circuits, selon que chaque élément exerce un effet positif sur son propre développement (circuit positif) ou que chaque élément exerce un effet négatif sur son propre développement (circuit négatif). *On distingue les deux types de circuits d'après la parité du nombre d'interactions négatives: si ce nombre est pair, il s'agit d'un circuit positif, s'il est impair, négatif.*

Les propriétés des deux types de circuits sont profondément différentes: les circuits négatifs sont impliqués dans la stabilité (ponctuelle ou périodique) et, en biologie, sont responsables de l'homéostasie. Les circuits positifs sont impliqués dans la multistationnarité, dont les manifestations biologiques les plus essentielles sont la différenciation et la mémoire.

État caractéristique d'un circuit

A chaque circuit (ou union de circuits disjoints) peut être associé un état logique défini par l'intersection des seuils impliqués dans le circuit, en

[2] La définition rigoureuse d'un circuit se situe au niveau de la matrice jacobienne du système: toute séquence d'éléments a_{ij} non nuls de cette matrice, telle que la séquence de ses indices de ligne (i) soit une permutation circulaire de celle de ses indices de colonne (j). Nous appelons “nuclei” les circuits (ou unions de circuits disjoints) qui impliquent toutes les variables d’un système. L’importance de ce concept ressort du fait qu’en l’absence de tout nucleus un système n’a pas d’état stationnaire non dégénéré.

d'autres termes, un état pour lesquels chaque variable du circuit est localisée à la valeur du seuil au-delà duquel elle est active dans le circuit considéré. Si, comme c'est fréquemment le cas, le circuit (ou l'union de circuits disjoints) considéré n'implique qu'une partie des variables du système, cet état est un état partiel (voir p. 22).

Comme nous le verrons dans un instant, l' état localisé à l'intersection des seuils d'un circuit joue un rôle essentiel dans la dynamique du système, et il est qualifié, pour cette raison *d'état caractéristique* du circuit (ou de l' union de circuits disjoints).

Le rôle crucial des états caractéristiques résulte de ce qui suit. Il est apparu [42] et a pu être formellement démontré depuis [34] que:

1) *parmi les états singuliers d'un système, seuls les états dont la partie singulière est caractéristique d'un circuit (ou d'une union de circuits disjoints) peuvent être stationnaires.* Bien entendu, pour que ce soit effectivement le cas, les variables extérieures au circuit (qui forment la partie régulière de l'état) doivent aussi être stationnaires.
2) étant donné l'état caractéristique d'un circuit (ou union de circuits disjoints**),** *il existe des valeurs de paramètres logiques pour lesquelles cet état est stationnaire dans le sous-espace des variables considérées. Lorsque ces conditions sont réalisées, le circuit est fonctionnel:* il engendre la multistationnarité s'il est positif, un cycle dans le graphe des séquences d'état s'il est négatif.

Plus concrètement, lorsqu'un circuit négatif est fonctionnel, son état caractéristique n'est autre que l'état stationnaire autour duquel s'organise l'homéostasie, avec oscillations stables si l'état stationnaire est lui-même instable, sans oscillations stables si l'état stationnaire est stable. Lorsqu'un circuit positif est fonctionnel, son état caractéristique n'est autre que le point de selle (col) localisé sur la séparatrice qui délimite les zones d'influence des différents attracteurs

Le problème de l'identification et de la caractérisation des états stationnaires d'un système est simplifié de manière drastique par ce théorème. En effet, si le nombre des états logiques singuliers croît rapidement avec le nombre des variables et le nombre de niveaux logiques de ces variables, le nombre de circuits reste en général très limité. Ceci résulte de ce que la connectivité des réseaux réels et, en particulier, biologiques, est généralement faible (KAUFFMAN, 1969). Au lieu de devoir déterminer, pour chaque état singulier d'un système, s'il est (ou dans quelles conditions il est) stationnaire, il suffit d'identifier l'état caractéristique de chaque circuit, et de déterminer pour quelles valeurs des paramètres logiques cet état est stationnaire.

Il faut insister sur le fait, déjà mentionné, qu'un état singulier ne peut être stationnaire que si non seulement sa partie singulière est stationnaire dans le sous-espace des variables du circuit, mais qu'en outre les variables formant

la partie régulière de l'état sont elles -mêmes stationnaires. Si ce n'est pas le cas, on a affaire à un "état stationnaire partiel" et, le cas échéant, à une "multistationnarité partielle", c'est à dire, limitée au sous-espace des variables du circuit. Signalons que ces notions d'état stationnaire partiel et de multistationnarité partielle ont été retrouvées depuis dans la description différentielle.

6. En pratique: analyse logique généralisée d'un système simple

Nous avons décrit plus haut un système comprenant deux circuits conjoints, l'un négatif à deux éléments, l'autre positif à un élément. Du fait de l'introduction des paramètres logiques, la table des états classique a été généralisée (Table 5).

Une manière d'exploiter ces tables d'état généralisées consiste à attribuer à chacun des paramètres K une valeur déterminée. Quelques unes des possibilités ont été présentées plus haut (Table 6). Elles permettent de voir que, malgré la simplicité de sa structure, le modèle couvre une série de situations intéressantes, depuis la simple périodicité jusqu'au choix entre deux attracteurs, l'un stable, l'autre périodique, en passant par le choix entre deux états stables. Il faut cependant remarquer que le nombre de combinaisons de valeurs des paramètres logiques est déjà passablement élevé (28) dans le cas présent et que le choix des cas réellement intéressants peut ne pas être évident à première vue.

C'est ici que nous allons à nouveau être aidés par le concept d'état caractéristique de circuits. Nous savons que le système comporte un 2-circuit négatif et un 1-circuit positif, qui sont responsables, respectivement, des possibilités de périodicité et de multistationnarité du système. C'est pourquoi nous nous posons la question: quelles sont les conditions pour que l'un, l'autre ou les deux circuits soient fonctionnels? La manière de répondre à cette question consiste à identifier les états caractéristiques des circuits et à voir pour quelles valeurs des paramètres logiques l'un, l'autre ou les deux sont stationnaires.

L'état caractéristique du 2-circuit négatif est $s^{(1)}s^{(1)}$. Pour calculer l'image de cet état, il faut la comparer aux images des états réguliers adjacents. Ces états sont au nombre de 4, mais un théorème dû à Snoussi [34] montre qu'il suffit de considérer deux de ces états, celui dont l'image est "maximale", et celui dont l'image est "minimale", et de voir pour quelles valeurs des paramètres K l'état $s^{(1)}s^{(1)}$ est compris dans l'intervalle de ces deux images. On peut tirer du théorème une "recette" simple pour identifier les états

adjacents dont les images sont maximale et minimale[3]: on donne à chaque variable les valeurs logiques juste inférieure et juste supérieure à la valeur du seuil considéré si la variable agit positivement, la valeur logique juste en dessous et juste au dessus de la valeur de seuil si la variable agit négativement.

Dans le cas de notre 2-circuit négatif, les deux états adjacents à considérer sont donc 01 et 10. La table des états (Table 5) montre que les images de ces états sont 00 et K_1K_2. Pour que le circuit soit fonctionnel il faut que $s^{(1)}$ soit inclus dans l'intervalle]0, K_1[et que $s^{(1)}$ soit aussi inclus dans l'intervalle]0, K_2[. En d'autres termes, il faut que $K_1 \geq 1$ et que $K_2 \geq 1$. Comme la variable x est binaire et la variable y ternaire, cela revient aux conditions: $K_1 = 1$, $K_2 = 1$ ou 2. Si ces conditions sont remplies, l'état $s^{(1)}$ $s^{(1)}$ est stationnaire, et le 2-circuit négatif, fonctionnel: il y aura donc un cycle dans le graphe des séquences d'états. Ceci peut être vérifié sans difficulté.

Passons maintenant au 1-circuit positif. Ce circuit ne fait intervenir que la variable y. Son état caractéristique est l'état partiel $-s^{(2)}$, où le tiret rappelle qu'il convient de considérer la stationnarité de la variable y pour les différentes valeurs régulières (entières) de la variable x. Nous envisagerons donc la stationnarité de y dans les deux états: $0s^{(2)}$ et $1s^{(2)}$. Voici les états réguliers adjacents à ces états singuliers, et leurs images:

$0s^{(2)}$

xy	XY
01	00
02	$0K_3$

$1s^{(2)}$

xy	XY
11	$0K_2$
12	$0K_{23}$

On peut voir que la variable y est stationnaire, dans le premier cas si $s^{(2)} \in$]0, K_3[c'est à dire, $K_3 = 2$, dans le second cas si $K_2 < s^{(2)} < K_{23}$, c'est à dire, $K_2 = 0$ ou 1 et $K_{23} = 2$. Cependant, la situation est très différente dans les deux cas. En effet, pour l'état:

- $0s^{(2)}$ la variable extérieure au circuit, x, est stationnaire (la variable et son image ont tous deux la même valeur, 0). L'état $0s^{(2)}$ est donc stationnaire pourvu que la condition $K_3 = 2$ soit remplie. Que cet état, caractéristique d'un circuit positif, soit stationnaire, implique la multistationnarité[4].

[3] Comme l'a fait remarquer Snoussi [34], cette méthode n'est valable que si les seuils attribués à une même variable sont distincts. Si ce n'est pas les cas, il faut comparer deux à deux celles des paires d'états dont la relation est impliquée dans le circuit considéré.

[4] Nous montrons dans [45] et [15] que la multistationnarité requiert en fait soit deux nuclei de signe opposé (au sens de Eisenfeld & de Lisi, [4]), soit un nucleus "variable" (c'est-à-dire, dont le "sign pattern dépend de la localisation dans l'espace des phases). Comme nous l'avons indiqué plus haut, notre description logique implique la présence, dans la matrice jacobienne du système différentiel homologue, de termes diagonaux négatifs linéaires, non

- En ce qui concerne l'état $1s^{(2)}$, la variable y est stationnaire si $K_2 = 0$ ou 1 et $K_{23} = 2$, mais la variable x ne peut être stationnaire en aucun cas puisqu'elle a la valeur 1 alors que son image a la valeur 0. Dans ce cas, les conditions mentionnées impliquent bien que le circuit positif soit fonctionnel (et entraîne donc la multistationnarité). Cependant, cette multistationnarité n'est que partielle: elle n'existe que dans le sous-espace des variables (ici, de la variable) du circuit.

Ce qui précède nous conduit à établir une distinction entre les conditions de fonctionnalité des circuits et les conditions de pleine stationnarité des états caractéristiques (et de pleine multistationnarité dans le cas d'un circuit positif).

Dans le cas considéré, le 2-circuit est fonctionnel (et le graphe des séquences d'états comporte un 4-cycle) si et seulement si $K_1 = 1$ et $K_2 = 1$ ou 2. Le 1-circuit positif est fonctionnel (et il y a multistationnarité au moins partielle) si et seulement si $K_3 = 2$ ou ($K_2 = 0$ ou 1 et $K_{23} = 2$).

Les conditions de stationnarité pleine des états caractéristiques (et de multistationnarité pleine) sont un peu plus restrictives. La condition pour que l'état $s^{(1)}s^{(2)}$ soit stationnaire est, comme indiqué plus haut, $K_1 = 1$ et $K_2 = 1$ ou 2. Par contre, l'état $1s^{(2)}$ ne peut être stationnaire et la condition pour que l'état $0s^{(2)}$ le soit est: $K_3 = 2$.

Lorsque les deux états caractéristiques sont stationnaires, il y a trois états stationnaires: l'état stable [02], l'état $s^{(1)}s^{(1)}$ qui engendre le cycle et l'état $0s^{(2)}$ qui, dans la description continue, est un col localisé sur la séparatrice entre les deux bassins d'attraction (Table 6c).

Lorsqu'un système comporte plusieurs circuits, comme c'est fréquemment le cas, on peut analyser séparément les conditions de fonctionnalité de chacun des circuits: il peut exister une gamme de valeurs de paramètres telle que tous les circuits soient fonctionnels, et cette situation est celle qui assure au système sa dynamique la plus complexe. Il est fréquent, par contre que deux circuits soient incompatibles. C'est le cas si les conditions de fonctionnalité sont exclusives.

7. "Logique inverse": emploi inductif (synthétique) de la méthode logique, des faits vers les modèles

Jusqu'ici, nous sommes partis d'un modèle et avons tenté d'en dériver toutes les prédictions, de manière à pouvoir les comparer à la réalité biologique. Cependant, quoique notre méthode ait été conçue initialement comme outil dans cette approche analytique, déductive, elle peut aussi être utilisée de manière inverse, inductive, synthétique. Plus concrètement: étant donné ce que l'on sait du comportement réel d'un système, dans quelle mesure peut-

explicitement représentés dans la description logique. Ces termes diagonaux constituent un circuit plein négatif, qui s'ajoute au circuit plein positif décrit dans le texte.

on procéder rationnellement vers l'élaboration d'un modèle? Ces questions sont traitées dans [39, 41, 43, 45].

Il faut être clair d'emblée: même si on s'adresse à un type bien défini de modèle, comme, par exemple, la description d'un circuit de neurones, en termes d'automates asynchrones, un comportement expérimental donné est le plus souvent compatible avec plusieurs, voire, de nombreux, modèles spécifiques. Ainsi, l'on connaît par les travaux de Friesen et Stent [6] la séquence périodique exacte des allumages et extinctions d'une série de neurones impliqués dans la locomotion de la sangsue. Ces auteurs ont développé des modèles qui rendent compte du caractère périodique de la dynamique, mais pas de la séquence précise des commutations. En admettant que les faits observés sont bien dus a des interactions entre les neurones identifiés, on peut trouver par un processus inductif rigoureux:

1) des modèles, aussi simples que possible, qui rendent compte de manière précise de la séquence expérimentale
2) quelles modifications minimales d'un modèle préexistant lui permettent de rendre compte de cette séquence exacte
3) quelles sont les contraintes absolues pour qu'un modèle puisse rendre compte du comportement réel. En fait, ce point n'est pas entièrement distinct du précédent: la liste des contraintes peut être confrontée avec le modèle pour voir où il doit être modifié.

Voici un exemple concernant la genèse du rythme pylorique chez le Homard, un système décrit par Friesen et Stent [6] d'après Maynard & Silverston [21].
Le diagramme de phase expérimental, c'est-à-dire, la séquence observée des allumages et extinctions des trois neurones considérés, est montré à la figure 5.

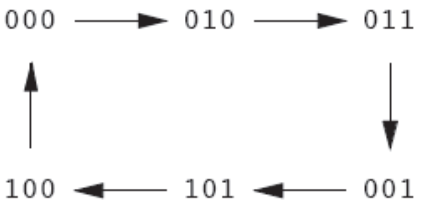


Figure 5: La séquence cyclique observée des allumages et extinctions de trois neurones contrôlant la genèse du rythme pylorique chez le Homard (Maynard & Silverston, 1975).

Pour tenter de rendre compte de cette séquence, les auteurs proposent le modèle représenté par la Figure 6a. Selon notre analyse, ce modèle prévoit deux états stables, mais pas d'oscillations. Comme le font remarquer Friesen et Stent, des oscillations seraient rendues possibles si l'un des neurones, par exemple le neurone X, oscillait par lui-même. Ce serait le cas si X exerçait

un contrôle négatif sur sa propre expression ($X = \overline{x}$). Cependant, le cycle obtenu ne reproduit pas la séquence expérimentale (Figure 6b).
Nous raisonnons comme suit. A supposer que le comportement observé résulte effectivement des interactions entre les trois neurones considérés (ce qui n'est pas évident), comment déterminer : a) quelles interactions *imposeraient* le parcours de la séquence et b) quelles interactions *permettraient* le parcours de la séquence sans pour cela l'imposer? Considérons la table des états (Table 9a) et voyons comment remplir la colonne de droite de cette table (images) pour trouver les interactions qui imposent la séquence expérimentale. Si l'état 000 *doit* avoir comme état suivant l'état 010, nous inscrivons comme image 010; si l'état 010 doit avoir comme état suivant 011, nous notons comme image 011, et ainsi de suite. Les états 110 et 111 ne font pas partie du cycle imposé et ne sont donc l'objet d'aucune contrainte. Leurs images sont donc indifférentes, et notées "- - -". Cette table des états (Table 9a) comporte 6 indéterminations; il y a donc $2^6 = 64$ manières de la remplir.

Table 9

(a)

x y z	X Y Z
0 0 0	0 1 0
0 0 1	1 0 1
0 1 0	0 1 1
0 1 1	0 0 1
1 0 0	0 0 0
1 0 1	1 0 0
1 1 0	- - -
1 1 1	- - -

(b)

x y z	X Y Z
0 0 0	- 1 -
0 0 1	1 - -
0 1 0	- - 1
0 1 1	- 0 -
1 0 0	0 - -
1 0 1	- - 0
1 1 0	- - -
1 1 1	- - -

Table 9: (a) Cette table des états indique quelles transitions sont requises pour que la séquence d'états de la Figure 5 *doive* être suivie. Les états 110 et 111 ne font pas partie du cycle, et leur image est donc indifférente (- - -), c'est-à-dire que se éléments peuvent être remplacés indifféremment par on "0" ou par un "1". (b) Cette table indique quelles transitions sont requises pour que la séquence observée *puisse* être suivie. Les autres éléments de la table sont indifférents (-), c'est-à-dire qu'ils peuvent être remplacés indifféremment par on "0" ou par un "1". Parmi les nombreuses possibilités (2^{18}, puisqu'il y

a 18 indifférences), nous avons choisi celle qui se rapproche le plus du modèle de Friesen et Stent [6] (modèle de la Figure 6c).

Des méthodes classiques permettent de trouver sans difficulté les solutions les plus simples (voir, par exemple, [5, 19, 39, 43]).

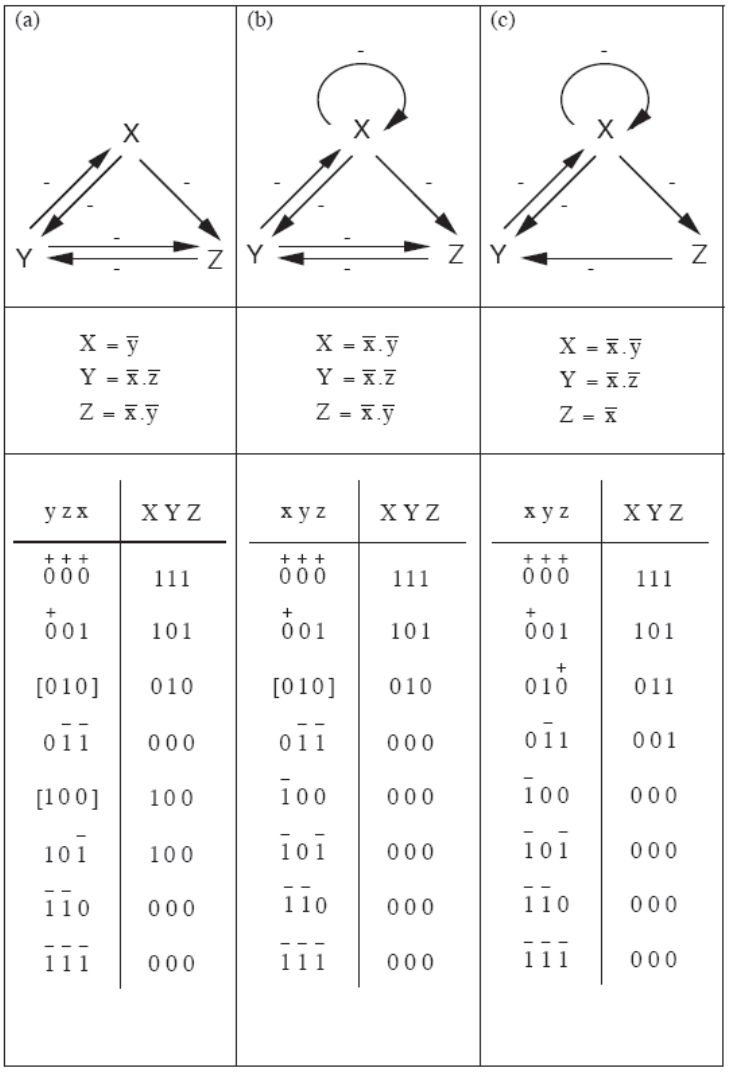


Figure 6: Trois schémas proposés pour rendre compte de la séquence d'états de la Figure 5: (a) le schéma de Maynard & Silverston [21]. Selon notre analyse, ce schéma prévoit deux états stables, mais pas d'oscillations. (b) le schéma modifié par Friesen et Stent [6] rend

compte de l'existence d'oscillations, mais pas de la séquence précise observé. (c) schéma proposé, qui ne diffère de celui de Friesen et Stent que par la suppression d'une interaction. Ce schéma tient compte de la Table 9b, qui donne les contraintes à respecter pour que la séquence observée puisse être suivie.

Voyons maintenant comment remplir la colonne de droite (images) de la table des états pour trouver les interactions qui *permettent* la séquence observée sans cependant l'imposer. On peut se rendre compte sans difficulté que l'état 000 peut avoir pour état suivant 010 si et seulement si son image est -1-. En effet, si 000 a pour image -1-, l'état suivant peut être 010, 110, 010 ou 011; par contre, si le deuxième élément de l'image n'est pas "1", l'état suivant ne peut être 010. De même, pour que l'état suivant 010 puisse être 011, il faut donner pour image à 010, - -1, et ainsi de suite. On obtient ainsi la Table 9b. Ici, la démarche la plus efficace consiste à confronter les modèles avec cette table, et à voir où ils entrent en contradiction avec elle.

On voit aussitôt que la table des états du modèle (a) est en contradiction avec deux des directives de la Table 9b: pour l'état 010, l'image de z est 0 au lieu de 1, et pour l'état100, l'image de x est 1 alors qu'elle devrait être 0. Le modèle (b) de Friesen et Stent corrige la seconde discordance (en ajoutant le contrôle négatif du neurone X sur lui-même), mais pas la première. Pour que le modèle soit compatible avec la séquence expérimentale, il suffit de corriger la discordance encore présente dans le modèle (b) de la figure 6, en supprimant le contrôle négatif de Y sur Z. On obtient ainsi le modèle (c) de la figure 6, dont la Figure 7 montre qu'il permet la séquence observée.

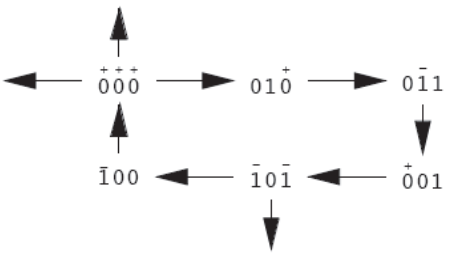


Figure 7: Graphe d'incidence dérivé du schéma de la Figure 6c. On peut voir que ce graphe permet (sans l'imposer) la séquence observée expérimentalement (Figure 5).

8. Davantage sur les délais

Les paramètres logiques et les délais introduisent dans la description logique généralité et flexibilité. Il est essentiel de cerner les rôles respectifs de ces deux éléments.

Les paramètres logiques concernent le *rapport* des constantes cinétiques impliquées dans la production (k) et la dégradation spontanée (λ) d'un

constituant. Quand nous écrivons K = n, cela signifie que $k/\lambda > s^n$, c'est-à-dire que le rapport des taux de production et de dégradation spontanée d'une variable excède de manière significative le n^e seuil dans l'échelle de discrétisation de cette variable.

Dans la description différentielle, ce même rapport apparaît dans les équations d'état stationnaire et tout ce qui en dérive, à savoir les nullclines et leurs intersections, qui définissent les états stationnaires. Si on multiplie toutes les constantes cinétiques par un même facteur dans l'une des équations différentielles, ni les équations d'état stationnaire, ni le tracé des nullclines, ni la localisation des états stationnaires du système ne sont modifiés. Par contre, la stabilité des états stationnaires et, par conséquent, toute la dynamique, peuvent être modifiés, car elles dépendent non pas du rapport, mais des *valeurs individuelles* des paramètres.

Il en résulte qu'en l'absence d'information supplémentaire, nos équations logiques nous informent sur le nombre et la localisation des états stationnaires, mais pas toujours sur leur stabilité. En particulier, en présence d'un circuit négatif fonctionnel, l'analyse logique identifie un cycle et en infère une périodicité, mais elle ne dit pas si un foyer est stable (périodicité amortie) ou instable (périodicité stable).

A l'opposé des paramètres logiques, les délais dépendent des valeurs individuelles de paramètres, et pas seulement de leur rapport [40]. Ceci explique pourquoi les délais nous apportent une information qui n'est pas déjà contenue dans les paramètres logiques.

Cycles logiques

En général, les graphes de transition fournis par l'analyse logique d'un système ne sont pas de simples arbres: ils contiennent fréquemment des cycles. Les méthodes décrites dans la section 2 permettent d'analyser les conditions du choix entre deux ou plusieurs séquences d'états et, en particulier, les conditions pour entrer dans un cycle et y rester ou non. Pour une description détaillé, voir [39, 41, 43].

Considérons un état logique différent de son image et, par conséquent, sujet à une commande de changement de valeur d'une ou plusieurs variables. Comme il a été mentionné dans la section 2, un tel ordre peut avoir été donné 1, 2, … n états en amont. Ce point doit être pris en considération pour le calcul des contraintes exercées par les délais sur le chemin effectivement suivi par le système.

Il est souvent commode d'"étiqueter" les variables qui ont reçu une commande de commutation en amont par un souscrit 1, 2, … n, comme dans la Figure3. Cette notation est particulièrement utile pour distinguer les "faux" cycles (dans lesquels la valeur de l'un des souscrits change chaque fois que l'état nominal considéré se reproduit) et les cycles vrais, dans

lesquels chaque état retrouve invariablement les mêmes souscrits. Nous appelons "cycle logique stable" un cycle qui, une fois entamé, persiste pour une *gamme* de valeurs des délais; certains cycles persistent même quels que soient les délais. Les conditions de maintien de ces cycles sont un ensemble d'*inégalités* entre délais ou sommes de délais. Par opposition, il existe des cycles logiques dont la persistance requiert une ou plusieurs *égalités* entre délais (ou sommes de délais). Nous les appelons "cycles logiques instables". Ils peuvent être immédiatement identifiés par le fait que chaque état d'un tel cycle porte au moins un souscrit [41, 43].

Notons que la relation entre cycles logiques stables ou instables et les cycles limite stables ou instables de la description différentielle n'est pas simple.

Chaos dans la description logique

A première vue, un automate fini ne peut décrire une dynamique chaotique parce que le nombre des états est fini. Cependant, dans notre description logique le temps est continu et un état nominal donné peut représenter un nombre illimité de variants selon les valeurs des restes de commutation des délais. Un exemple à quatre variables comprenant deux 3-circuits de signes opposés est décrit dans [41]. Pour des valeurs appropriées des délais, la description logique de ce système révèle une dynamique typiquement chaotique. Le comportement complexe de ce système avait déjà été mentionné (section 2) comme exemple de la différence entre les descriptions logiques asynchrone et synchrone.

9. Discussion

L'un des intérêts majeurs de la description logique décrite ici est qu'au lieu de devoir tester une large gamme de valeurs de nombreux paramètres, nous considérons des paramètres logiques dont chacun ne peut revêtir qu'un nombre limité, généralement petit, de valeurs discrètes. Ainsi, les innombrables combinaisons de valeurs des paramètres réels sont remplacées par un ensemble limité de combinaisons de valeurs des paramètres logiques. L'important est que chaque situation dynamique qualitativement significative occupe un ou plusieurs pavés dans l'espace des paramètres logiques. Ceci permet une vue complète des comportements qualitatifs d'un modèle. En fait, il n'est le plus souvent pas nécessaire de balayer toutes les combinaisons de valeurs des paramètres logiques. En pratique, il est commode de commencer par identifier l'état caractéristique de chaque circuit ou union de circuits disjoints, puis de calculer les gammes de valeurs des paramètres logiques pour lesquelles chacun de ces circuits est fonctionnel ou non. La procédure est entièrement informatisée [11, 28, 36].

9. 1. Les idéalisations impliquées dans notre description logique

9. 1. 1. L'assimilation des fonctions sigmoïdes à des fonctions à seuil

L'idéalisation la plus visible de la description logique consiste à traiter toutes les interactions en termes de fonctions à seuil. En pratique, pourvu que les interactions réelles soient des sigmoïdes un peu raides (ce qui est fréquemment le cas) l'essentiel de la dynamique est conservée si l'on caricature la sigmoïde par une fonction à seuil [8, 9, 13, 16 17].

En fait, ce qui importe, c'est moins la pente des interactions sigmoïdes individuelles d'un circuit que celle de la sigmoïde résultant de la composition de ces sigmoïdes individuelles. Comme suggéré il y a longtemps [43] et démontré depuis (Reigner et al., en préparation), la composition de sigmoïdes résulte en une sigmoïde croissante ou décroissante selon qu'elle comprend un nombre pair ou impair de sigmoïdes décroissantes. La pente de la sigmoïde résultante croît avec le nombre de sigmoïdes impliquées dans la composition. En pratique, un circuit peut être décrit de manière appropriée en termes de fonctions à seuil même si certaines des interactions sont linéaires, pourvu que la pente de la fonction composée soit suffisante.

Dès que les combinaisons intéressantes de valeurs des paramètres logiques ont été établies, on peut, si on le désire, revenir à une description différentielle utilisant des fonctions sigmoïdes ou seuil. Dans ce dernier cas, la concordance entre la description logique généralisée et la description différentielle est excellente. Si on adopte des sigmoïdes de pente modérée, les relations entre les constantes cinétiques et les valeurs de seuils qui déterminent la fonctionnalité d'un n-circuit doivent être adaptées en conséquence (voir, par exemple Kaufman, dans [43]).

9. 1. 2. Expression de fonctions logiques comme sommes arithmétiques de termes impliquant chacun une seule variable

Tant que nous utilisons la description logique "naïve", toute fonction logique, qu'il s'agisse de la somme logique (OU inclusif), du produit logique (ET), du OU exclusif, etc., peut être utilisée sans problème. Par contre, dans les équations de Snoussi, chaque fonction est décrite comme une somme arithmétique de termes. L'emploi de valeurs appropriées des paramètres permet de reproduire une série de fonctions logiques classiques, y compris le "ET", et aussi de rendre compte de situations non décrites par les fonctions logiques classiques. Cependant, certaines fonctions, comme le OU exclusif, ne peuvent être décrites de cette façon. En fait, rien n'empêche d'utiliser ces fonctions logiques dans des équations à la Snoussi. Encore faudrait-il s'assurer que les théorèmes existants s'appliquent à ces situations "non-linéaires". Finalement, il faut aussi savoir que même des affirmations aussi raisonnables que "K_{12} ne peut être inférieur ni à K_1 ni à K_2" trouvent des contre-exemples en biologie.

9. 1. 3. La règle du jeu pour la définition des délais

La transition des équations logiques à la table des états et de celle-ci au graphe des séquences d'états n'implique pas d'approximations. Par contre, l'analyse logique "au second degré" basée sur les inégalités entre sommes de délais implique le choix d'une règle du jeu. Nous avons choisi la règle suivante: la commande de changer la valeur d'une variable est exécutée après un délai caractéristique de la transition, à moins que la commande ne soit annulée avant d'avoir été exécutée; et si cette annulation survient, nous raisonnons comme si la commande n'avait pas eu lieu[5]. Cette règle du jeu a l'avantage d'une extrême simplicité, mais il est clair que l'idéalisation ne sera pas appropriée dans le cas d'une succession rapide de commandes et de contre-ordres: quand un gène synthétise de manière répétée une quantité significative mais insuffisante de son produit, et que le taux de dégradation de celui-ci est faible, ce produit finira par atteindre son niveau fonctionnel, en désaccord avec l'application de la "règle du jeu".

9.2. Un point de vue stochastique

L'analyse logique "au second degré" basée sur les inégalités entre sommes de délais permet d'approcher de manière rationnelle le problème du choix entre les chemins possibles. Si on ne dit rien des délais, tout le graphe des transitions reste ouvert. Si, à l'opposé, on définit entièrement les valeurs des délais, un seul chemin, bien défini, peut être suivi. La réalité biologique diffère de ces deux situations extrêmes. Dans une population cellulaire, même homogène, un délai donné n'aura pas une valeur exactement déterminée, mais plutôt une valeur moyenne et une distribution. Ceci introduit dans la description logique un élément stochastique: chaque fois que c'est nécessaire, on peut donner une description logique des cellules individuelles et réaliser de nombreuses simulations en tenant compte de la variabilité individuelles des délais. Dans un tel cas, si l'on ignore momentanément les interactions entre cellules, on peut se forger une vue du comportement global de la population par sommation des comportements individuels. Cette approche a été utilisée pour l'analyse des processus d'intégration et d'excision des bactériophages tempérés [39].

---

[5] Par exemple, dans le cas du système simplifié qui modélise l'enclenchement ou non du gène cI du bactériophage lambda, à l'état initial $\overset{+}{0}\,0\,\overset{+}{0}\,0$ succède l'état 100 pourvu que le délai t_x soit inférieur au délai t_y; et comme l'état 100 est stable, cela implique que l'exécution de l'ordre de passage de x = 0 à x = 1 a annulé la commande de passage de y = 0 à y = 1.

Une remarque finale sur le choix entre les descriptions logique et différentielle. On peut avoir le sentiment intuitif que la description différentielle est proche de la réalité biologique et que la seconde n'est rien d'autre qu'une caricature de la première. Ce n'est certainement pas toujours le cas. Considérons, par exemple, une population bactérienne suspendue dans un milieu nutritif. Classiquement, la description différentielle traite un tel système comme une suspension homogène des constituants des bactéries. Dans cette suspension idéale, chaque espèce moléculaire est présente à raison d'un grand nombre de copies, en sorte que les lois des grands nombres, y compris la loi d'action de masse, sont d'application. Cependant, dans la réalité, on se trouve en présence d'un grand nombre de très petites boites (les bactéries individuelles) au sein de chacune desquelles certaines des espèces moléculaires sont présentes en un très petit nombre de copies, voire une seule, voire même aucune. De ce point de vue, le cas le plus frappant est celui de la molécule d' ADN, qui, typiquement, n'est présente qu'à un exemplaire dans chaque cellule bactérienne. Clairement, les lois classiques de la cinétique chimique s'appliquent difficilement telles quelles à ces situations. De plus, à moins que l'on n'y introduise des délais, les équations différentielles sont habituellement telles qu'un produit - par exemple, une protéine - est supposé apparaître sans délai aussitôt que le gène correspondant a été allumé. En fait, il faut un certain temps, dépendant notamment de la longueur du message, avant que la première molécule de l'ARN correspondant soit complétée, et quelques minutes de plus avant que la toute première molécule de la protéine codée par le gène soit achevée. Dans de telles situations, une description logique peut être extrêmement utile. Comme cette description n'est pas tenue de respecter les lois de la cinétique chimique en milieu homogène, on peut l'utiliser pour décrire d'abord la situation dans les cellules individuelles. Des considérations stochastiques sur les délais rendront alors compte des différents chemins qui peuvent être suivis dans différentes cellules individuelles, menant ainsi à différentes situations de régime.
Voici, pour terminer, les références de quelques travaux, biologiques ou autres, qui utilisent notre méthode logique et ne sont pas cités dans le corps de ce travail: [1-3, 13, 17, 18, 22-25, 31, 37, 44, 47].

Références

[1] Boon, F. et de Palma, A., Boolean formalism and urban development, Lecture Notes in Biomathematics 29, 402-439 (1979).
[2] Dee, D. et Ghil, M., Boolean difference equations, I: formulation and dynamic behaviour, Siam J. Appl. Math., 44, 11-126 (1984).
[3] Demongeot, J., R. Thomas, R. et Thellier, M., A mathematical model for storage and recall functions in plants, C. R. Acad. Sci, Sciences de la vie, 323, 93 (2000)

[4] Eisenfeld, J. et de Lisi, C. "On conditions for qualitative instability of regulatory circuits with application to immunological control loops", In Mathematics and computers in biomedical applications, Eisenfeld, J. and de Lisi, C. Eds. Elsevier, 39-53 (1985).
[5] Florine, J. La synthèse des machines numériques et son automatisation, Dunod, Paris (1974).
[6] Friesen, W. O. et Stent, G. S., Biol. Cybernet. 28, 40 (1977).
[7] Glass, L., Combinatorial and topological methods in nonlinear chemical kinetics, J. Chem. Phys. 63, 1325-1335 (1975).
[8] Glass, L. et Kauffman, S. A., Co-operative components, spatial localization and oscillatory cellular dynamics, J. Theor. Biol. 34, 219-237 (1972).
[9] Glass, L. et Kauffman, S. A., The logical analysis of continuous, non-linear biochemical control networks, J. Theor. Biol., 39, 103-129 (1973).
[10] Glass, L. et Pasternack, J. S., Stable oscillations in mathematical models of biological control systems, J. Math. Biology. 6, 207-223 (1978).
[11] Gonzales, A., Sánchez, L., Thieffry, D. et Chaouiya, C., GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks, BioSystems, 84, 91-100 (2006); http://gin.univ-mrs.fr/GINsim.
[12] Kauffmann, S. A., J. , Metabolic stability and epigenesis in randomly constructed genetic nets, J. Theoret. Biol. 22, 437-467 (1969).
[13] Kaufman, M. Role of multistationarity in an immune response model: a combined discrete and continuous approach, in Theoretical Immunology, A. S. Perelson, ed., Addison Wesley p. 199-222 (1988).
[14] Kaufman, M., Andris, F. et Leo, O., A logical analysis of T cell activation and anergy, Proc. Natl. Acad. Sci. USA, 96, 3894-3899 (1999).
[15] Kaufman, M., Soulé, C. et Thomas, R., A new necessary condition on interaction graphs for multistationarity, J. Theor. Biol. 248, 675-685 (2007).
[16] Kaufman, M. et Thomas, R., Model Analysis of the bases of multistationarity in the humoral immune response, J. Theor. Biol., 129, 141-162 (1987).
[17] Kaufman, M., Urbain, J. et Thomas, R., Towards a logical analysis of the immune response, J. Theor. Biol. 114, 527-561 (1985).
[18] Leclercq, J. et Dumont, J., Boolean analysis of cell regulation networks, J. Theor. Biol. 104, 507-534 (1983).
[19] Leussler, A. et Van Ham, P., Combinatorial systems, Lecture Notes in Biomathematics 29, 62-85 (1979).
[20] McAdams, H. H. et Arkin, A., Stochastic mechanisms in gene expression, Proc. Natl. Acad. Sci. USA, 94, 814 (1997).
[21] Maynard, D. M. et Silverston, A. I., J. Comp. Physiol., 100, 161-162 (1975).

[22] Mendoza, L., Thieffry, D. et Alvarez-Buylla, E.R., Genetic control of flower morphogenesis in Arabidopsis thaliana: a logical analysis, Bioinformatics, 15, 593 (1999).
[23] Nicolis, C., A Boolean approach to climate dynamics, Quart. J. R. Met. Soc. 108, 707-715 (1982).
[24] Nicolis, G., Mathematical aspects of biological regulatory processes, Lecture Notes in Biomathematics, 29, 178-211 (1979).
[25] Pastoret, P. P., Thiry, E. et Thomas, R., Logical description of Bovine Herpesvirus type I latent infection, J. Gen. Virol. 67, 885-897 (1986).
[26] Rajewsky, N., Mathematical Biophysics, The University of Chicago Press (1948).
[27] Ratner, V. A., Molekulargenetische Steuerungssystem, Academie-Verlag, Berlin (1977).
[28] Richard, A., Comet, J.-P. et Bernot G., smbionet : Selection of Models of Biological Networks, http://smbionet.lami.univ-evry.fr.
[29] Richelle, J., Analyse booléenne et continue de systèmes comportant des boucles de rétroaction. I. Analyse continue de boucles de rétroaction simples., Bull. Cl. Sci. Acad Roy. Belg. 66, 890-912 (1980).
[30] Richelle, J., Multilevel logic as a tool in modelling: application to glycolytic reactions, in Mathematical Modelling in Science and Technology, Avula et al. eds., Pergamon (1986).
[31]Sanchez, l., van Helden, J. et Thieffry D., Establishement of the dorso-ventral pattern during embryonic development of drosophila melanogasater: a logical analysis, J Theor Biol. 189, 377-389 (1997).
[32] Slack, J. M. W., Serial thresholds and homeotic variation, Lecture Notes in Biomathematics, 29, 473-480 (1979).
[33] Snoussi, E. H., Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach, Dynamics and Stability of Systems, 4, 189-207 (1989).
[34] Snoussi, E. H. et Thomas, R., Logical identification of all steady states: the concept of feedback loop characteristic states, Bull. Math. Biol. 55,973-991 (1993).
[35] Sugita, M., Functional analysis of chemical systems *in vivo* using a logical circuit equivalent. II The idea of a molecular automaton, J. Theoret. Biol. 4, 179-192 (1963).
[36] Thieffry, D., Colet, M. et Thomas, R., Formalisation of regulatory networks: a logical method, and its automatization, Math. Modelling Sci. Computing, 2, 144-151 (1993).
[37] Thieffry, D. et Thomas, R., Dynamical behaviour of biological regulatory networks II. Immunity control in bacteriophage lambda, Bull. Math. Biol., 57, 277 (1995).
[38] Thomas, R., Boolean formalization of genetic control circuits, J. Theor. Biol. 42, 563-585 (1973).

[39] Thomas, R., Some biological examples, Lecture Notes in Biomathematics, 29, 354-401 (1979).
[40] Thomas, R., Logical vs continuous description of systems comprising feedback loops:the relation between time delays and parameters, chemical applications of topology and graph theory, R. B. King, ed., Elsevier, pp.307-321 (1983).
[41] Thomas, R., Logical description, analysis and synthesis of biological and other networks comprising feedback loops, Adv. Chem. Phys. 55, 247-282 (1984).
[42] Thomas, R., Regulatory networks seen as asynchronous automata: A logical description, J. Theor. Biol.,153, 1-23 (1991).
[43] Thomas, R. et D'Ari, R., Biological Feedback, C.R.C. Press, Boca Raton, Florida., 316 pp., (1990).
[44] Thomas, R., Gathoye, A.-M. et Lambert, L., A complex control circuit. regulation of immunity in temperate bacteriophages, Eur. J. Biochem. 71, 211-227 (1976).
[45] Thomas, R. et Kaufman, M., Multistationarity, the basis of cell differentiation and memory I. Structural conditions of multistationarity and other nontrivial behavior, et II. Logical Analysis of Regulatory Networks in Terms of Feedback Circuits, Chaos 11, 170-179 et 180-195 (2001).
[46] Thomas, R., Thieffry, D. et Kaufman, M., 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, 247-276 (1995).
[47]Thomas, R. et Van Ham, P., Analyse formelle de circuits de régulation génétique: contrôle de l'immunité chez les bactériophages lambdoïdes, Biochimie, 56, 1529-1547 (1974).
[48] Van Ham, P., How to deal with variables with more than two levels, Lecture notes in Biomathematics, 29, 326-343 (1979).
[49] Wolpert, L. et Lewis, J. H., Towards a theory of development, Fed. Proc. 34,14-20 (1975).

Programming with models

Jeremy Gunawardena[1]

[1]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue Boston, MA 02115, USA

Abstract

Mathematical models are increasingly used in systems biology to understand how phenotypes emerge from the collective interactions of molecular and cellular components. While several methodologies have been put forward for creating and analysing models, they largely reflect the perspectives of the physical sciences (mathematics, physics, engineering, computer science). Integrating model building into a programme of experimental biology suggests that a different perspective on models may be needed. They are better seen as formalisations of biological knowledge rather than as descriptions of some "external" reality. Accordingly, models need to be built in such a way that their biological assumptions can be readily altered in response to new knowledge or new data or simply as a way to explore alternative hypotheses. Among other things, mechanisms for modularity and for abstraction become essential. Although these are well-understood in engineering and computer science, their implementation in systems biology imposes interesting new challenges. We have developed a prototype computational infrastructure which supports such a flexible methodology for biological model building.

Computer-based analysis of bacterial chemotaxis

Dennis Bray[1]

[1] Physiology, Development and Neuroscience, Downing Street, Cambridge CB2 3DY, UK

Abstract

The set of biochemical reactions by which an *E. coli* bacterium detects and responds to distant sources of attractant or repellent molecules is probably the simplest and best understood example of a cell signalling pathway. The pathway has been saturated genetically and all of its protein components have been isolated, measured biochemically, and their atomic structures determined. We are using detailed computer simulations, tied to experimental data, to ask how the pathway works as an integrated unit. Increasingly we find that the physical location of molecular components within the molecular jungle of the cell interior is crucial for an understanding of their function. Signal amplification, for example, appears to depend on the propagation of activity across clusters of receptors and associated molecules. By combining a molecularly detailed ODE model of the chemotaxis pathway with graphical displays of bacterial swimming we create ”surrogate” bacteria that can be treated as though experimental organisms. Reponses of these virtual bacteria to twodimensional gradients of attractants reveals novel features of chemotactic behavior that would be difficult if not impossible to detect experimentally.

For more information please see: **`www.pdn.cam.ac.uk/compcell`**

Tensegrity-based integration of cellular structural and information processing networks

Donald E. Ingber[1]

[1] Departments of Pathology and Surgery, Harvard Medical School and Children's Hospital, Boston, MA, USA

Abstract

This lecture will focus on the mechanism by which mechanical forces and chemical signals interplay to control how individual cells decide whether to grow, differentiate, move, or die, and thereby direct pattern formation during tissue development. Pursuit of this challenge has required development and application of new nano- and micro-technologies, as well as theoretical formulations, computational models and bioinformatics tools. These approaches have been used to apply controlled mechanical stresses to specific cell surface molecules and to measure mechanical and biochemical responses; to control cell shape independently of chemical factors; and to handle the structural, hierarchical and informational complexity of living cells. Results of these studies have changed our view of how cells and tissues control their shape and mechanical properties by demonstrating that cells use tensegrity architecture to structure themselves at the molecular level, as well as at larger size scales. They also have led to the discovery that integrins and the cytoskeleton play a central role in cellular mechanotransduction, and that cell fates represent multi-dimensional 'attractor' states of the genome-wide gene regulatory network. Recognition of these critical links between mechanics, cellular biochemistry and genomic information processing should lead to novel strategies for development of therapeutics, as well as the design and fabrication of novel biomimetic microdevices and nanotechnologies that can function effectively within the context of living tissues.

References

[1] Wang N, Butler JP, Ingber DE. Mechanotransduction across the cell surface and through the cytoskeleton. *Science* 1993; **260**:1124-1127.

[2] Chen CS, Mrksich M, Huang S, Whitesides G, Ingber DE. Geometric control of cell life and death. *Science* 1997; **276**:1425-1428.

[3] Ingber, DE. The Architecture of Life. *Scientific American* Jan 1998; **278**:48-57.

[4] Ingber DE. Cellular tensegrity revisited I. Cell structure and hierarchical systems biology. *J. Cell Sci.* 2003; **116**: 1157-1173.

[5] Ingber DE. Tensegrity II. How structural networks influence cellular information processing networks. *J. Cell Sci.* 2003; **116**: 1397-1408.

[6] Huang S, Eichler G, Bar-Yam Y, Ingber DE. Cell fates as attractors in gene expression state space. *Phys. Rev. Lett.* 2005; **94**: 128701-12802.

[7] Ingber DE. Cellular mechanotransduction: putting all the pieces together again. *FASEB J.* 2006 **20**: 811-827.

[8] Mannix RJ, Kumar S, Cassiola F, Montoya-Zavala M and Ingber DE. Magnetic actuation of receptor-mediated signal transduction. *Nature Nanotech* **3**, 36-40 (2008) Published online: 23 December 2007.

Shape, Polarity and Division Axis in Animal Adherent Cells

Michel Bornens[1]

[1]UMR144, CNRS, Institut Curie, 26 rue d'Ulm 75248 Paris Cedex 05, France

Abstract

The molecular role of cell-cell or cell-extracellular matrix (ECM) contacts on the polarity of epithelial cell has been well characterized. How the spatial distribution of extracellular environment affects cell asymmetry is less studied. Micro-patterned substrates imposing cells to spread on various combinations of adhesive and non-adhesive areas have previously been used to control the orientation of cell division and to demonstrate that cells can generate identical total traction force on the substrate but that individual stress fibers are differently spaced depending on the spatial distribution of its adherence contacts. The reproducible effect on overall cell compartmentalization enabled the quantification of the spatial organization of intracellular compartments. Analyzing the organization of individual cells plated on different micropatterns, one can show that ECM distribution can predictably modulate cell asymmetry and polarity axes [3]. The growth of MTs appears modulated by the asymmetric composition of the cortex. The respective positions of the nucleus and the centrosome-Golgi apparatus indicate that the internal cell polarity of these non-migrating cells is in register with the polarity axis of the adhesive environment. Interestingly the cortical asymmetry did not affect the centrosome positioning at the cell centroid. Thus in addition to ECM molecular composition and mechanical properties, ECM geometry could play a key role in developmental processes. We will argue that the shape of adherent cells, which reflects a tensional force field, undertakes a continuous transformation through the cell division process compatible with the maintenance of tissue integrity during growth or turn over [4].

We have further showed that the observed spindle orientation in adherent cells can be understood as the result of the action of cortical force generators which interact with spindle microtubules [5]. We assumed that these force generators are activated by cortical signals which are associated with retraction fibers that connect the round cell body to the cell adhesive microenvironment. The force generators exert a torque on the spindle which depends on its relative orientation with respect to the microenvironment geometry and leads to spindle rotation. We developed a simple physical description of this spindle mechanics which allows us to calculate angular profiles of the torque acting on the spindle, as well as the angular distribution of spindle orientations. Our model could account for the spindle orientations observed

in a large variety of different adhesive patterns. Remarkably it also describes the sharp transition from symmetrical to asymmetrical orientation observed for two patterns of similar type. Our results suggest that a slight modification of cell microenvironment is sufficient to provide either two main polarizing signals and stimulate symmetrical orientation or a single polarizing signal and stimulate asymmetrical orientation of the spindle. It supports the hypothesis that, in vivo, niche architecture is established so as to embed stem cell is in a geometrical configuration close to this transition.

References

[1] Théry M, Racine V, Pepin A, Piel M, Chen Y, Sibarita JB, Bornens M. The extracellular matrix guides the orientation of the cell division axis. *Nat Cell Biol.* 2005 Oct;7(**10**):947-53. Epub 2005 Sep 18.

[2] Théry M, Pepin A, Dressaire E, Chen Y, Bornens M. Cell distribution of stress fibres in response to the geometry of the adhesive environment. *Cell Motil Cytoskeleton*, **63**: 341-355, (2006)

[3] Théry M, Racine V, Piel M, Pepin A, Dimitrov A, Chen Y, Sibarita Jb, Bornens M. Anisotropy of cell adhesive microenvironment governs cell internal organization and orientation of polarity. *Proc Natl Acad Sci USA*. **103**:19771-6, (2006). Epub 2006 Dec 18.

[4] Théry M, Bornens M. 2006. Cell shape and cell division. *Curr Opin Cell Biol* **18**(6):648-57

[5] Théry M, Jiménez-Dalmaroni A, Racine V, Bornens M, Jülicher F Experimental and Theoretical Study of Mitotic Spindle Orientation *Nature*, **447**, 493-496, (2007).

Genetic analysis of electric signal-directed cell movement

Min Zhao[1]

[1]University of California Davis, School of Medicine, 1515 Newton Ct., Davis CA, USA

Abstract

Directional cell migration plays an important role in development, wound healing and inflammation. We provide experimental evidence that endogenous electric fields may serve as an important signaling system in guiding cell migration. The underlying mechanisms are starting to emerge.

Modern techniques have been used to confirm and extend the understanding of the existence of endogenous electric fields associated with wounds. Such type of electric fields was first detected more than 150 years ago by Du Bois Reymond. At sites of wounds and regeneration, small dc electric fields are readily detectable.

Many types of cells, including human cells respond to an electric signal by directional migration, a phenomenon called electrotaxis or galvanotaxis. More importantly, recent experiments have demonstrated a predominant role for the electric signals in guiding cell migration in epithelial wound healing. When applied in the presence of other well accepted guidance cues, such as contact inhibition release, wound void etc., electric fields of physiological strength override those cues in directing cell polarization and migration.

Some elements of the molecular machinery are being identified using transgenic mice. We showed that PI3 kinase / Pten play important roles in electric field directed cell migration. A genetic tractable model organism– *Dictyostelium discoideum* is been used to elucidate the molecular genetics of electrotaxis. With novel screen strategies, it is possible to glean into the complex signaling mechanisms for cells to sense and respond to an electric signal.

Rule-based modelling of cellular signalling

Vincent Danos[1]

[1]Équipe Preuves, Programmes, Systèmes, CNRS UMR 7126, Université Denis Diderot, Case 7014, 75205 PARIS Cedex 13

Abstract

Modelling is becoming a necessity in studying biological signalling pathways, because the combinatorial complexity of such systems rapidly overwhelms intuitive and qualitative forms of reasoning. Yet, this same combinatorial explosion makes the traditional modelling paradigm based on systems of differential equations impractical. In contrast, agent-based or concurrent languages, such as Kappa or the closely related BioNetGen language describe biological interactions in terms of rules, thereby avoiding the combinatorial explosion besetting differential equations. Rules are expressed in an intuitive graphical form that transparently represents biological knowledge. In this way, rules become a natural unit of model building, modification, and discussion. We illustrate this with a sizeable example obtained from refactoring two models of EGF receptor signalling that are based on differential equations. An exciting aspect of the agent-based approach is that it naturally lends itself to the identification and analysis of the causal structures that deeply shape the dynamical, and perhaps even evolutionary, characteristics of complex distributed biological systems. In particular, one can adapt the notions of causality and conflict, familiar from concurrency theory, to Kappa, our representation language of choice. Using the EGF receptor model as an example, we show how causality enables the formalization of the colloquial concept of pathway and, perhaps more surprisingly, how conflict can be used to dissect the signalling dynamics to obtain a qualitative handle on the range of system behaviours. By taming the combinatorial explosion, and exposing the causal structures and key kinetic junctures in a model, agent- and rule-based representations hold promise for making modelling more powerful, more perspicuous, and of appeal to a wider audience.

Constraint-based Modeling of Human Metabolism

Eytan Ruppin[1]

[1]School of Computer Science & School of Medicine Tel Aviv University, Tel Aviv, Israel

Abstract

My presentation will consist of two parts. First, I shall provide a general overview of the paradigm of constraint-based modeling of metabolism, focusing on flux-balance analysis (FBA) of the metabolic networks of microorganisms, describing what's it's about, what has been achieved, and the challenges that lie ahead. As times permits, I shall very briefly describe a few contributions from my lab to this just endeavor. The second, research part of my talk, will focus on describing recent work from my lab that has constructed a large-scale, tissue-specific description of human metabolism. Our approach is based on integrating tissue-specific gene and protein expression data with a comprehensive existing reconstruction of the global human metabolic network. Applying our method to predict tissue-specific metabolic activity for 10 human tissues (including the liver, kidney, heart and brain) has revealed the central role of post-transcriptional regulation in shaping tissue-specific metabolic activity profiles. The predicted tissue specificity of metabolic disease-causing genes and of metabolite exchange with biofluids are shown to go markedly beyond that manifested solely in gene expression data, and are validated via large-scale mining of tissue-specificity data. Overall, these results lay down a computational basis for the genome-wide study of normal and abnormal human metabolism in a tissue-specific manner.

Metabolic Control Theory and the analysis of biological regulation

Daniel Kahn[1]

[1]INRIA HELIX project-team, Laboratoire de Biométrie et Biologie Évolutive, Université Lyon 1, France

Abstract

Metabolic Control Theory provides a formalised framework for dissecting the system-level behaviour of an arbitrarily complex metabolic network, as a function of individual enzyme properties [1, 2]. The talk will introduce the main concepts and results of Metabolic Control Theory and discuss its relevance for the analysis of biological regulation.

In particular I will review different ways that have been proposed to decompose complex regulatory systems into modules, what has been termed 'hierarchical control analysis'. Such a formalism can describe how the control of a metabolic steady-state is distributed between the regulation of gene expression and enzyme activity. Regulation coefficients were defined and could be measured in some instances in order to quantify the relative importance of these levels [3]. I will also review a successful reverse engineering approach that has been used to deduce the topology of a regulatory network through systematic external perturbations [4].

Finally I will give an outline of a project that we have initiated towards understanding the interrelation between gene regulation and metabolic regulation. Indeed gene regulation is affected by metabolite concentrations and, conversely, changes in gene expression affect metabolic activities. Therefore a genuine understanding of cellular regulation requires the embedding of gene regulatory networks in cellular metabolism. A quantitative model combining metabolic and gene regulation is needed to provide insight into the biological rationale underlying the distribution of regulation between these two levels.

References

[1] Fell D. (1997) Understanding the control of metabolism. Portland Press.

[2] Heinrich R. and Schuster S. (1996) The Regulation of Cellular Systems. Chapman & Hall.

[3] Rossell S. *et al.* (2006) Unraveling the complexity of flux regulation: a new method demonstrated for nutrient starvation in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. USA* **103**:2166-2171.

[4] Kholodenko B.N. et al. (2002) Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci. USA* **99**:12841-12846.

Stable isotope tracing in metabolic pathways: linkage to the genome and proteome

Andrew N. Lane[1]

[1]JG Brown Cancer Center, Department of Chemistry and Center for Regulatory Environmental Analytical Metabolomics, University of Louisville, KY, USA

Abstract

Mammals and other multicellular organisms pose numerous experimental and computational problems fro understanding cellular homeostasis. Our focus has been on the functional biochemistry of cells under defined conditions, and how they respond to environmental changes such as nutrient supply, hypoxia and toxic agents, and their interactions with genetic lesions associated with cancer [1, 2]. In particular, we have been using cancer cells in monoculture to define their biochemical phenotype, and the repertoire of responses to external perturbation, using different genetic backgrounds. The main approach has been to trace individual atoms through metabolic pathways using stable isotopes monitored with mass spectrometry and NMR [3, 4, 5]. By choosing different labeled precursor molecules, various pathways can be traced at the atomic level, and thus how the metabolic fluxes alter in response to changes in demand. The challenge is to make use of the metabolic information to make sense of gene expression data and alterations in enzyme activities, which may reflect changes in concentration (e.g. expression), changes in post-translational modification, or via alterations in allosteric regulation. Once the repertoire of responses of cells in monoculture are understood, it is possible to search for similar detailed metabolic flux profiles of such cells in tissues of model organisms, or even in a clinical setting.

The experimental approaches will be briefly outlined, and illustrated with specific examples from breast and lung cancer metabolomics coupled with gene expression analysis and cell physiology. Approaches to modeling and computational (informatics) requirements will be treated with some examples of new programs developed in our laboratories to deal with isotopomer distributions in complex mixtures.

References

[1] Fan, T. W-M.,,Bandura, L.L. Lane, A.N., & Higashi, R.M., (2006) " Integrating Genomics and Metabolomics for Probing Se Anticancer Mechanisms" *Drug Metabolism Reviews* **38**, 1-25

[2] Telang S. Yalcin A., Clem A. L.,Bucala R., Lane A. N., Eaton, J. W., & Chesney J. A. (2006) Ras Transformation Requires Metabolic Control by 6-Phosphofructo-2-Kinase. *Oncogene* **25**: 7225-7234

[3] Lane, A.N. & Fan, T. W-M. (2007) Determination of positional isotopomers in metabolites. *Metabolomics* **3**, 79-86 (On line March 15 2007)

[4] Lane, A.N., Fan, T-WM. & Higashi, R.M. (2007) "Isotopomer-based metabolomic analysis by NMR and mass spectrometry". Methods in Cell Biology, Vol 84. *Biophysical Tools for Biologists, Volume 1*, (John J. Correia and H. William Detrich, Eds). Elsevier Science & Technology Books Ch. 18. pp 541-588.

[5] Fan, T.W-M. & Lane, A.N. (2007) Structure-based profiling of Metabolites and Isotopomers by NMR, *Prog. NMR Spectrosc.* In Press. doi:10.1016/j.pnmrs.2007.03.002

Metabolomics and Systems Biology: from basics to basic science

Marc-Emmanuel Dumas[1]

[1]Centre de RMN à Très Hauts Champs (FRE 3008 CNRS / ENS-Lyon / UCBL1), Ecole Normale Supérieure de Lyon, 69100 Villeurbanne, France

Abstract

The study of human multifactorial diseases like insulin resistance, or complex biological processes such as ageing, represent a real healthcare challenge for the western and developing world [1]. In this regard, high-throughput "omics" biotechnologies like genomics, transcriptomics and metabolomics are invaluable tools for investigating insulin resistance-related (type 2 diabetes, obesity, non-alcoholic fatty liver diseases) and ageing-related pathologies.

In a first introductory part, I will summarize the basis of metabolic profiling and chemometric analysis of spectral data. Metabolic profiling is usually achieved using nuclear magnetic resonance spectroscopy (NMR) or mass spectrometry (MS). I will mainly focus on NMR-based metabolic profiling. The information content in spectra is such that chemometric modeling is compulsory to process and analyse the data.

In a second part, I will describe how animal models of human disease can be used to decipher genetic from environmental variation. Metabolic phenotypes generated by NMR or MS, or metabotypes, can be used as a structuring tool to reveal the latent signature associated with silent mutations in *Caenorhabditis elegans* [2]. Metabolic profiling can also help understanding the effect of microbial metabolism in insulin-resistance [3]. In fact, the integration of metabolic profiles with genome-wide genotyping [4] and expression profiling data extends the spectrum of rodent genetics to the study of transgenomic interactions between the symbiotic gut flora and the mammalian host [5].
Finally, I will present data showing how human metabotypes are influenced by environmental factors [6].

References

[1] Zimmet, P., Alberti, K.G.M.M. & Shaw, J. Global and societal implications of the diabetes epidemic. *Nature* **414**, 782-787 (2001).

[2] Blaise, B.J. et al., Metabotyping of Caenorhabditis elegans reveals latent phenotypes. *Proc Natl Acad Sci USA* **104**, 19808-12 (2007).

[3] Dumas, M.E. et al. Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc Natl Acad Sci USA* **103**, 12511-6 (2006).

[4] Dumas, M.-E. et al. Direct quantitative trait locus mapping of mammalian metabolic phenotypes in diabetic and normoglycemic rat models. *Nat Genet* **39**, 666-672 (2007).

[5] Martin, FPJ, et al., A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. *Molecular Systems Biology* **3**, (2007).

[6] Dumas, M.E. et al. Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. *Anal Chem.* **78**, 2199-208 (2006).

Genome architecture and dynamics studied by computer simulations

Jörg Langowski[1]

[1]Biophysics of Macromolecules, German Cancer Research Center,
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

Abstract

The folding of the genomic DNA in the eukaryotic cell nucleus is one of the major factors determining gene activation in cell differentiation and development. To understand its structure, structural models have to be developed on many different length and time scales, ranging from an atomic-level description of DNA and protein structure to the distribution of interphase chromosomes and the transport of macromolecules between them. Such models have to be developed in close connection with experiments that determine characteristic quantities of the genome network, such as the positions of genomic markers, the flexibility of DNA and chromatin stretches, structures of the constituents (DNA and proteins) and their interactions, and transport properties of biomolecules within the cell nucleus.

I shall present here an introduction to the most important methods and concepts for modelling and simulation of the genome structure. The common thread is multiscaling: as one proceeds from a description of DNA and protein components at the atomic level towards more and more complicated structures (nucleosome chain, chromatin fiber, interphase chromosome), low-level details must be left out and merged into a coarse- grained description.

Examples covered will be:

- the analysis of nucleosome dynamics on a microsecond timescale, using single beads as approximation for amino acids and nucleotides;
- the unrolling of DNA from the histone core using an elastic-chain model for DNA;
- the folding of the 30 nm chromatin fiber and stability analysis of nucleosomal repeats;
- the organization of the interphase chromosome and intranuclear transport of macromolecules, using a discrete-lattice model for the chromatin chain.

Measuring *in vivo* binding affinities by quantitative FRAP

Jim McNally[1]

[1]Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, Bethesda, MD 20892, USA

Abstract

An essential requirement for accurate modeling of cellular networks is knowledge of the binding affinities between interacting components in the network. Our research focuses on the network of interactions that gives rise to transcription. We are studying this process in live cells using GFP-tagged proteins that can be visualized binding to both specific promoter target sites and nonspecific chromatin sites throughout the nucleus. We are developing methods to measure the binding interactions of transcription factors with chromatin. A primary tool thus far has been fluorescence recovery after photobleaching (FRAP). FRAP reveals the dynamics of fluorescently tagged molecules within live cells. These molecular dynamics are governed by diffusion of the molecule and its *in vivo* binding interactions. As a result, quantitative estimates of the association and dissociation rates of binding can be extracted from FRAP. However, there are no benchmarks for these *in vivo* binding estimates, and so it remains uncertain how accurate the estimates are. FRAP analyses of three different transcription factors interacting with chromatin have yielded significantly different estimates of both the binding rates and the number of predicted binding states. We have now found that these discrepancies are not due to fundamental differences among the site-specific transcription factors, but rather arise from errors in FRAP modeling. The two principal errors are a neglect of diffusion's role and an oversimplified approximation of the photobleach profile. Accounting for these errors eliminates most of the previous discrepancies in the binding estimates for the three different transcription factors. These results provide improved quantitative FRAP protocols and further suggest that different transcription factors may exhibit similar *in vivo* interactions with native chromatin.

Methods to improve the kinetic evaluation of fluorescence intensities and locations of chemical signals within single living cells

Howard R. Petty[1,2] and Andrea Clark[1]

[1] Department of Ophthalmology and Visual Science
[2] Department of Microbiology and Immunology
The University of Michigan Medical School Ann Arbor, MI 48105

Abstract

Chemical signaling in living cells is a highly complex process that, in general, remains poorly understood. Although Ca^{2+} has been the subject of thousands of scientific studies, fundamental questions remain concerning its mechanistic properties; for example, how can one ion mediate the great diversity of signals attributed to it? To better understand signal transduction, we have improved microscopic Ca^{2+} detection strategies. In one approach we minimized illumination and detector noise to provide precise kinetic measurements of Ca^{2+} probe ratios. In the second approach, probe movement during excitation and detection are reduced using a flash-lamp as a light source. These experimental methods are illustrated by the physiological processes of neutrophil adhesion and migration.

1 Introduction

Experimental biology is limited principally by the scientific tools available to investigators. For example, cells were only discovered after the optical microscope was built. For many years, microscope experiments were dominated by studies requiring a sample's chemical fixation. On the other hand, biochemical studies of purified proteins necessarily remove the biological entities from their far-from-equilibrium physiological context and thereby remove crucial regulatory elements and pathways from the analysis. For example, purified pyruvate kinase is studied *in vitro* as a kinase, but *in vivo* it is a phosphatase as its product is pyruvate. Consequently, there is a growing appreciation of the fact that live cell and *in vivo* experiments are crucial in understanding complex biological processes as well as exploiting this understanding in drug development. To achieve this, microscope methods are needed to monitor dynamic cell activities in real time.

Microscope-based fluorescence ratio methods are now widely used in experimental biology for both quantitative microfluorometry and ratio imaging [1]. There are many technical advantages of ratioing. Most importantly,

ratioing corrects for variations in cell thickness. Ratioing also corrects for variations in dye concentration, such as those due to compartmentalization and leakage. Ratiometric studies generally have an improved dynamic range because the emission (or excitation) intensity at one wavelength increases while the intensity of the second wavelength decreases. Ratioing compensates for variations in the level of excitation light. Finally, many dyes are available for many different biological signals, such as Ca^{2+} and pH.

2 Experimental Methods and Findings

To obtain the most meaningful data, experiments must be designed to best reflect the expected environmental conditions *in vivo*. As our experiments are intended to reflect the early events in immunologic activation, the key buffer components listed in Table 1, part A were employed. Early in sepsis responses, arterial pH and glucose are elevated. Later stages of sepsis as well as cancer are associated with lactic acid accumulation and therefore lower pH values (which are more likely to be associated with immunosuppression). Pyruvate is a metabolite associated with Ca^{2+} signaling in leukocytes [2], which is included in the buffer to better reflect proper pyruvate levels at the plasma membrane. As pyruvic acid is a small organic anion, its intracellular concentration is expected to fall when cells are *in vivo*, thereby perturbing both metabolic and signaling pathways.

A. Biological

pH 7.6
10 mM glucose
5 mM pyruvate

B. Chemical

Simultaneous Fluo-4 and Fura Red labeling
30 min. washout to remove unmodified probes from cells.
Simultaneous measurement of Fluo-4 and Fura Red intensities

Table 1: Key Buffer and Labeling Conditions of Experiments

Cell labeling was optimized using Fluo-4 and Fura Red, whose fluorescence intensities increase and decrease, respectively with increases in Ca^{2+}. Both excite at 480nm, but fluoresce at different wavelengths in the visible. Compared to other Ca^{2+} probes, these molecules have good quantum yields, excitation spectra that correspond to reasonably bright regions of Hg-Xe and Xe lamps and emission spectra near the optimal wavelengths of many detectors. The disadvantage of these two molecules is that their sensitivity is not good below 500nM [3]. After labeling, a washout period was used (Table 1).

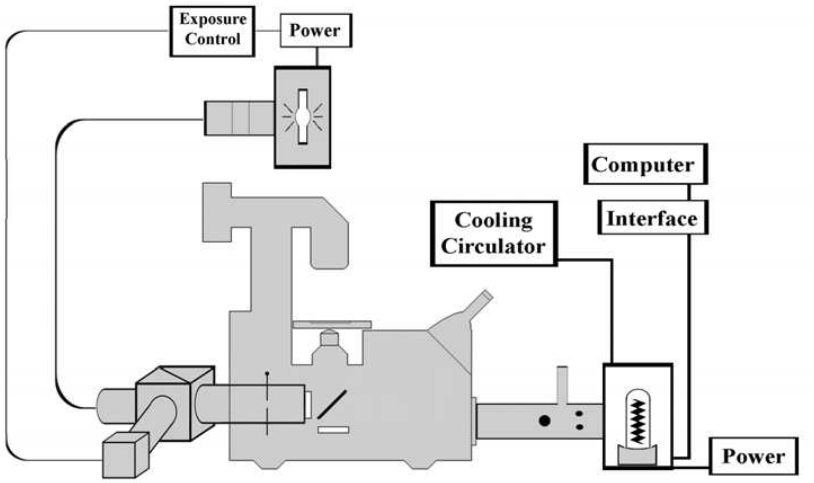


Figure 1: Schematic illustration of the super-quiet microfluorometry system. (For clarity, only one photomultiplier tube (PMT) is shown in this diagram.)

Figure 1 shows an illustration of the microfluorometry apparatus, which has been previously described [4]. The illumination and detection systems dramatically reduce sources of noise in the experimental measurement. Illumination is provided by a Hamamatsu (Bridgewater, NJ) super-quiet 200W mercury-xenon bulb, which displays ultra-low arc fluctuation due to barium-impregnated (BI) electrodes. The bulb was held in a Newport (Irvine, CA) lamp housing and powered by a low noise Newport model 69907 arc lamp power supply. The light was delivered to the microscope via a liquid light guide. As disturbances to the liquid light guide were a source of noise, it was enclosed for most of its length. The lamp was generally ignited about 1 hour before use to allow the system to stabilize. A custom Flash-Cube (Rapp OptoElectronic, Hamburg, Germany) reflected 15% of the broad-spectrum illuminating light onto a thermoelectrically-cooled photodiode, which provided a feedback signal to an Oriel light intensity controller (Newport model 68950). The light intensity controller, in turn, regulated the power supply, thereby closing the control loop. In all cases, light intensity management was performed much faster that data collection. For Ca^{2+} studies, an HQ475/40x excitation filter and a 510 nm dichroic mirror (Chroma Technology Corp., Rockingham, VT) were used in the microscope's filter cube. The highest numerical aperture objectives available were employed to deliver the excitation light and collect the emission light from the sample.

When experiments were performed at 37° C, a non-immersion 63x/.9n.a. objective or a 100x/1.45n.a. objective with a warming collar was used. The latter was necessary because the temperature gradient between a sample at

37°C and an immersion objective at room temperature cause differences in oil density that introduce noise into the measurement. This and other sources of noise were suppressed (e.g., 5). For quantification of light levels, the image was directed into a Photon Technology International (PTI, Birmingham, NJ) D-104 microscope photometer. To detect emission levels, the photometer utilized a 600nm dichroic mirror and the emission filters HQ530/30m and HQ670/50m (Chroma), which maximize both transmission of the appropriate signal and rejection of out- of-band light. The photometer's apertures were adjusted to remove irrelevant cells from the measurement. The intensities of the two wavelength bands were measured using R1527P single photon counting photomultiplier tubes held in refrigerated housings (Products for Research, Danvers, MA). Dark count rates of about 5 counts per second were routinely observed. The microscope, photometer, and photodetectors were housed in an aluminum enclosure to minimize stray light, the influence of air ducts, etc. All optical components were securely attached to the optical table. Except for the photodetectors, no electronics or moving parts were placed on the optical table. The system was checked for potential ground loops. A PTI interface and computer were utilized to collect the data.

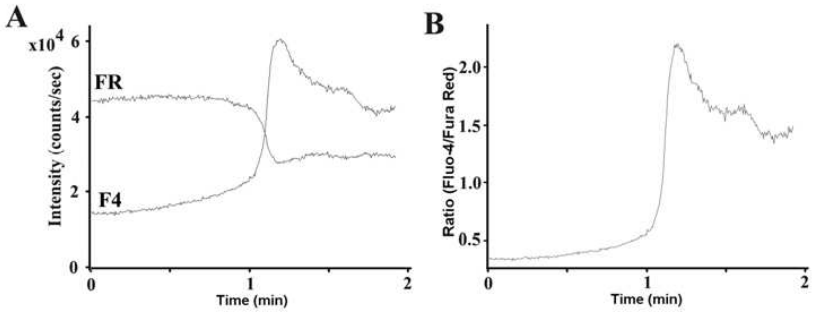


Figure 2: Fluo-4 (F4) and Fura Red (FR) analysis of neutrophils during cell attachment to a glass surface. Panel A shows the raw data whereas panel B presents the data as a ratio.

An example of super-quiet microfluorometry of Ca^{2+} probes is shown in Figure 2. In this experiment, neutrophils were labeled with fluo-4 and fura red, washed, purged for 30 min., then transferred to the imaging buffer (Table 1). Cells were then allowed to spread on glass surfaces during microscope observations. Intensity data from a single neutrophil labeled with Fluo-4 and Fura Red were collected simultaneously, as shown in the plots given in Figure 2A. In this particular experiment, low excitation light levels were used to minimize photobleaching, which resulted in the low emission intensity rates ($\sim 10^4$ counts/sec.). The noise observed in these raw data are at the same order of magnitude as the quantum or shot noise of fluorescence emission. (For

the purposes of noise calculation, it should be noted that the data of Figure 2A were collected at three points per second, thus making the total number of counts in any one channel 1/3 of the rate shown in the graph.) Thus, minimal instrument noise is found in these data. As expected, these two channels of raw data mirror one another. Figure 2B shows these same data after ratioing. As these data show, a high signal-to-noise ratio is observed. The ratio rapidly increases followed by a series of oscillations, one of which is shown at 1.6 min. in Figure 2B. Thus, we have made very precise measurements of the fluo-4/fura red ratio during cell attachment.

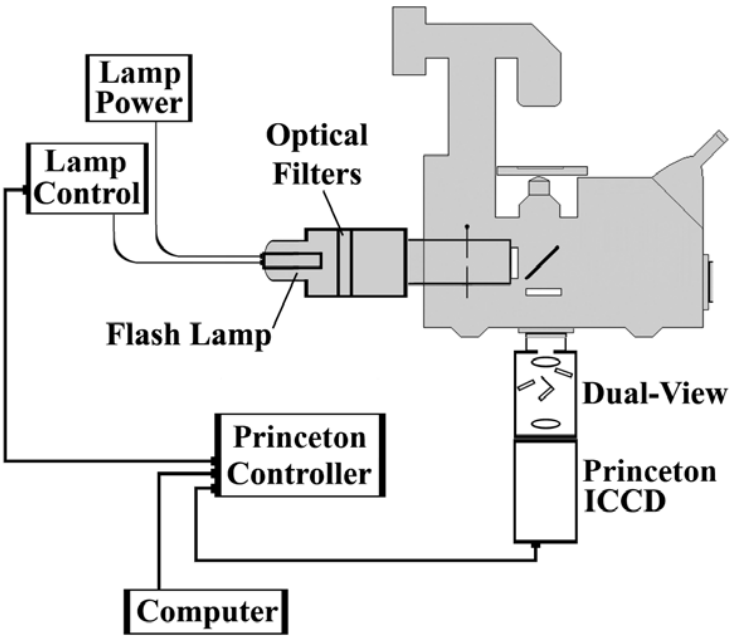


Figure 3: Schematic illustration of flash-lamp-based excitation system.

Although microfluorometry provides precise measurements of the fluorescence intensities of cells, it cannot provide images. Microfluorometry data are often complemented by fluorescence microscopy experiments. However, imaging experiments typically integrate over a time period on the order of a second, which necessarily leads to the randomization small molecules within cells. To eliminate signal and probe diffusion as a source of potential error in imaging experiments, we employed a flash-lamp-based excitation system, as shown in Figure 3. This configuration has been previously described [6, 7]. Briefly, a Perkin-Elmer ElectroOptics (Salem, MA) FX-4400 flash-lamp

generates a 1J 6 sec. pulse of light. One strength of the Perkin-Elmer system is that its electrical noise is low compared to other flash-lamps. To attenuate the IR and UV regions of its output spectrum, a 3 cm water filter was used. Although the flash is very bright, the duty cycle is long enough to provide a cumulative light exposure that is roughly equal to a conventional mercury lamp. Light is delivered to the sample via a 100x/1.45n.a. objective. A filter cube containing an HQ475/40x excitation filter and a 510 nm dichroic mirror (Chroma) was used for fluo- 4/fura red imaging. The emission was directed into a Dual-View (Optical Insights, Tuscon, AZ) apparatus containing a 600nm dichroic mirror and the emission filters HQ530/30m and HQ670/50m (Chroma). Two images are simultaneously projected on a Princeton Instruments (Trenton, NJ) PI- Max II ICCD camera (Gen 3, 1K x 1K), which is cooled by a Peltier sytem. An ICCD camera is chosen as it has better performance than an EMCCD at extremely low light levels [8]. The Princeton Instruments controller interfaces to the camera, a computer, and the lamp control trigger. The timing is set to open the electronic shutter of the camera roughly 200 sec. before the flash is triggered. At full resolution, the frame speed is roughly 8 fps, although this can be greatly increased by defining an appropriate region of interest.

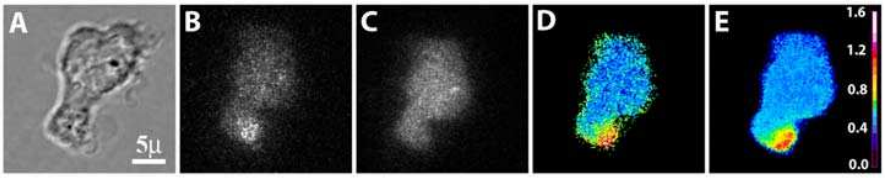


Figure 4: Representative microscope images of a fluo-4/fura red labeled neutrophil. Panel A shows a DIC image. Panels B, C, and D show individual frames exposed with a 6 sec. flash of the fluo- 4, fura red and the fluo-4/fura red ratio image. In panel E, 10 consecutive frames were averaged. The scale at the right gives the ratio.

To illustrate our findings with the system described in Figure 3, neutrophils were labeled with fluo-4 and fura red as described above (Table 1, Figure 2) [6]. Cells were then placed on glass coverslips for observations at 37°C. Figure 4A shows a DIC micrograph of a morphologically polarized neutrophil. Panels B and C show images of the fluo-4 and fura red channels collected simultaneously during a single 6 μsec. flash; as can be seen, these raw data dramatically differ. Panel D shows a fluo-4/fura red ratio image for one frame. Ten consecutive frames were averaged in the micrograph of panel E, which improved the overall signal-to-noise ratio without affecting the conclusions. As these data illustrate, the ratio varies from about 0.4 at the center of the cell to a value of 1.2 at a distal edge of the uropod. As the color LUT in panel E illustrates, a gradient in the fluo-4/fura red ratio is observed from the outer edge of the uropod 1.2 (red) to 0.9 (yellow) to 0.6 (green) and

then to a background level of 0.4 (blue) at the farthest distance from the site. As described in the following section, we interpret these patterns as a Ca^{2+} gradient associated with the uropod.

3 Discussion

In this brief methodological review, we have focused on recent improvements in the kinetic evaluation of fluorescence emission characteristics of living cells. As living cells are far-from-equilibrium systems constantly undergoing change, sensitive kinetic measurements are a necessity in dissecting such dynamic systems. We have utilized wide field microscopy because it is inherently more efficient in kinetic experiments than other modalities, although its resolution is inherently lower as well. The methodologies discussed were directed toward opposite extremes of wide-field microscopy: precise measurements of either intensity or location. Super-quiet microfluormetry provides kinetic intensity measurements whose noise levels are very low near a sample's quantum noise but it provides no spatial information. On the other hand, the flash-lamp-based imaging system provides better insight into the location of signals, but the output is noisy. These approaches provide more precise data than more conventional tools. Although the kinds of information they provide are very different (precise temporal versus precise spatial), these tools nonetheless yield complementary data.

Although generally applicable, super-quiet microflourometry was initially developed for use in applications where ratioing was not possible and therefore the signal is degraded by lamp fluctuations which contribute large amounts of noise. Examples of these situations include NADH and flavoprotein autofluorescence levels. This approach should also be useful in kinetic studies of Ca^{2+} using excitation ratioing, as the excitation light is essentially constant. It is certainly true that this approach can quantify statistically significant changes when other approaches would be compromised by noise. As one can calculate the expected noise in the ratio data, it is possible to ascertain statistically important changes in the ratio. This may, for example, allow the evaluation of excess noise in the signal transduction apparatus and its relevance to disease.

In general, Ca^{2+} imaging experiments of immune cells report that the Ca^{2+} level rises uniformly in the cell. In some cases the center of the cell appears to be the brightest region. As Ca^{2+} is known to enter cells from the extracellular medium and to be released from Ca^{2+} stores near the plasma membrane, this seems contrary to common sense. Depending upon the magnitude of the rate constants involved, it should be possible to detect these regions of Ca^{2+} entry and release. However, to detect these events, image acquisition times should be small in comparison to molecular transport times due to diffusion. As the diffusion coefficient of small molecules and ions is

$\sim 10^{-5}$cm^2/sec, the acquisition times should be $>$100 μsec. to avoid losses in wide field resolution. Although intracellular diffusion coefficients are somewhat smaller, if this physical condition is met then diffusional blurring of the fluorescence image is not physically possible. The Perkin-Elmer flash-lamp used in this study exceeds this criterion by a factor of ten. Therefore, for the live cell images of Figure 4B-D, molecular motions of neither the probes nor Ca^{2+} ions could have contributed to the image.

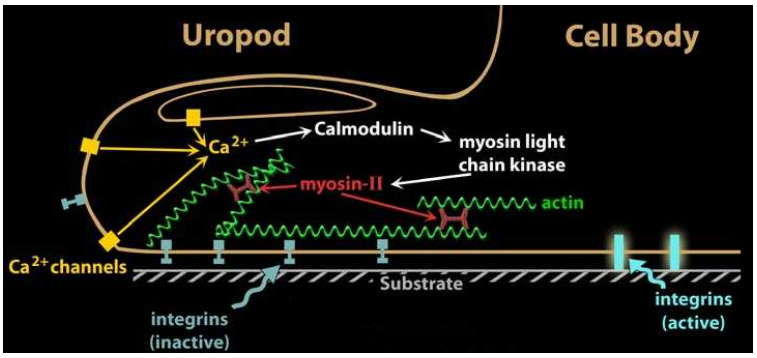


Figure 5: A hypothetical model of Ca^{2+} entry at the uropod and its physiological consequences.

The experimental findings of Figure 4 demonstrate that the uropod of the polarized neutrophil exhibits a very high fluo-4/fura red ratio and is therefore rich in Ca^{2+} signaling. As a consequence of these data, a hypothetical model for the biological role of these Ca^{2+} domains is shown in Figure 5. Ca^{2+} enters the cell cytoplasm via plasma membrane channels and/or cytoplasmic stores near the plasma membrane, thereby setting up the gradient in the fluo-4/fura red ratio reported in Figure 4. This Ca^{2+} signal leads to the local activation of calmodulin and myosin light chain kinase. Myosin II becomes activated thereby leading to force transduction in the cytoskeleton, as suggested by Eddy et al. [9].

Although the findings discussed above have focused on neutrophils, we anticipate that other physiological settings and cell types would benefit from analyses of Ca^{2+} signals using super-quiet or high- speed flash-lamps as excitation sources in wide field microscopy. Moreover, as ratiometric probes are available for many additional biological applications, such as pH values and membrane potentials, the tools described above will likely have applications in the analysis of additional biological signals.

References

[1] Bolsover, S. R., R. A. Silver, & M. Whitaker, (1993) Ratio imaging microscopy of intracellular calcium and pH. *Electronic Light Microscopy* (D. Shotton, ed.) Wiley-Liss, New York. pp. 181-210.

[2] Bakowski, D. & A. B. Parekh, (2007) Regulation of Store-Operated Calcium Channels by the Intermediary Metabolite Pyruvic Acid. *Current Biology* **17**:1076-81.

[3] Floto, R. A., M. P. Mahaut-Smith, B. Somasundaram, & J. M. Allen, (1995) IgG-induced Ca^{2+} oscillations in differentiated U937 cells; a study using laser scanning confocal microscopy and co-loaded fluo-3 and fura-red fluorescent probes. *Cell Calcium* **18**:377-89.

[4] Clark, A. J. & H. R. Petty, (2007) Super-Quiet Microfluorometry: Examples of Tumor Cell Metabolic Dynamics. *Modern Research and Educational Topics on Microscopy*. (Eds., A. Méndez-Vilas, J. Díaz) **3**:403-408

[5] Petty, H. R., (2007) Fluorescence Microscopy: Emerging Methods and Strategies to Extract Weak Signals with Applications in Immunology. *Microscopic Research and Technique* **70**: 687-70.

[6] Clark, A. J. & H. R. Petty, (2008) Observation of Calcium Microdomains at the Uropod of Living Morphologically Polarized Human Neutrophils using Flashlamp-Based Fluorescence Microscopy. *Cytometry*, in press.

[7] Hinkovska-Galcheva, V., A. J. Clark, M. Hiraoka, A. Abe, M. Borofsky, J. A. Shayman, F. Lanni, H. R. Petty & L. A. Boxer, (2007) Ceramide Kinase Mediates Store Operated Calcium Signals and Enhances Phagolysosomal Fusion. *Journal of Lipid Research*, in press.

[8] Dussault, D. & P. Hoess, (2004) ICCDs Edge Out Electron-Multiplying CCDs in Low Light. *Laser Focus World* **40** (4).

[9] Eddy, R. J., L. M. Pierini, F. Matsumura, & F. R. Maxfield, (2000) Ca^{2+}-dependent myosin II activation is required for uropod retraction during neutrophil migration. *Journal of Cell Science* **113**:1287-98.

PART II ARTICLES

Toward a computer-aided methodology for topology-based simulation of the Golgi apparatus

Mathieu Poudret[1,4], Jean-Paul Comet[2], Pascale Le Gall[1,3]
François Képès[1], Agnès Arnould[4], Philippe Meseure[4],
Jean-Marc Verbavatz[5], Alain Rambourg[5]

[1] Epigenomics Project, Genopole® and University of Evry, F-91000 Evry
[2] I3S, UMR 6070 CNRS, Univ. of Nice-Sophia-Antipolis, F-06903 Sophia-Antipolis
[3] MAS, Ecole Centrale Paris, Grande Voie des Vignes - F-92295 Châtenay-Malabry
[4] XLIM/SIC, UMR 6172 CNRS, University of Poitiers, F-86962 Futuroscope
[5] IBITEC-S CEA and CNRS URA2096, F-91191 Gif-sur-Yvette
and LRA17V, University of Paris-Sud 11, F-91405 Orsay

Abstract

The Golgi apparatus is the place in the cell where proteins remain for a maturation phase before their excretion. Several hypotheses on its dynamics have been formulated under the form of schematic models. The GolgiTop working group initiated by the Epigenomics Project (Genopole, Evry) aims at studying the 3-dimensional structure and dynamics of the Golgi apparatus by building 3D structures with the help of computer graphics tools and by animating them using transformation rules which depend on topological, geometric and/or biochemical data. Finally, we introduce a first proposition for a computer-aided methodology for helping biologists to understand and choose hypotheses about the topology and dynamics of the Golgi apparatus.

1 Introduction

The Golgi apparatus is the place in the cell where proteins remain for a maturation phase before their excretion. The generic shape of the 2-dimensional projection of the Golgi apparatus is widely known and rather easily recognizable since it is often located near the cell nucleus and composed of a stack of 5 or 6 flattened cisternae surrounded with numerous vesicles. Moreover, biochemical studies have established that the main role of the Golgi apparatus is the maturation and transportation of proteins. In particular, the Golgi apparatus is the central point for ensuring the excretion of proteins towards the cell environment. In fact, this few and basic biological knowledge about the Golgi apparatus is shared by the whole scientist community.

It remains some open questions about the Golgi apparatus. In particular, answering both following interlinked questions would be essential for a better understanding of the functioning of the Golgi apparatus: first, what are the

plausible 3-dimensional structure of the Golgi apparatus both from a topological and geometric point of view? Secondly, what are the dynamic topological transformations of the Golgi apparatus which ensure the protein excretion? These questions are widely studied and several hypotheses have been formulated under the form of schematic models. These last ones give some intuitions on plausible transformation rules likely to explain the functioning of the Golgi apparatus. In fact, since no experiments allow one to fully access neither the 3D Golgi structure nor its dynamics, some information is missing or misleading to discriminate among the ongoing hypotheses proposed by biologists which one is the most convincing with respect to the biological observations. Numerous experiments, in particular concerning locations or flow of enzymes or various proteins, are achieved to complete the biological knowledge about the Golgi apparatus. Results issued from these experiments can be interpreted according to each hypothesis. As a consequence, the controversy is continuously sustained since supporters of a given hypothesis emphasize the biological results which better fit with their favorite hypothesis and neglect the other ones. At the moment, it is then not possible to discriminate between these hypotheses. Finally, the large amount of works on the Golgi apparatus indicates that the 3D structure and the dynamics of the Golgi apparatus involve really complex and subtil mechanisms.

The GolgiTop working group initiated by the Epigenomics Project (Genopole, Evry) aims at studying both the 3D structure and the dynamics of the Golgi apparatus by building 3D structures with the help of computer graphics tools and by animating them using transformation rules which depend on geometric and/or biochemical data. For each considered hypothesis, we build a 3D structure and animate it with rules which capture the essence of the underlying Golgi apparatus hypothesis. An analysis of 2-dimensional sections helps us in calibrating them according to available biological observations. The initial postulate of the GolgiTop working group is that using computational tools to model and simulate such systems is essential. This entails recognition of the main characteristics of the phenomenon, choice of the appropriate level of abstraction, and comparison of different models. In the case of the Golgi apparatus modelling, compartmentalization is an important issue and a spatial representation of the compartments is needed to describe both static and dynamic characteristics [10].

A variety of approaches have been used to model cellular systems. In particular, rule-based modelling has already been advocated for biochemical reactions since biochemical reactions can be easily translated into transformation rules. In the case of rule-based modelling, formal methods like model checking [1] or symbolic execution [7] have been fruitfully applied to verify that the model satisfies a known property of the biological system. However, many rule-based models ignore compartmentalization and treat the system,

unrealistically, as a homogeneous environment. Recent rule-based modelling takes into account different compartments (see Brane calculi [3], Bioambients [13] and BioCham [2]). In these models, the compartmentalization only captures static topology or simple topological modifications (resulting, for example, from endocytosis or exocytosis) but not geometric aspects (such as the position and shape of the objects).

Topology-based geometric modelling [6] is particularly adapted to represent compartmentalization and is widely advocated for computer graphics. It deals with the representation of the structure of objects (their decomposition into topological units: vertices, edges, faces and volumes) and of the neighbourhood relations that exist between topological units. It treats topological structure and geometry separately; this means that the topological properties of objects can be studied without knowledge of their geometry. In a previous work [9], we formally expressed basic topological operations in terms of generic rules that can be applied to a large family of topological objects and we illustrated this topology-based approach using as examples, resp. a simple interaction between two cells.

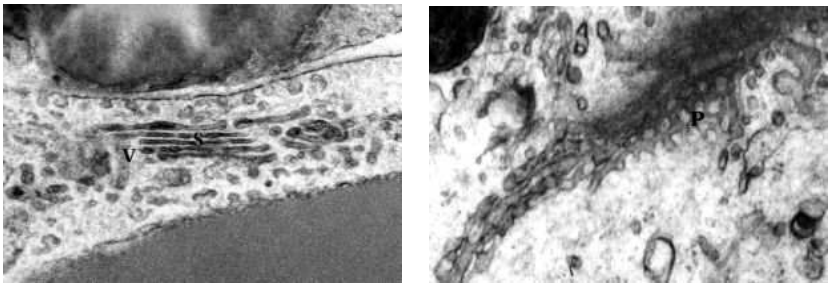
The paper is a short version of [8]: in this paper, we intentionally discard all technical elements related to topology-based geometric modelling (see [8] or [6]). On the contrary, we focus on our methodological approach devoted to help biologists to analyse their hypothesis with computer-aided model animation. The main difficulty is to find an appropriate trade-off between simplicity and fullness. Indeed, models should be sufficiently simple to allow manipulation and reasoning on them but also sufficiently complete to incorporate pertinent elements which are involved in the phenomenon under modelling.

Section 2 introduces some basic features of the Golgi apparatus. Section 3 presents three ongoing hypotheses on the dynamics of the Golgi apparatus. Section 4 describes our rule-based topological approach for the simulation of the models. Finally, in Section 5, we briefly present our computer-aided methodology for helping biologists to understand and choose hypotheses about the topology and dynamics of the Golgi apparatus.

2 The Golgi Apparatus : general description

Discovered by Camillo Golgi in 1898 in the cytoplasm of nerve gaglion cells, the Golgi apparatus (or dictyosome in plants) is an organelle that formed an extensive perinuclear network. Thanks to the use of electron microscope, [4] confirmed in 1954 that in the juxtanuclear area of mammalian cells, the Golgi apparatus usually appears as a system of stacks of closely apposed lamellae also known as saccules or cisternae. It is now widely known that the Golgi apparatus is present in most cells as an organelle made up of stacked

flattened saccules and vesicles. As an illustration, let us observe the electron micrographs 1(a) and 1(b) (this last one can also be found in [12]). The Golgi apparatus appears on 1(a) as a stack of 5 disconnected cisternae (the saccules) bounded with a phospholipidic membrane (see the image part **S**). This saccule stack is usually surrounded by small vesicles that bud out from the saccules (see the image part **V** on Figure 1(a)). Notice that on some pictures like the one given by Figure 1(b), the saccules appear perforated : the image part **P** depicts such regularly perforated saccules.



(a) Saccules stack **S** and vesicles **V** (b) Saccules perforation **P**

Figure 1: The Golgi apparatus

Using histochemical techniques at the light microscope scale, [11] has observed in 1969 an irregular presence of carbohydrates within the Golgi apparatus. Indeed, a carbohydrate gradient appears from the face close to the nucleus, (which contains few such carbohydrate stains) to the other face (which contains a lot of carbohydrate stains). Thus, the Golgi apparatus is a polarised object: the *cis* face is directed to the endoplasmic reticulum while the opposite *trans* face is often directed to the plasma membrane. Subsequent radioautographic and biochemical studies revealed that the Golgi apparatus is involved in the elaboration of complex carbohydrates, by progressively adding carbohydrate elements from its *cis* face to its *trans* face. This may be detailed according to the two following points:

- The main function of the Golgi apparatus is to sort proteins synthesized by the cell and then to transport them from the endoplasmic reticulum to adapted locations as the plasma membrane or lysosomes.

- During the transportation inside the Golgi apparatus, proteins are subjected to a maturation phase by the means of loss of peptidic sequences and addition of sugars (glycosylation) or sulfate (sulfatation).

3 Three hypotheses on the dynamics of Golgi apparatus

Because of observation limitations, the complete structure of the Golgi apparatus is not precisely known. Indeed, with optical microscopy techniques, biologists observe the dynamics at the cost of a small resolution that does not allow them to observe the structure. By contrast, electron microscopy provides high resolution pictures but the observation is done on thin and inert sections of the Golgi apparatus. Last but not least, those thin sections lead to many interpretation mistakes when a 3-dimensional reconstruction is performed (for instance, both spheres and tubes section can appear as discs on a picture).

In particular, the path that proteins follow from the endoplasmic reticulum to the plasmic membrane or lysosomes is not well known. Consequently, three main hypotheses exist [5].

3.1 Vesicular excretion

The first hypothesis views the Golgi apparatus as a static organelle composed of a stack of disconnected flattened saccules surrounded with numerous vesicles. Vesicles are supposed to play a major role in the excretion of proteins. In this vesicular secretion hypothesis (see Figure 2), an aggregate of endoplasmic reticulum (**ER**) fragments generates disconnected saccules (**S**). Proteins migrate through the stack by means of vesicles (**V**) that jump from one saccule to another. They are finally evacuated by the means of secretory granules (**G**) that bud out from the *trans* face. We know that enzymes in charge of the activation and the maturation of proteins are located near the *cis* face of the Golgi apparatus. In this first hypothesis, those enzymes may stay in the first saccules that are motionless by definition.

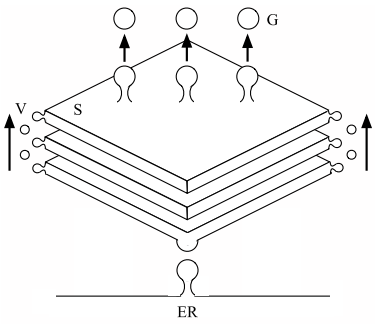


Figure 2: Vesicular excretion

3.2 Saccule maturation

The second hypothesis appears quite similar to the first one since they both suppose that vesicles play a major role in the excretion of proteins. In the saccule maturation hypothesis (see Figure 3), saccules are still disconnected but follow an anterograde movement from the *cis* face to the *trans* face which supports the transport of proteins. Here, vesicles move along a retrograde flow in order to return enzymes that are useful at the beginning (near the *cis* region), of the protein pathway to ensure protein maturation.

Vesicles issued from the endoplasmic reticulum fuse together to form the *cis* saccules. Saccules are then shifted forward to the *trans* face when new saccules are created at the *cis* face.

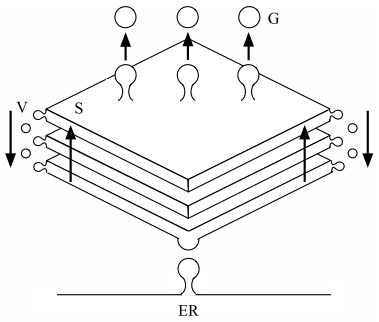


Figure 3: Saccule maturation

3.3 Continuous membrane flow

The third hypothesis promotes a continuous 3D structure for the Golgi apparatus. The saccules are no more isolated but connected together to form a unique continuous structure. Such a view can be represented by the Figure 4. This continuous hypothesis does not rely on any vesicle transportation. On the contrary, it considers a continuous membranes flow (see Figure 4) emerging from the endoplasmic reticulum. Indeed, observed endoplasmic reticulum fragments and vesicles are interpreted in this hypothesis as small sections of a tubular network that connects the saccules (**T**). In this case, proteins may follow the membrane flow and diffuse from one saccule to another along the tubes while enzymes may diffuse following a retrograde movement. Moreover, in this last hypothesis, the saccules perforation may explain the creation of the secretory granules by the rupture of the junctions resulting from the perforation.

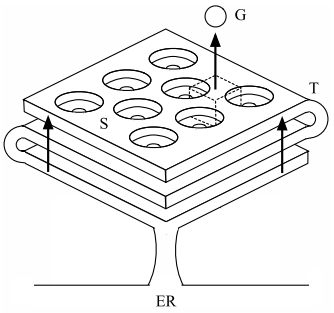


Figure 4: Continuous membrane flow

4 Topology-based simulation technics

4.1 Topology-based geometric modelling

It seems clear that among the numerous features involved in the Golgi apparatus (from the precise shape of the object to the different molecule flows), the role played by the topology is decisive. Thus, a relevant abstraction of the previous Golgi Apparatus hypotheses must handle this component. In [9, 8], we have already proposed a topology-based abstraction dedicated to the animation of simple biological processes.

In order to take the biological compartments in our model into account, we rely on the topology-based geometric modelling (topological modelling for short). This field of the computer graphics deals with the representation of the object structure (their decomposition into topological units: vertices, edges, faces and volumes) and of the neighbourhood relations that exist between topological units. Among numerous topological models, we choose the n-dimensional generalised map [6] (n-G-maps for short). It defines the topology of an n-dimensional space subdivision and allows the representation of a large class of objects[1]. This topological model has the advantage of providing a homogeneous mathematical definition for all dimensions. In this paper, we do not give details about the G-maps model for describing topological structures.

4.2 Topology-based abstraction of Golgi Apparatus hypotheses

In Section 3, we introduced three hypotheses that may explain the behavior of the Golgi apparatus. Two of these hypotheses implicate vesicles in the transport of proteins while the third hypothesis involves a continuous membrane flow in a tubular network. In this section, we use our topology-based approach to model on one hand both the vesicular excretion and the saccule maturation

[1] Quasi-manifolds, orientable or not.

hypotheses and on the other hand the continuous membrane flow hypothesis. Indeed, the vesicular excretion and saccule maturation hypotheses are strictly identical from the topological point of view while the continuous membrane flow introduces significant topological differences (connected and perforated saccules).

Figure 5 illustrates the 3-G-map topological representation of the vesicular excretion and saccule maturation hypotheses. We call it the plate stack model. Figure 6 illustrates the 3-G-map topological representation of the continuous membrane flow. We call it the tower model.

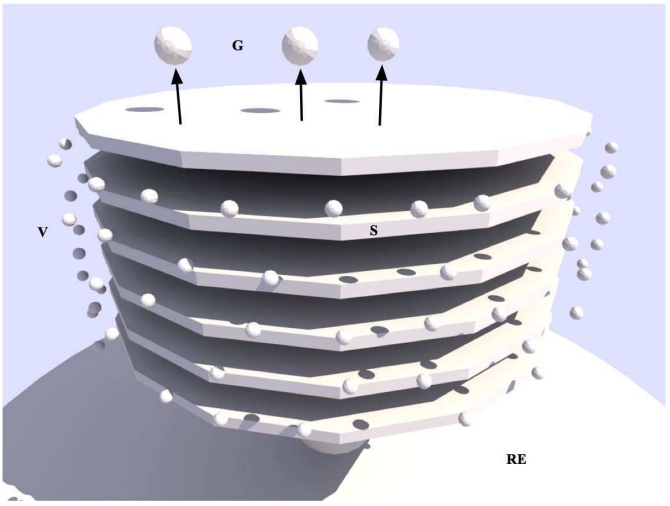


Figure 5: Plate stack model

The topology-based geometric modelling allows one to easily abstract geometry and to focus on pure topology which is, as we said, the most relevant distinction between studied Golgi apparatus hypotheses. When it is necessary, geometric shapes can be associated with the topological units. On Figure 5 and 6, the geometry is basic (here, the object are said to be polyhedric) but the topological differences between hypotheses are captured. The first distinction is the connection between the saccules (**S**). The proteins are transported through vesicles (**V**) in the plate stack while they diffuse into tubes (**T**) that connect the saccules in the tower model. As we see, we choose to abstract both saccules and vesicles with volumes with which we associate concentrations that abstract proteins (concentration gradients are modelled by subdividing volumes and associating different concentrations with each subdivision). Moreover, from a topological point of view, a tube between two saccules is represented with a volume stuck between the connected saccules. Because topological models allow one to handle border of volumes,

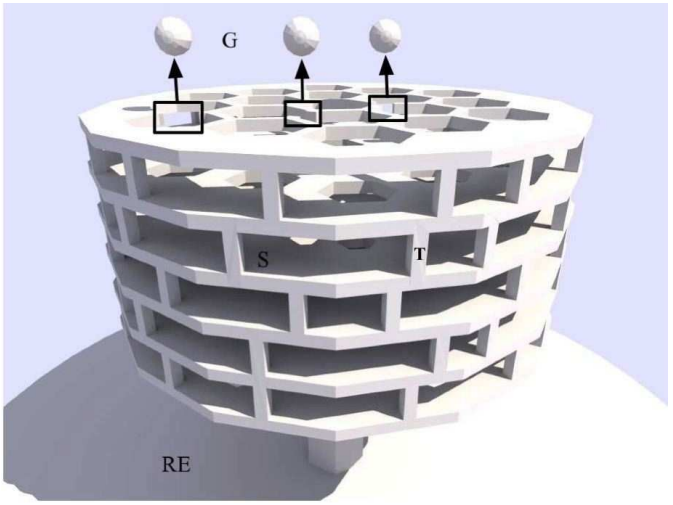


Figure 6: Tower model

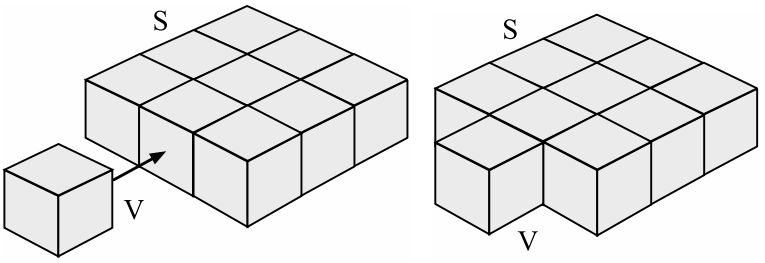
we abstract the transport of proteins by associating permeability on faces that connect saccules to tubes. The second topological distinction concerns the creation of the secretory granules. In the plate stack, the secretory granules (**G**) bud out from the *trans* face (see arrows of Figure 5) while in the tower model, they are constituted of saccule pieces that result from the rupture (framed on Figure 6) of the bee nest structure that abstracts the perforation (according to the biologists, the perforation appears progressively from the *cis* face to the *trans* face). Finally, small parts of the endoplasmic reticulum aggregate into saccules in the first model, while the endoplasmic reticulum is connected to the *cis* face in the second one.

4.3 Topological transformation rules

In order to edit topological objects, computer scientists have defined many topological operations on the n-G-maps. In [9], we have formally expressed the basic operations as graph transformation rules. In order to model biological cellular processes we may want to attribute different kinds of information to the topological units. For instance, we may want to attach types, biochemical data, geometric data (when the biological observation allows it), etc. to the volumes that abstract the biological compartments. Thus, we may want to write transformation rules whose application depends on values of these different data and that modify them.

In order to animate the plate stack model and the tower model, we have to write the rules that capture their dynamics. We succinctly give two examples of such rules. They are detailed in [8] and are illustrated in Figure 7 and

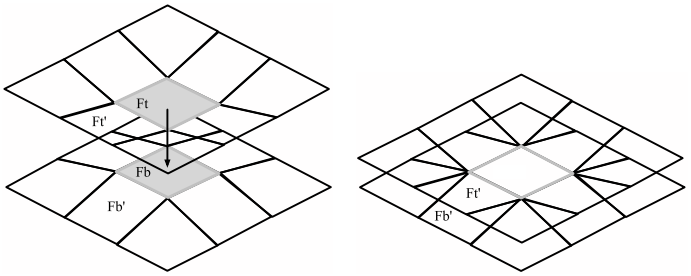
Figure 8.



(a) Matched pattern (b) Result

Figure 7: Gluing a vesicle with a saccule

The first rule is dedicated to the plate stack model. It models the gluing of a vesicle with a saccule which initiates their fusion. Figure 7(a) introduces a simplified representation of the matched pattern, it contains a vesicle (**V**) close to a saccule (**S**). The transformation rule glues them and updates the position of the glued vesicle (see Figure 7(b)).



(a) Matched pattern (b) Result

Figure 8: Perforating a saccule

The second rule is dedicated to the tower model. It provides a means to perforate the saccules (which is one of the behaviors at the root of the continuous membrane flow hypothesis). The matched pattern (see Figure 8(a)) contains two close faces (**Ft**) and (**Fb**) that belong to the same saccule (one is on the top, the other on the bottom). The rule executes the perforation removing faces (**Ft**) and (**Fb**) and linking their neighbours (see Figure 8(b)).

These two examples are well-representative of the transformation rules we need in order to animate the plate stack and tower models. For instance, in the plate stack model, most of the topological operations consist in sticking (it is the case of the example rule) or unsticking topological objects. Notice that the stick operation can also be used to aggregate the pieces of endoplasmic

reticulum that constitute a new saccule while the unstick operation is used to abstract the budding out of vesicles and secretory granules.

We should notice that geometry plays a decisive role in animation processes. The geometric data that influence the biological function we are abstracting are handled in the condition associated with the rules. For instance, when a rule is applied it can take the proximity of objects into account. Other phenomena, e.g. collision detection between vesicles or secretory granules, should be ignored. In fact, many of them only influence the visual rendering, and we do not consider this issue as significant in our context.

5 Toward a topological discrimination of Golgi Apparatus hypotheses

5.1 Iterative construction of the topological models of the Golgi apparatus

Both models, the plate stack one and the tower model, have been elaborated by following a loop of topological model refinements. Biologists have deeply analysed intermediate models by proposing many topological updates. This is particularly true for the continuous membrane flow for which the tower model presented in Figure 6 gives a first insight of its precise topological structure. The study of vertical sections of tower model has played an important role in the biologists' validation process (see [8] for an illustration of such a vertical section). Our tower model, which has been initiated by the biologist observation of perforations and by the interpretation of vesicles as small sections of tubes, is fully compatible with the electron micrographs that are usually used to promote the vesicular excretion and saccule maturation hypotheses. Thus, the tower model is a first original contribution of our topological abstraction process since it has been shown to be consistent with the biological observations.

5.2 Discrimination methodology

The definition of adequate transformation rules is mandatory for animating the topological models but is not sufficient to simulate such complex systems. The rules only define the syntactic part of the simulations, in other words, they define what kind of transformations the simulator performs. Our ongoing work consists in exploring what kind of strategies have to be taken into account when applying transformation rules in order to play simulations. In order to help the biologists to better understand a given biological complex system, we furthermore aim at introducing a computer-aided methodology for analysing topology and dynamics of different hypotheses associated to the biological system. Our goal is not to build an accurate model, but instead, we

Table 1: Simulation parameters

Parameters	Value	Plate stack model	Tower model
Membrane tickness	$7\eta m$	Yes	Yes
Number of saccules	6	Yes	Yes
Saccule thickness	$30\eta m$	Yes	Yes
Saccule length	$50 \times 30\eta m$	Yes	Yes
Secretory granule diameter	$120\eta m$	Yes	Yes
Number of proteins in a granule	600	Yes	Yes
Vesicle diameter	$60\eta m$	Yes	No
Tube diameter	$60\eta m$	No	Yes
Membrane quantity	OUT	Yes	Yes
ATP consumption	OUT	Yes	Yes

would rather discriminate between the different models and choose the one which best approximates the observed phenomena. However, to properly define our discrimination methodology, we need at first to define the parameters of the models.

Table 1 gives an example of parameters that have been discussed with biologists about discrimination of Golgi apparatus hypotheses. The first column displays the name of the parameters. An approximation of their value when they exist (according to the biological state of the art) appears in the second column. A value OUT means that the parameter is computed within the simulation. The third and fourth columns tell whether a given parameter makes sense in respectively, the plate stack model and the tower model. The given set of parameters is not exhaustive but contains the parameters considered by the biologists as the most relevant for the comparison of the two models.

The first six parameters are input parameters (their values are given by the biologists, according to the observations) and are used to initialize the topological models. They are associated with the topological units at the beginning of the simulation and can be refined as we discuss in the next section. The vesicle and tube diameter are also input parameters but fit to only one topological model (respectively the plate stack model and the tower model). Let us remember that depending on the hypothesis we are considering (vesicular excretion, saccule maturation or continuous membrane flow hypothesis), the same pieces of an electron micrograph can be considered as vesicle or tube section. Thus, vesicle and tube diameter must be the same. In other words, updating one of them implies to update the other one. Finally, the last two parameters: membrane quantity and ATP (it is the energetic unit

of the cell) consumption have been chosen among others to discriminate the topological models. Biologists think that the quantity of membrane within the Golgi apparatus (vesicles and secretory granules take part of it) must be constant in time. Thus, if reaching a wanted quantity of transported proteins within the Golgi implies to break this property in one model, this could allow the discrimination of the two models. In the same manner, the fact that the quantity of consumed ATP reaches a critical level could be discriminating too.

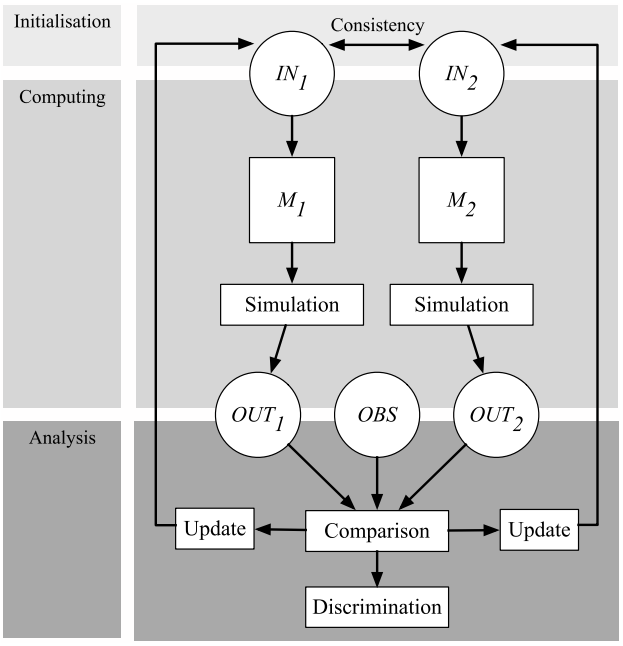


Figure 9: Models discrimination loop

The proposed methodology, which is based on successive simulations of the topological models that implement the hypotheses, is illustrated on Figure 9. The figure only takes into account two topological models but can easily be extended. The different kinds of parameters described in the previous sections are introduced. The input parameters (IN_1 and IN_2) are used to initialize the simulations of, respectively, topological models M_1 and M_2 that implement the selected hypotheses. Note that, as discussed earlier, some parameters can be specific to only one model, but are correlated to parameters of the other model. This consistency between parameters of M_1 and M_2 is necessary for the models discriminating process. OUT_1 and OUT_2 parameters result from the simulations of, respectively, M_1 and M_2 (for instance, the

flow of excreted proteins are output parameters for both models). The results of the simulations are compared with biological experimental observations (OBS on the figure). Our methodology then consists in a refinement process, that is modifying the set of input parameters according to the observations. Note that the models must not be refined independently: the updates still guarantee the consistency between IN_1 and IN_2. Thereafter, we reiterate the simulation, comparison and refinement processes. This loop aims at making both models converge toward the experimental observations. We then select the model that better approximates the observation or eliminate the one that do not converge toward the desired output values. More precisely, a model is given up when one can no longer update parameters in a satisfactory manner. If such a model is not consistent with available biological data, it is refuted and only the other model is considered for further analysis. Our methodology is intentionally simplistic. Our goal is to consider models as simple and abstract as possible to be analysed by biologist experts. When necessary, according to observations or in order to discriminate models, detailed elements are gradually incorporated in the models in order to make models more complex, up until the moment biologist experts can discard one of the models. Thus, there is a trade-off between abstraction of the models and the need of information to discriminate models. The development of finer models becomes useless even if these finer models would be more plausible with respect to the real biological processes under consideration.

Conclusion

In this paper, the Golgi apparatus is widely recognized as a complex biological system where topology plays a key (but poorly-understood) role. In the GolgiTop working group, we develop a topology-based method of modelling cellular processes. This method puts in place transformation rules that allow simultaneous simulation of topological, geometric and biochemical mechanisms. This facilitates a better understanding of the dynamics of these cellular processes that strongly depend on compartmentalization. We first study two very different topological representations based on the three principal hypotheses about the topology of the Golgi: in the plate stack model corresponding to both vesicular excretion and saccule maturation hypotheses, saccules are disconnected and proteins move from one saccule to another *via* vesicles, while in the tower model corresponding to the so-called continuous membrane flow hypothese, saccules are connected with tubes that allow proteins to cross the Golgi. Finally, these topological models can be animated using transformation rules that are determined by the geometric and biochemical data and that determine both these data and the topology itself.

References

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

[2] L. Calzone, F. Fages, and S. Soliman. Biocham: an environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics*, 22(14):1805–1807, 2006.

[3] L. Cardelli. Brane calculi. interactions of biological membranes. In *Proc. Computational Methods in Systems Biology*, volume 3082 of *Lecture Notes in BioInformatics*, pages 257–280. Springer, 2005.

[4] A.J. Dalton and M.D. Felix. Cytological and cytochemical characteristics of the Golgi substance of epithelial cells of the epididymis - in situ, in homogenates and after isolation. *Am Journal of Anatomy*, 94(171-208), 1954.

[5] F. Képès, A. Rambourg, and B. Satiat-Jeunemaître. Morphodynamics of the secretory pathway. *International review of cytology*, 242:55–120, 2004.

[6] P. Lienhardt. Subdivision of n-dimensional spaces and n-dimensional generalized maps. In *SCG'89*, pages 228–236. ACM Press, 1989.

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

[8] M. Poudret, A. Arnould, J.-P. Comet, P. Le Gall, P. Meseure, and F. Képès. Topology-based abstraction of complex biological systems: Application to the Golgi apparatus. *Journal of Theory in Biosciences*, 2008.

[9] M. Poudret, J.-P. Comet, P. Le Gall, A. Arnould, and P. Meseure. Topology-based geometric modelling for biological cellular processes. In *1st Int Conf on Language and Automata Theory and Applications, LATA'07*, April 2007. Available at: http://grammars.grlmc.com/LATA2007/proc.html.

[10] J. F. Presley, T. H. Ward, A. C. Pfeifer, E. D. Siggia, R. D. Phair, and J. Lippincott-Schwartz. Dissection of COPI and Arf1 dynamics in vivo and role in Golgi membrane transport. *Nature*, 417:187–193, May 2002.

[11] A. Rambourg. L'appareil de golgi : examen en microscopie électronique de coupes épaisses (0,5-1μ, colorées par le mélange chlorhydrique-phosphotungstique.). *C.R. Acad. Sci. Paris (série D)*, 269:2125–2127, 1969.

[12] A. Rambourg, Y. Clermont, and Hermo L. Three-dimensional architecture of the golgi apparatus in sertoli cells of the rat. *American Journal of Anatomy*, 154(4):455–476, 1979.

[13] A. Regev, E. M. Panina, W. Silverman, L. Cardelli, and E. Shapiro. Bioambients: an abstraction for biological compartments. *Theor. Comput. Sci.*, 325(1):141–167, 2004.

Importance of the distances between the redox centers and the quinones binding sites in the mitochondrial bc complex

Importance des distances entre les centres rédox et les sites de fixations des quinones dans le complexe bc_1 mitochondrial

Stéphane Ransac[1] and Jean-Pierre Mazat[1]

[1] INSERM U688 Université de Bordeaux 2, 146 rue Léo-Saignat, F-33076, Bordeaux-cedex, France.

Abstract

The bc_1 complex is a central complex in the mitochondrial respiratory chain. It links the transfer of electrons from ubiquinol (QH_2) to cytochrome c and proton translocation across the inner mitochondrial membrane (reviewed in [1-5]). It is widely agreed that the "Q-cycle mechanism" proposed by Mitchell [6-8] correctly describes the bc_1 complex functioning. It is based on an unexpected separation of the two electrons coming from the QH_2 molecule bound at the Q_o site of the bc_1 complex. One electron is transferred to the iron-sulfur centre (FeS) of the iron sulphur protein (ISP) and the second to the lower potential heme b_L. The electron on heme b_L moves within the cytochrome b to reduce the higher potential heme b_H, which in turn reduces an ubiquinone (Q) or a semiquinone (SQ) at a second ubiquinone binding site Q_i (see Fig. 1).

Using a stochastic approach (Gillespie [9]) based on the known spatial structure of bc_1 complexes and the kinetic parameters described by Moser & Dutton [11-13] we have demonstrated in a previous paper [10] the natural emergence of the Q-cycle mechanism and the quasi absence of short-circuits in the functional dimer of bc_1 complex without the necessity to invoke any additional mechanism.

Because the rate constants are highly dependant upon the distances between the reaction centres, it was interesting, using the same stochastic approach, to look for those distances which will appear critical in this mechanism.

In this paper we modulate in turn each distance on the electrons way (see table 1). We show that the rate of the first electron transfer from the ubiquinol to the Iron-Sulfur centre of the ISP (varied by the distance between both centres, Q_o site and FeS centre) is the only limiting step in the normal conditions (Fig. 2 and 3). Three other distances are critical (see Fig. 2): the b_L-b_H, the Q_o-b_L, and the FeS-c_1 distances. The increase in the first

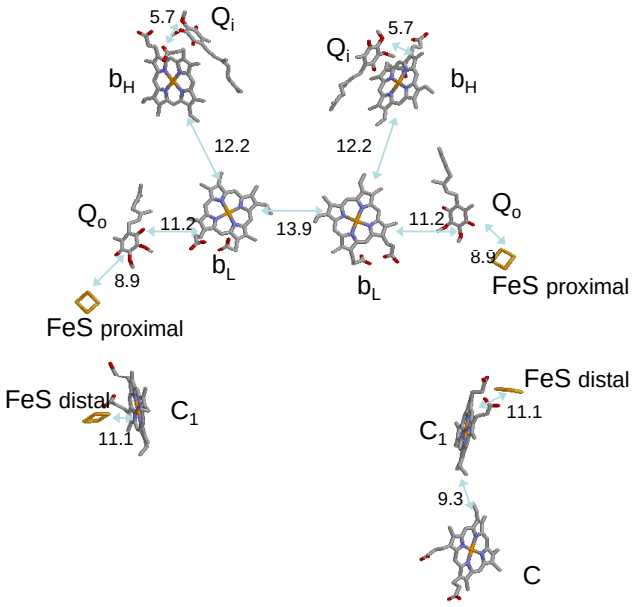


Figure 1: **Structure d'un dimère de complexe** bc_1. Les distances indiquées entre les sites Q_o et Q_i et les centres de réactions rédox (centre Fer-Soufre sur la protéine de Rieske et hèmes b_L et b_H) sont une moyenne calculée à partir de 26 structures de la base PDB. La tête de la protéine de Rieske contenant le centre Fer-Soufre peut adopter deux positions : une position proximale (du site de fixation Q_o) et une position distale (éloignée de Q_o et proche de l'hème du cytochrome c_1). *(Structure of a dimer of bc_1 complex. The distances indicated between the Q_o and Q_i and the redox centres (Iron-Sulfur (FeS) centre on the Rieske protein and et hemes b_L and b_H) are a mean calculated on 26 pdb structures. The head of the Rieske protein which contains the FeS centre can take to position: a proximal position (close to Q_o site) and a distal position (far from Q_o site and close to the cytochrome c_1)*

one decreases the rate of transfer of the second electron in the low potential pathway increasing its return probability on a new semiquinone formed in Q_o site and eventually its direct transfer on the cytochrome c_1. The increase of the Q_o-b_L distance reduces the electron transfer to b_L, the diminution of the FeS-c_1 distance accelerates the return of the FeS in the vicinity of the Q_o where the second electron can still be there; in both cases the second electron jumps directly to the high potential pathway towards the cytochrome c_1. All these three changes increase the short-circuits in which the second electron is more or less transfered directly to cytochrome c_1 and then to cytochrome c.

It has to be noticed that the natural distances measured on the crystallographic structures minimize these short-circuits favouring an energetic "Q-cycle" mechanism.

Résumé

Le complexe bc_1 est central dans l'organisation des oxydations phosphorylantes. Il est largement accepté aujourd'hui que le fonctionnement de ce complexe obéit au schéma décrit par P. Mitchell sous le nom de "cycle Q". Nous avons montré, en nous appuyant sur un modèle stochastique des réactions d'oxydo-réduction intervenant dans le complexe bc_1 que la structure de ce complexe permet naturellement l'émergence du cycle Q et minimise les courts-circuits non générateurs d'un gradient de protons transmembranaire. La vitesse des réactions d'oxydo-réduction est très dépendante de la distance entre les centres réactionnels. Dans cet article nous avons modélisé une variation de ces distances et nous montrons qu'il existe une étape limitante dans les conditions normales qui est le transfert du premier électron sur le centre Fer-Soufre de la protéine de Rieske. Nous montrons d'autre part que tout ce qui ralentit le passage du deuxième électron par la voie de basse énergie favorise son retour vers le site Q_o et son passage direct, non énergétique, vers le cytochrome c (court-circuit).

1 Introduction

Le complexe bc_1 est un complexe central dans la chaîne respiratoire mitochondriale. Il permet le transfert des électrons entre l'ubiquinol (QH_2) et le cytochrome c accompagné d'un transfert de protons de l'intérieur vers l'extérieur de la mitochondrie (revues [1-5]). Il est maintenant largement accepté que son fonctionnement suit la proposition de Mitchell d'un "cycle Q" [6-8] dans lequel les deux électrons venant de l'ubiquinol fixé au site Q_o ont un destin différent. Le premier électron est transféré sur la protéine fer-soufre (ISP ou protéine de Rieske) puis sur le cytochrome c_1 et finalement sur le cytochrome c.

Le deuxième électron emprunte une toute autre voie (voie à bas potentiel) et passe sur l'hème b_L puis l'hème b_H avant de réduire une quinone (Q) ou une semiquinone (SQ) fixée au site Q_i (voir Fig. 1). Cela permet finalement de faire passer tous les électrons des QH_2 vers le cytochrome c mais de plus, de prélever 2 protons dans la matrice mitochondriale et d'en relarguer 4 dans l'espace inter-membranaire.

En utilisant l'approche stochastique décrite par Gillespie [9], les données de distances des structures cristallographiques et la connaissance des potentiels de demi-réduction entre les centres rédox et les sites de fixations nous avons pu démontrer dans un article précédent [10] l'émergence naturelle du fonctionnement du complexe bc_1 selon le "cycle Q" décrit par Mitchell et la quasi-absence de courts-circuits sans qu'il soit nécessaire d'invoquer des hypothèses supplémentaires.

e^- transferts	D (Å)	Variations de D(Å)
Q_o à b_L	11.2	$\pm 0.56; \pm 1.12; \pm 1.68; \pm 2.24; \pm 2.8$
Q_o à FeS	8.9	$\pm 0.445; \pm 0.89; \pm 1.335; \pm 1.78; \pm 2.225$
FeS à c_1	11.1	$\pm 0.555; \pm 1.1665; \pm 1.68; \pm 2.22; \pm 2.775$
b_L à b_H	12.2	$\pm 0.61; \pm 1.22; \pm 1.83; \pm 2.44; \pm 3.05$
b_H à Q_i	5.7	$\pm 0.285; \pm 0.57; \pm 0.855; \pm 1.14; \pm 1.425$
c_1 à c	9.3	$\pm 0.465; \pm 0.93; \pm 1.395; \pm 1.86; 2.325$

Table 1: Les distances considérées pour les différents transferts sont les moyennes de mesures effectuées sur 26 structures cristallographiques différentes. Les modulations effectuées de part et d'autre de cette distance représentent $\pm 5\%$; $\pm 10\%$; $\pm 15\%$; $\pm 20\%$ et $\pm 25\%$ de la distance moyenne. *(The distances for the various transfers are the mean of the measurements on 26 different pdb crystallographic structures. Modulations ($\pm 5\%$; $\pm 10\%$; $\pm 15\%$; $\pm 20\%$ and $\pm 25\%$) are effectuated on both sides of these average values)*

Dans ce travail, la constante de vitesse de transfert des électrons (ou la probabilité dans notre modèle) est donnée par l'équation approchée de Moser & Dutton [11-13] qui est particulièrement sensible aux distances de transfert. Il était donc intéressant de savoir dans quelle mesure des variations de ces distances affectaient le mécanisme global de la réaction, c'est-à-dire l'existence du cycle Q et des courts-circuits. Pour cela dans cet article, nous avons systématiquement varié chaque distance de 0 à $\pm 25\%$ par pas de $\pm 5\%$ à partir de sa valeur initiale (les autres distances étant maintenues constantes) et nous avons mesuré d'une part les effets sur la constante catalytique k_{cat} de la réaction globale et d'autre part le pourcentage d'un certain nombre de courts-circuits déjà décrits par Osyczka et al. [14].

Les résultats montrent l'importance de l'ajustement précis des distances tel qu'il apparaît conservé dans les structures cristallographiques du complexe bc_1 dans différentes espèces.

2 Méthodes et Modèle

Le modèle stochastique que nous employons pour décrire le fonctionnement du complexe bc_1 mitochondrial s'appuie sur la méthode proposée par Gillespie [9]. Elle comprend essentiellement deux choix probabilistes : le choix aléatoire d'une réaction parmi toutes les réactions possibles et la détermination (aléatoire) du temps auquel la réaction choisie aura lieu. Un paramètre important est donc la probabilité d'occurrence d'une réaction donnée qui est liée à la constante de vitesse de réaction k_{et}^{exer}, donnée par l'équation empirique de Moser & Dutton [11-13] :

$$log k_{et}^{exer} = 15 - 0.6D - 3.1\frac{(\Delta G^o + \lambda)^2}{\lambda}$$

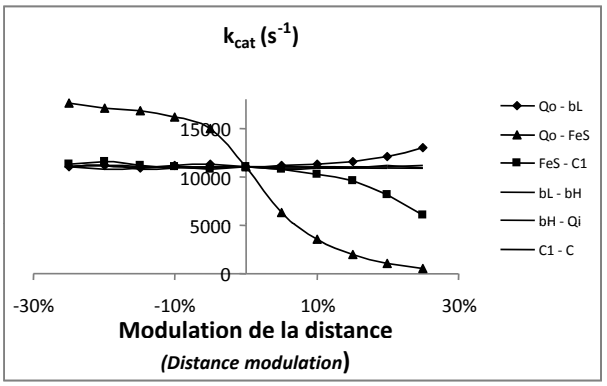


Figure 2: **k_{cat} en seconde^{-1} en fonction d'une modulation des distances.** Les distances entre les différents centres redox ou sites de fixations sont énumérées à droite de la figure ($\pm$5%; $\pm$10%; $\pm$15%; $\pm$20% et $\pm$25% de la distance moyenne). Seules les modifications de distances ayant un effet sur le k_{cat} sont repérées par des signes : distance Q_o-b_L : losanges, distance Q_o-FeS : triangles et distance FeS-c_1 : carrés. Une seule distance est variée à la fois les autres distances ayant les valeurs moyennes calculées (voir tableau 1). *(The distances between the different redox centres or the binding sites are enumared on the right of the figure ($\pm$5%; $\pm$10%; $\pm$15%; $\pm$20% and $\pm$25% of the average distance). Only the distances modifications which have an impact on the k_{cat} are represented with symbols: distance Q_o-b_L: diamonds, distance Q_o-FeS : triangles and distance FeS-c_1: squares. Only one distance is changed at a time, the other distances having the mean values calculated as indicated in table 1)*

k_{et}^{exer} est fonction du potentiel de demi-réduction ΔG^o (en eV) et de la distance D (en Angström) entre les centres réactionnels (calculée bord à bord voir tableau 1 dérivé de [12]).

Le fait que ce soit le logarithme de la constante de vitesse qui dépende linéairement de la distance D entraîne une grande sensibilité de la constante de vitesse par rapport à la distance. Dans les diverses simulations, nous avons étudié l'effet d'une augmentation ou d'une diminution de 5%, 10%, 15%, 20% et 25% sur la vitesse de la réaction. Les simulations sont effectuées avec le modèle décrit dans [10].

La constante catalytique globale est calculée en prenant le nombre de cytochromes c réduits par le temps nécessaire pour effectuer 10 000 réactions élémentaires dans la simulation stochastique. Les valeurs indiquées dans les figures résultent de la moyenne de 10 simulations.

3 Résultats

La figure 2 montre qu'une seule étape est réellement limitante pour la vitesse de fonctionnement du complexe bc_1, il s'agit de l'étape du transfert du premier

électron du quinol QH_2 fixé au site Q_o sur le centre fer-soufre (FeS) de la protéine de Rieske. Lorsque la distance entre Q_o et le centre Fer-Soufre (FeS) dans sa position proximale augmente, la vitesse de la réaction tend vers zéro. La figure 3 confirme que k_{cat} est égal dans ces conditions au k_f de ce transfert du premier électron puisque la pente initiale de k_{cat} en fonction de k_f est de 1 pour des valeurs de k_f comprises entre 0 et 10 000 s^{-1}. Lorsque cette distance (entre Q_o et FeS) diminue, k_f augmente mais d'autres étapes limitent à ce moment là la réaction globale (Fig. 2 et 3).

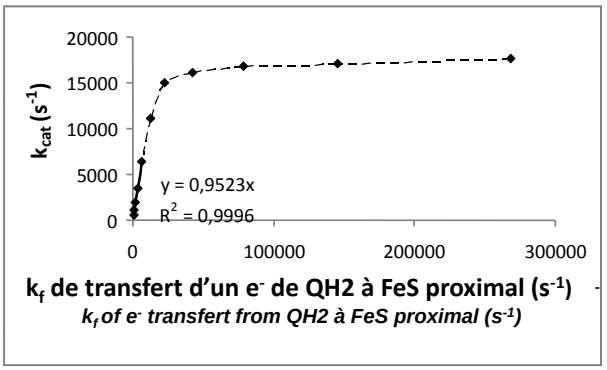


Figure 3: **Représentation de k_{cat} en fonction de la vitesse k_f** (k forward) de transfert des électrons du quinol QH_2 fixé au site Q_o au centre FeS en position proximale. La partie initiale de la courbe (entre 0 et 10.000 s^{-1}) a une pente qui est pratiquement de 1 indiquant une complète dépendance de k_{cat} en fonction de ce k_f. Au-delà de cette valeur de k_f. , la courbe atteint un plateau indiquant que ce k_f n'est plus l'étape limitante. *(Representation of k_{cat} as a function of k_f (k forward) of the transfer reaction of the first electron from the quinol QH_2 bound at Q_o towards the FeS centre in proximal position. The initial part of the curve (between 0 and 10,000 s^{-1}) exhibit a value of 1, which indicates a complete control of k_f on k_{cat}. Beyond this k_f value, the curve reaches a plateau, indicating that this step is no longer limiting)*

La figure 2 (carrés) montre aussi que le transfert de l'électron du FeS sur le cytochrome c_1 peut devenir une étape limitante (elle ne l'est pas au départ avec les distances qui sont celles des données cristallographiques) lorsque la distance entre FeS (dans sa position distale) et c_1 augmente.

La diminution de la distance de Q_o à b_L n'a aucun effet. Par contre son augmentation (losanges dans la figure 2) entraîne une augmentation de la vitesse catalytique mesurée comme la réduction du cytochrome c. Comme on le voit sur la figure 4 cela est dû à l'augmentation des courts-circuits. En effet, dans ces conditions, le deuxième électron n'est pas transféré rapidement sur b_L et demeure au site Q_o sous forme de semiquinone SQ. La tête de la protéine de Rieske a donc le temps de revenir en position proximale et d'accepter ce deuxième électron qui sera ensuite transféré sur le c_1. Dans ces conditions,

il n'y a pas de cycle Q. On remarque sur la figure 4 que la diminution de la distance de FeS (en position distale) à c 1 à pour effet d'augmenter ce même court-circuit en accélérant l'autre mécanisme en compétition c'est-à-dire le transfert du premier électron de FeS à c_1 ce qui permet au centre FeS de revenir plus rapidement en position proximale et de prendre le deuxième électron avant qu'il ne soit passé sur b_L. On remarquera que les distances cristallographiques sont telles qu'elles minimisent ces courts-circuits (Fig. 4).

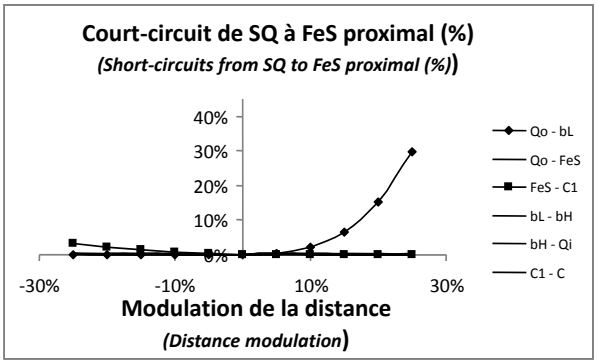


Figure 4: **Pourcentage de courts-circuits de SQ en Q_o à FeS proximal.** Dans ces conditions les deux électrons passent directement sur FeS. Cela peut résulter de l'allongement de la distance entre Q_o et b_L. Le deuxième électron ne passe pas rapidement sur b_L et peut attendre que la tête de la protéine de Rieske portant le FeS revienne en position proximale (losanges dans la partie droite). Ce court-circuit peut aussi résulter de l'accélération du transfert de l'électron sur c_1 permettant un retour plus rapide de la tête de la protéine de Rieske, avant que le deuxième électron ne soit parti sur b_L (carrés dans la partie gauche de la figure). Les symboles ont la même signification que dans la figure 2. *(Short-circuit percentage between SQ in Q_o and FeS in proximal position. In these conditions both QH_2 electrons pass directly on the FeS. It can be the result of the increase in the distance between Q_o and b_L. The second electron does not go rapidly on b_L and can wait until the head of the Rieske protein which carry the FeS comes back in proximal position (diamonds in the right part of the figure). This short-circuit can also be the result of the increase in the rate transfer of the electron on c1 allowing a faster coming back of the head of the Rieske protein, before the second one passes on b_L (squares on the left part of the figure). The symbols have the same signification as in figure 2)*

Enfin la figure 5 montre l'augmentation d'un deuxième type de court-circuit lorsque la distance b_L-b_H augmente. Dans ces conditions, le deuxième électron est transféré normalement sur b_L, mais va y demeurer plus longtemps dû au ralentissement de son transfert sur b_H. Il a alors la possibilité de revenir réduire la semiquinone résultant de la fixation d'une deuxième molécule de quinol $(QH_2)_2$ au site Q_o, laquelle a déjà transmit son premier électron sur le FeS. Ce deuxième électron de la molécule $(QH_2)_1$ qui est revenu au site Q_o

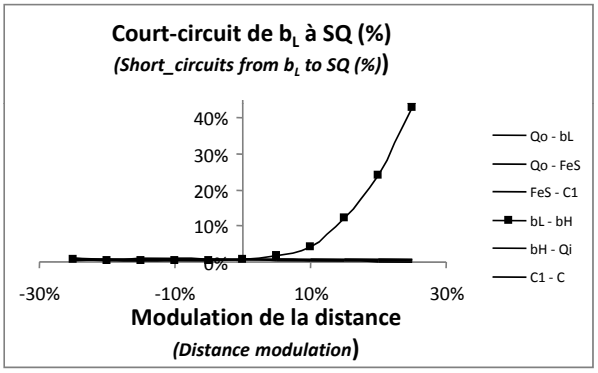


Figure 5: **Pourcentage de courts-circuits de b_L à SQ (%)** Ce court-circuit correspondant au retour du deuxième électron fixé sur b_L et provenant d'une première molécule de $(QH_2)_1$ sur la semiquinone $(SQ)_2$ en Q_o formée par la fixation d'une deuxième molécule de $(QH_2)_2$ ayant transféré son premier électron à FeS. Ce deuxième électron vient réduire de nouveau $(SQ)_2$ en $(QH_2)_2$, ce qui lui permet d'être transféré sur FeS en position proximal. Ce court-circuit résulte de l'allongement de la distance de b_L à b_H qui peut maintenir l'électron suffisamment longtemps sur b_L pour lui permettre de revenir sur une nouvelle molécule de semiquinone $(SQ)_2$. *(Short-circuit percentage between between b_L and SQ This short-circuit corresponds to the return of the second electron bound on b_L which comes from a first molecule $(QH_2)_1$ on the semiquinone $(SQ)_2$ in Q_o formed by the binding of a second molecule $(QH_2)_2$ which has already transferred its first electron on the FeS centre. This second electron reduces $(SQ)_2$ en $(QH_2)_2$, an is then transferred on FeS in proximal position. This short-circuit is the result of the distance increase between b_L and bH . In these conditions, the second electron of $(QH_2)_1$ can remains long enough on b_L and retrun on the new semiquinone molecule $(SQ)_2$)*

pour refaire la deuxième molécule de quinol $(QH_2)_2$ peut être alors transmit au FeS. Il en résulte que les deux électrons de la première molécule $(QH_2)_1$ ont servi à réduire directement le cytochrome c_1 et donc n'ont pas participé à la formation d'un cycle Q.

4 Discussion

Cette étude montre que la seule étape limitante dans le fonctionnement normal du bc_1 complexe (c'est-à-dire avec les distances données par les structures cristallographiques et les potentiels d'oxydo-réduction par la physico-chimie) est le transfert du premier électron du quinol QH_2 sur le centre FeS de la protéine de Rieske et non le transfert de cet électron entre FeS et le cytochrome c_1. Toutefois, il faut remarquer que les fixations et relargages des quinols et quinones sur les sites Q_o et Q_i ont été pris comme très rapides en absence de données précises sur ces étapes qui peuvent intervenir dans la détermination du k_{cat}.

Enfin cette étude montre que le jeu des distances tel qu'il résulte des structures cristallographiques est particulièrement bien adapté pour éviter les courts-circuits et donc favoriser l'émergence naturelle et systématique d'un cycle Q qui est le garant d'un transfert énergétique de protons de l'intérieur vers l'extérieur de la mitochondrie.

Cette étude met bien en évidence l'intérêt d'une approche stochastique des ces mouvements d'électrons qui demeurent des évènements isolés à l'intérieur du complexe bc_1.

Remerciements

Ce travail a reçu le soutien financier de l'ACI IMPBio de l'ANR SysBio/BBSRC 2007 et des ateliers épigénomique, génopole et Université d'Evry.

References

[1] Trumpower, B.L. and Gennis, R.B. (1994) Energy transduction by cytochrome complexes in mitochondrial and bacterial respiration the enzymology of coupling electron-transfer reactions to transmembrane proton translocation. *Ann. Rev. Biochem.* **63**, 675716.

[2] Berry, E.A., Guergova-Kuras, M., Huang, L.-S. and Crofts, A.R. (2000) Structure and function of cytochrome bc complexes. *Annu. Rev. Biochem.* **69**,10051075.

[3] Hunte, C., Palsdottir, H. and Trumpower, B.L. (2003) Protonmotive pathways and mechanisms in the cytochrome bc_1 complex. *FEBS Lett.* **545**,3946.

[4] Crofts, A.R. (2004) The cytochrome bc_1 complex: function in the context of structure. *Ann. Rev. Physiol.* **66**, 689733.

[5] Mulkidjanian, A.Y. (2005) Ubiquinol oxidation in the cytochrome bc_1 complex: reaction mechanism and prevention of short-circuiting. *Biochim. Biophys. Acta* **1709**, 534.

[6] P. Mitchell, Possible molecular mechanisms of the protonmotive function of cytochrome systems, *J. Theor. Biol.* **62** (1976) 327 367.

[7] B.L. Trumpower, The Protonmotive Q Cycle, *J. Biol. Chem.* **265** (1990) 11409-11412.

[8] P.R. Rich, The quinone chemistry of bc complexes, *Biochim. Biophys. Acta* **1658** (2004) 165-171.

[9] D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **8** (1977) 2340-2361.

[10] Nicolas Parisey, Stéphane Ransac, Jean-Pierre Mazat The loneliness of the electrons in the bc_1 complex, soumis.

[11] C.C. Moser, C.C. Page, P.L. Dutton, Darwin at the molecular, *Phil. Trans. R. Soc.* B **361** (2006) 1295-1305.

[12] C.C. Moser, T.A.Farid, S.E. Chabot, P.L. Dutton, Electron tunneling chains of mitochondria, *Biochim. Biophys. Acta* **1757** (2006) 1096-1109.

[13] C.C. Moser, J.M. Keste, K. Warnke, R.S. Farid, P.L. Dutton, Nature of biological electron transfer, *Nature* **355** (1992) 796-802.

[14] A. Osyczka, C.C. Moser, P.L. Dutton, Fixing the Q cycle, *Trends Biochem. Sci.* **30** (2005) 176182.

Modelling Complexity using Hierarchical Interaction Networks

Abdallah Zemirline[1]

[1] LISyC, EA3883, Département Informatique, UFR Sciences et Techniques, Univ. de Bretagne Occidentale, C.S. 93837 BP 809, F-29238 Brest Cedex 3, France.

Abstract

For apprehending *complex systems*, the understanding of *emergent* phenomena and especially the processes leading to the global properties and behaviour of a system from the individual properties and behaviours of its constituents requires in particular to investigate how precisely these components interact together. Besides, a more appropriate organization of the system's constituents and their interactions would be also needed. Indeed, badly organized, an interaction network cannot represent efficiently the interactions, binary or multiple, between entities not necessarily located at the same levels. In this paper, we consider a hierarchical structuring of the observables (entities and interactions) of a complex system, called *hierarchical interaction network*, which has the advantage to allow the representation and modelling of a complex system problem from different views of the data.

Keywords

complex system, emergence, hierarchical hypergraph, interaction network, modelling.

1 Interaction networks of Complex systems

In the following, *complex systems* are dynamic systems in which the components interact simultaneously in ways that are often not precisely known. For example, these systems are a cell, the ensemble of cells that constitute a body, the ensemble of plants and animals that constitute a natural ecosystem, human agents in an economy or in a stock-market, etc.

A characteristic of dynamic systems is the phenomenon of emergence [1] by which properties and behaviour emerge that cannot be readily predicted from a knowledge of the individual properties and behaviours of the system's constituents. This very unpredictability makes emergence hard to model and to work with. This problem is acute in the case of biological systems that contain thousands of different molecules participating in a myriad interactions. Indeed, biological processes [2] are complex systems, where the distribution

of their components is very heterogeneous, where the structures evolve dynamically and their components are in perpetual interaction, and where the interactions may often be between entities located at different levels.

To characterize *emergence*, in the sense where properties and behaviour of a complex system emerge that cannot be readily predicted from a knowledge of the individual properties and behaviours of the system's constituents, two factors which appear essential, the *non-separability* property and the process of *multiple interaction*, were described in [3]. A multiple interaction is an interaction between at least three constituents. For example, if S is a biochemical reaction set, a binary interaction is a reaction between one enzyme and one metabolite whilst a multiple interaction is a reaction involving one enzyme and at least two metabolites.

An *interaction* function is often defined for the interaction set of a given biological system. In this function, to each interaction is associated some value. For example, an interaction function could be defined in terms of the *stoichiometric coefficients* of the biochemical reaction set S.

In a biological system, the interaction set and the interaction function constitute a network of interactions between the system's constituents, called *interaction network*. Knowledge of the full network of interactions between the system's constituents is crucial for a full understanding of a biological process. Such knowledge would provide new insights into the structures and properties of biological systems and would allow useful computations. Several instances of networks of biomolecular interactions are described in [4]. As examples, we can mention the interaction networks where the interactions are respectively between *proteins*, between *enzymes* and *metabolites*, and between *regulatory proteins* and DNA *regulatory regions*.

It should be noted that, in a biological system, the existence of multiple interactions implies that the interaction network has not a *graph* structure but a *hypergraph* [5] structure, and the associated *interaction function* is non-separable. This means the following: if f denotes an interaction function associated to an interaction network $N = (H, f)$, where $H = (V, A)$ is a hypergraph having a node set V and a hyperarc set A, and if $a = (u_1, ..., u_p, v_1, ..., v_q)$ is an element of A, i.e. a is a hyperarc with $u_1, ..., u_p, v_1, ..., v_q$ belonging to V, then f(a) may be very different from the *sum* (resp. *product*, etc) of the $f(u_i, v_j), i = 1, ..., p$ and $j = 1, ..., q$, whenever all the arcs (u_i, v_j) exist in N.

Furthermore, when the interaction network is not a graph but a hypergraph, computing the interplays that would have every two non-neighbouring nodes, in this network, is known to be a non-polynomial computational task. An example of this computing complexity is given by the computation of elementary modes in metabolic networks [6].

To gain an insight into the biological systems requires the construction of suitable models which in general will be *continuous* or *discrete* or even *hybrid*, often *multi-scale*, and which will be also *stochastic*, since a *deterministic* model is only a particular instance of stochastic models. Here, a *model* is a representation of some phenomenon by mathematical objects such as data sets, tables, matrices, relations, rules, equations, trees, graphs, hypergraphs, networks, functions, linear and non-linear functional operators, etc. Typically a model will refer in general only to some aspects of the phenomenon in question. Having built a model for some perceived aspect of reality, this *modelling* can serve for some *computation* or as the basis for *simulation*, an efficient way to complement real-world experiments for non- invasive examination of physical reality.

This is why in the next section, Section 2, we outline a way in which the modelling of a complex system has to be accomplished, when the *classical reductionist approach* "understand the parts in order to understand the whole" is used, and when a *predictive model* is devised for some given objective by means of post-genomic methods. A more appropriate organization of the system's constituents and their interactions will be described. Indeed, badly organized, an interaction network cannot represent efficiently the interactions, binary or multiple, between entities not necessarily located at the same levels. In this paper, we consider a hierarchical structuring of the observables (entities and interactions) of a complex system, called *hierarchical interaction network*, which has the advantage to allow the representation and modelling of a complex system problem from different views of the data.

2 Hierarchical interaction networks for modelling complexity

To understand a complex system and its behaviour, it is often necessary to decompose it into its (known) parts, understand their individual behaviour and their respective functions, and investigate how these components interact together. This "reductionist" approach may need the elaboration of a relevant model (specifying the assumed behaviour of the system), as well as the design of experiment plans for verifying a posteriori the predictions of this model.

In general, as a representation of some phenomenon by mathematical objects, typically a model will refer only to some aspects of the phenomenon in question. Furthermore, a model for biology would be very helpful if it can be validated by a set of biological experiments feasible at a reasonable cost. And it could be useless when it presents a lack of some properties like *observability* and *operability* [7]. The observability is the capability of observing a revealing behaviour by means of biological experiments; the operability is the capability of doing biological experiments on the constituents of a modelled biological entity.

On the other hand, it should also be noted that accuracy is essential in the construction of a network of interactions and in the derivation of its interaction function. Indeed as mentioned in [8], biological function depends upon the exact conformation of molecules and macromolecules; consider for example the difference between D- and L-amino acids [9], and the difference between the enzymes lipases and dehalogenases [10].

Sometimes, an efficient modelling uses a strategy of knowledge integration, that is the task of combining knowledge arising from different independent sources, and of structuring this resource of information, before processing these diverse data. And sometimes when the data are not abundant, a relevant modelling compensates the lack of data by using simulation, in order to process the considered problem. In both of these situations, it is often very useful to structure and organize well the observables [11] of the model, in particular the entities of the system and their interactions. One way to accomplish a good organization of the system's constituents and their interactions consists in using an interaction network structured as a *hierarchical hypergraph*.

2.1 Hierarchical interaction network

In a *hierarchical graph* [12], nodes, arcs or both of them may serve as a grouping primitive. The hierarchical structure can be coded and integrated in the graph itself. A nesting relation between complex nodes, represented by an arc, allows building a hierarchy.

Some main ideas of hierarchical hypergraphs are those of hierarchical graphs which are the following:

- *Grouping*: entities can be grouped together from a certain characteristic, for instance the size, to form higher level structures.
- *Encapsulation*: entities can be hidden inside the hierarchy.
- *Restrictions*: the hierarchical structure and the visibility information can introduce certain restrictions on the underlying data.

Like for hierarchical graphs, the advantage of using a hierarchical hypergraph for structuring an interaction network, that is a hierarchical structuring of observables (entities and interactions) of the system, is that we can represent different views of the data, according to the point in the hierarchy from which we look at it (size, molecular weight, importance of the functions of the system's components, etc).

Thus, in a hierarchical interaction network $R = (N, I, f)$,

- the *node set* N is composed of entities, and is partitioned into levels: each level contains only the nodes in the same hierarchy. A node Y which represents some main component of an entity X belongs to the

level following that of node X; In this way, nodes of the next level are constituted by the main components of the nodes belonging to the courant level, and so on.

Furthermore at each level, except the first eventually, every node is represented by a box where are mentioned all the its characteristics: its name, the names of its constituents belonging to the next level, etc.

- each *hyperarc* of I represents a binary or a multiple *interaction*.
- To each hyperarc h is associated a value $f(h)$, where f is an interaction function defined on I and taking values in R.

2.2 Example of a hierarchical interaction network modelling a dynamic tissue environment

In [13], the authors depict the coexistence of cell, growth factors (GFs) and extracellular matrices (ECMs) in a dynamic tissue environment, and the multiple interactions among them. A knowledge of these multiple interactions is crucial for effectively raising the biological activities of GFs, regulating cell life cycle, designing and preparing exogenous matrices to control GFs release in tissue or organ regeneration by engineering means. The ECMs which are around the cells are largely composed of proteins and polysaccharide secreted by cells, while growth factors, which almost all exist in ECMs, are polypeptides synthesized and secreted by cells as well. In vivo, cell, ECMs and growth factors coexist, and multiple interactions among the three determine the dynamic environment in tissues. These multiple interactions are in particular the bilateral interactions between GFs and ECM (impacts ECM exerting on GFs, effects GFs exerting on ECM, and synergism and antagonism between GFs or between GFs and ECM), the manner and correlative signaling of GFs acting on cells, and the biomimetic requisites of matrices for controlled release of GFs.

For the dynamic tissue environment, and the multiple interactions described in [13], we can build a hierarchical interaction network. Indeed, the dynamic environment of a biological tissue can be represented by the following hierarchical network $R = (N, I, f)$, where:

- N is composed of
  - at the first level: the tissue,
  - second level : ECMs, cells, GFs, other extra-cellular proteins;
  - third and next levels : subcellular components (nucleus, nucleolus, ribosome, vesicle, rough endoplasmic reticulum (ER), Golgi apparatus, Cytoskeleton, smooth ER, mitochondria, vacuole, cytoplasm, lysosome, centrioles, other proteins, etc.

- I is the set of the interactions between elements of N, as described above (impacts ECM exerting on GFs, effects GFs exerting on ECM, and synergism and antagonism between GFs or between GFs and ECM, etc).
- $f\mathrm{f}$ is an interaction function which can be for example a transition function or other.

References

[1] J. Deguet, Y. Demazeau, and L. Magnin, *Elements about the Emergence Issue: A Survey of Emergence Definitions,* ComplexUs 2006, 3: 24-31, Complex Systems Part 2.

[2] V. Norris, A. Cabin and A. Zemirline, *Hypercomplexity*. Acta Biotheoretica, V 54, N 4, 313-330, December 2005.

[3] A. Zemirline and V. Norris, *Nonseparability, Multiple Interactions and Hypercomplexity*, Proceedings of the Evry Spring School on Modelling Complex Systems in the Context of Genomics, 99-102, May 2007.

[4] F. Képès, *Complex Networks of Biomolecular interactions*, Advanced Introduction to Genomic Biology, (October 2006).

[5] C. Berge, *Hypergraphs*, North-Holland Mathematical Library (1989).

[6] S. Klamt, J. Stelling, *Combinatorial complexity of pathway analysis in metabolic networks*, Mol. Biol. Rep., 29, 233-236, (2002).

[7] G. Bernot, J. Guespin-Michel, J.P. Comet, P. Amar, A. Zemirline, F. Delaplace, P. Ballet, and A.Richard, *"Modelling, observability and experiment: a case study Positive feedback loop in a genetic regulatory network"*, Proceedings of the Dieppe spring school on Modelling and simulation of biological processes in the context of genomics, pp 49-55, May 2003.

[8] E. Westhof, *Structural Biology Nucleic Acids*, Advanced Introduction to Genomic, (October 2006).

[9] http://en.wikipedia.org/wiki/Isomer

[10] R.M. de Jong, J.J.W. Tiesinga, A. Villa, L. Tang, D.B. Janssen, and B.W. Dijkstra, *Structural Basis for the Enantioselectivity of an Epoxide Ring Opening Reaction Catalyzed by Halo Alcohol Dehalogenase*, JACS Articles published on Web 09/03/2005

[11] J. L. Casti, *Reality Rules*, J. Wiley & Sons, (1993).

[12] M. Croitoru and E. Compatangelo, *Hierarchical knowledge oriented specification for information integration*, In the Twenty-fifth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, (2005).

[13] J. Bo, J. Yan, D. Zhu, J. Li and K. Yao, *The multiple interactions between growth factors and microenvironment in vivo*, Chinese Science Bulletin, Vol. 51, No. 7, pp 761-769, (2006).

From bioputing to bactoputing: computing with bacteria

Vic Norris[1,2], Abdallah Zemirline[1,3], Patrick Amar[1,4], Pascal Ballet[1,3], Eshel Ben Jacob[5], Gilles Bernot[1], Guillaume Beslon[1,6] , Eric Fanchon[1], Jean-Louis Giavitto[1,7], Nicolas Glade[1], Patrick Greussay[1], Yohann Grondin[1], James A. Foster[8], Guillaume Hutzler[1], François Képès[1], Olivier Michel[1,7], Gradimir Misevic[1,2], Franck Molina[1,9], Jacqueline Signorini[1], Pasquale Stano[10], Alain Thierry[1,11]

[1] Epigenomics Project, Genopole®, F-91000 Evry, France
[2] Laboratoire d'Assemblages moléculaires: modélisation et imagerie SIMS, FRE CNRS 2829, Faculté des Sciences de l'Université de Rouen, 76821 Mont Saint Aignan Cedex, France
[3] LISyC, EA3883, Département Informatique, UFR Sciences et Techniques, Univ. de Bretagne Occidentale, C.S. 93837 BP 809, F-29238 Brest Cedex 3, France.
[4] Laboratoire de Recherche en Informatique, Université Paris Sud, UMR CNRS 8623, F-91405 Orsay cedex, France
[5] School of Physics and Astronomy, Tel Aviv University, 69978 Tel Aviv, Israel
[6] Computer Sciences Dept., Institut National des Sciences Appliquées de Lyon, Bâtiment Blaise Pascal, F-69621 Villeurbanne cedex, France.
[7] Lab. IBISC - LIS project, CNRS FRE 2873, Univ. d'Evry Val d'Essonne, France
[8] Department of Biological Sciences, University of Idaho, Moscow, ID 83844-3051
[9] Laboratoire des Défenses Antivirales et Antitumorales, Université Montpellier2, Place Eugène Bataillon-Bat 24-CC 086, 34095 Montpellier Cedex 5, France
[10] "Enrico Fermi" Research Centre, Compendio del Viminale, 00184 Rome, Italy
[11] Sysdiag FRE3009 - CNRS / BIO-RAD, CAP DELTA 1682 Rue de la Valsiere, F-34184 Montpellier CEDEX 4, France

Summary

The relevance of certain biological materials and processes to computing or bioputing has been explored for decades. These materials include DNA, RNA, enzymes and other proteins whilst the processes include transcription and translation (as well as the control of these processes by protein and by small RNA) and signal transduction. Recently, other directions have been envisaged using bacteria themselves as living computers. Generally, these uses of bacteria fall within the classical paradigm of computing. Computer scientists, however, have a variety of problems to which they seek solutions whilst microbiologists are having new insights into the problems bacteria are solving and how they are solving them. Here, we envisage that bacteria might be used for new sorts of computing. These might be based on the capacity of bacteria to grow, move and adapt to a myriad different fickle environments

as both individuals and as populations of both bacteria and bacteriophage. This new computing may extend to developing a new high level language appropriate to using populations of bacteria and bacteriophage. Such new principles might be based on the way that bacteria explore phenotype space via hyperstructure dynamics and the fundamental nature of the cell cycle. Here we offer a speculative tour of what we term bactoputing, namely the use of the natural behaviour of bacteria and other cells for calculating.

1 Introduction

If the two species, microbiologists and computer scientists, are to interact fruitfully, microbiologists need to have an idea of some of the problems that are of interest to computer scientists whilst computer scientists need to see solutions perhaps to other problems in the knowledge and intuitions of microbiologists.

What is computing? Defined narrowly, it is the systematic study of algorithmic processes that describe and transform information: their theory, analysis, design, efficiency, implementation, and application. The fundamental question underlying all computing is 'What can be (efficiently) automated?' [Denning et al., 1989]. In essence, a Turing machine is a very simple computer. The Turing machine is further specified by a set of instructions which we can think of as a program. What can a Turing machine do or not do? To answer this, consider a Universal Turing machine which is a Turing machine able to read the description of any other Turing machine and to do what that other Turing machine can do. A Universal Turing machine can therefore perform any definite method and, importantly, it could do this without being extraordinarily complex provided it has an immense storage capacity. (Note that a modern computer runs a microprogram that allows its processor chip to take instructions from the main store and compute local functions of them so as to make these instructions resemble those of a particular processor; hence by changing the microprogram, the computer becomes a PC, or a Mac, or a Unix workstation, or any other known computer. Most modern computers are therefore Universal Turing machines). No-one has yet found a plausible model of computation which is more powerful than the Turing machine. Whether living systems constitute or could be turned into more powerful calculating devices than Turing machines is highly controversial (see for example [Ben-Jacob, 2003; Ben Jacob and Shapira, 2004]). In an investigation of how the green sulphur bacterium, Chorobium tepidum, transfers and traps light energy, it has been suggested that it actually performs a quantum computation in using a wavelike characteristic of the energy transfer within the photosynthetic complex to allow the complexes to sample simultaneously different states and find the most efficient path [Engel

et al., 2007]; this can be likened to an algorithm in quantum computing for searching an unsorted information database [Grover, 1997]. That said, as coauthors with differing opinions, we choose in the following to skirt the issue of whether cells offer an alternative to the paradigm of the Turing machine.

What is a cell? It can be argued that the cell is an autocatalytic network, or a neural net, or a tensegrity structure, or a pattern of connectivity with characteristics of Small Worlds and Self- Organised Criticality, or a giant oscillating dipole, or a unit of subjective experience etc. It seems evident that the cell is the creator and the creation of an extraordinarily high density of different organizing processes that have autocatalytic relationships with one another [Norris et al., 2004]. It is a system that produces self-organization and assembly by recruiting and dismissing a multitude of processes and molecules. An exciting question for bioputer designers is therefore what else does this? What else, in other words, could be modelled using a cell and, in particular, a bacterial cell?

What is bioputing? The relevance of certain biological materials and processes to computing has been understood for decades. These materials include DNA, where its value to different sorts of computing, such as the solution of combinatorial problems, is well-known [Adleman, 1994; Carbone and Seeman, 2002; Rothemund et al., 2004]. Such materials in combination with biological processes can constitute effective computers [Benenson et al., 2004]. Hence, bacteria and other cells can be used as a source of new materials with new properties for computing along traditional lines. They may also be used in an intact form for simple forms of such computing [Basu et al., 2005; Skretas and Wood, 2005]. Attempts to construct the minimal cell, inspired in part by origins of life studies and by biotechnological applications, may also produce cells that are amenable for sophisticated, albeit traditional, computing. All these approaches form part of the general approach of what we term *bioputing*.

What is bactoputing? Bactoputing is the use of the natural behaviour of bacteria for computing. As such, it is a subset of bioputing. Here we try to focus on a version of bactoputing in which bacteria are considered as computers. One common approach to computing with bacteria entails adapting them so that they become identical sets of logic gates. Each essentially identical bacterium is then a constituent of a computer; in other words, a homogeneous population of bacteria constitutes the computer. In this approach, one possibility is to use the logic systems that are native to the bacterium, to use, in other words, its original set of networks of gene expression and protein synthesis [Thomas, 1980; Gardner et al., 2000; Atkinson et al., 2003; Ozbudak et al., 2004; Wall et al., 2004]. One of the obvious attractions here would be the capacity of bacteria to multiply cheaply. One of the drawbacks is

that bacteria have a tendency to follow their own agenda and frustrate attempts to engineer them to follow human designs (but see below [Posfai et al., 2006]). An alternative approach is to consider each bacterium as different [Tolker-Nielsen et al., 1998; Booth, 2002; Balaban et al., 2004; Avery, 2006]. In this case, a heterogeneous population of bacteria constitutes the bactoputer. This is the tack we follow here. We choose to ignore a different version of bactoputing in which bacteria are considered as agents that both compute and act; for example, bacteria may be modified to recognise, invade and treat cancer cells or parasites within us [Baker, 2005]. This would involve adapting what certain species of bacteria do anyway, and is therefore in the spirit of bactoputing. More speculatively still, bacteria might be converted into new organelles in a remake of the origins of eukaryotic cells. Such organelles might function to repair the host cell and reverse ageing (Norris, in preparation). This would entail making full use of the capacity of bacteria to sense their environment and to modify it. But again we choose to ignore this in what follows.

In section 2, we mention a number of problems that may, one day, be amenable to bactoputing. These problems include: many combinatorial problems that are unsolvable by traditional computing since they entail polynomial increases in the number of steps needed; hardware problems due to lack of memory or the difficulty of construction in 3D; problems faced by many social groups in which a compromise must be found so as to survive in difficult conditions but to proliferate in favourable conditions; 'undecidable' problems that may require construction of a bactoputer or other novel brain; the problem of finding a new high level language appropriate for a bactoputer. In section 3, we review, for computer scientists interested in bactoputing, certain aspects of microbial physiology along with efforts to construct simplified bacteria by deleting 'superfluous' DNA from genomes, by use of wall-less variants (L-forms), by selecting bacteria via mutation and selection, and by origin-of-life experiments to make liposome-based systems. In section 4, we try to address the problems raised in section 2.

2 A few questions in computer science and the social sciences

2.1 NP problems

If the number of steps in the calculation is given by a function of N and each step takes a microsecond, for $N = 100$, functions such as $log_{10}N$ and N^5 are tractable since they take 2 microseconds and 3 hours, respectively, whilst N^N is not tractable (it would take $3\ 10^{186}$ years). This leads to the idea that an algorithm can be tractable if its behaviour depends polynomially (N^2, N^3 etc.) on the size, N, of input. This idea can be extended to the problem treated which is considered tractable if its worst case can be solved by a tractable

algorithm. The class **P** is the class of tractable decision problems. The class **P** is about polynomial time but there is a wider class of problems, **PSPACE**, that are solvable with a polynomial amount of memory. This has a direct relevance to a bacterial population which, in the right conditions, can undergo an exponential increase in mass.

The Hamiltonian circuit problem is whether there is a route which visits every village exactly once and which ends at the village where it started. The related Travelling Salesman problem (see below) is whether there is a route shorter than a given distance which visits every village at least once and which ends at the start. These are examples of the **NP** class of problems, which may be tractable but for which no polynomial-time algorithm is known. A decision problem (one that needs a 'yes' or a 'no' as an answer) is said to be in **NP** if there exists the equivalent of a lucky guess algorithm (a pseudo-algorithm) for instances of the problem needing a 'yes' that takes less than polynomial time to correctly answer 'yes'. The problem of whether or not P and **NP** are the same class of problems is a major question in mathematics and has economic repercussions. If they are the same class and a problem in **NP** is tractable without, as well as with, a lucky guess, then much larger instances of them can be tackled. If they are different classes and problems in **NP** can be shown to be intractable, the search for certain types of algorithm for them can cease.

Problems can sometimes be transformed into one another (this is the case for the Travelling Salesman and the Hamiltonian Circuit) and, since any polynomial function of a polynomial is itself also a polynomial, as long as the time taken to do the transformation is polynomial in N, the size of the input, the time taken by a polynomial-time algorithm for the transformed problem must also be polynomial. This notion of transformation is important because many **NP** problems can transformed into a problem that is itself in **NP**; the hard problems of **NP** are termed **NP**-complete problems (an **NP**-complete problem is in **NP** and every problem of **NP** can be reduced to it by means of a polynomial transformation) [Garey and Johnson, 1979]. This is a general term for a wide variety of many problems, indeed, every branch of mathematics has its **NP** problems involving networking, timetabling, packing, matrices, geometry, and combinatorial mathematics (note that DNA sequence comparisons are in **NP** if mismatches and gaps are allowed).

2.2 The problem of density

A recurrent problem is that computers have insufficient memory or run too slowly. One limitation to the speed at which computers can run is the distance between components. This limitation is due to what is essentially a 2D construction of the integrated circuit. The possibility of constructing a nanoscale 3D calculating device would therefore be very attractive.

2.3 Optimisation and constraint problems in organisations

Many social and economic problems require an organisation to steer between survival and growth. Companies and universities must survive hard (financial) times and expand in good ones. These appear to be contradictory constraints. No single optimal solution exists. For example, there may be no individual solution to the management problem of what proportion of the staff of a multinational group or of a research organisation should be permanent. One possibility is to consider the organisation as a collection of relatively independent units, such as research laboratories, that could offer a simultaneous diversity of independent solutions. A range of different solutions may be needed but what is this range? Other questions where we might look to bacteria for answers include the optimum number of decision-making levels and identification of the subsystem that actually makes the decision. Finally, many social organisations are constrained by the need to reconcile coherence with their present environment and coherence with their past environments. Research laboratories have to respond to new discoveries and to new funding initiatives but must reconcile these with their research history and, in particular with their skills, experience and interests. Perhaps bacteria have something to teach us here too.

2.4 Recognition and other problems

Electronic circuitries or even neuronal brains may be used to address complex problems that include many undecidable problems such as the recognition of shapes (e.g. is this a picture of a horse?) and optimisation problems with non-separable objective functions (e.g. the problem of attributing local rules to components so as to obtain a given global behaviour). A potential ability to address such problems is one motivation for research into the design of synthetic 'brains'. Some of these brains assemble readily into structures, are easy to understand and straightforward to control, which facilitates interfacing with users. They include self- organised networks of real neurones connected to electronic chips [Demarse et al., 2001]. There are also 'soft' networks where the circuits are not fixed and easy to reconfigure. These include bio-electronic hybrid architectures such as those based on dynamic circuits made of the slime mould Physarum polycephalum [Tsuda et al., 2006] or 'chemical brains' based on collisions between chemical waves in the Belousov-Zhabotinsky reaction [Adamatzky and de Lacy Costello, 2002]. What are the possibilities for a bacterial brain?

2.5 Beyond high level instructions

High level programming languages are written in terms of instructions that include loops (For ... Next), tests (If ... then .... else), and operate on variables

and modules (Gosub). In general, each instruction is specific and instructions are acted on sequentially. How might very different languages be developed? Biological systems have inspired imitation in conventional computing in the case, for example, of genetic algorithms. Might bacteria inspire an effectively different style of computing. Would it be possible, for example, to devise a new type of programming language based on bacterial actions?

3 Bactoputing tools

In the following section, we review some well-known facts about bacteria and mention some recent speculations with the idea that these may be useful for bactoputing. We then propose population-based approaches in which the bacterial population is a single computer but in which each bacterium is a different computing device.

3.1 Phenotypic diversity

Population diversity can pose a problem to those types of bactoputing that require bacteria to behave in standard, constant ways. One solution to this lies in constructing negative feedback circuits to limit the range over which the concentrations of network components fluctuate, as shown for simple genetic circuits in E. coli [Becskei and Serrano, 2000]. However, population diversity can be seen as a solution in search of a problem. One of these problems is species extinction where a possible solution would lie in preserving the small sub-groups in which a disproportionate fraction of the diversity is concentrated [Rauch and Bar-Yam, 2004].

Rather than think of bacteria as identical individuals, it is often useful to think of them in terms of populations of heterogeneous individuals that compete, collaborate and communicate. Bacteria already use peptides and other chemicals that they export into the media and that they then sense to determine population density in the phenomenon of quorum sensing which is implicated in processes that include symbiosis, virulence, competence, conjugation, antibiotic production, motility, sporulation, and biofilm formation [Miller and Bassler, 2001].

Many bacteria are confronted with the problem of a changing environment in which different and sometimes incompatible strategies are required for survival and for growth. This is resolved at the population level by the generation of both phenotypic diversity [Tolker-Nielsen et al., 1998; Booth, 2002; Balaban et al., 2004; Avery, 2006] and genetic diversity. In generating phenotypic diversity, transcription factors are clearly important and, since they are often present in small numbers, there is a role to be played by stochastic noise [Elowitz et al., 2002; Sato et al., 2003]; however, the key role, we and others

have argued, is played by the cell cycle [Norris and Madsen, 1995] which leads to the presence of two or more chemically identical chromosomes within the same cytoplasm that spontaneously adopt complementary patterns of expression to equip the future daughter cells for life in different environments [Minsky et al., 2002]. In generating genetic diversity, there is an interesting phenomenon whereby certain individuals in a stressed population undergo mutations in proof-reading genes that lead to a high level of mutations; when these unhealthy individuals lyse, fragments of their DNA can be taken up and used by other individuals which may thus acquire a beneficial mutation [Matic et al., 2004]. It turns out that for environmental stresses to induce programmed cell death in cultures of E. coli, the bacteria must secrete a specific pentapeptide that is derived from the degradation of glucose-6-phosphate dehydrogenase, a metabolic enzyme [Kolodkin-Gal et al., 2007]. This may mean that there is yet another connection to be understood between individual metabolism and population signalling.

3.2 Plasmids, bacteriophage and transposons

Bacteria possess small 'chromosomes' or plasmids that are replicated independently of their principal chromosome and that can be transferred readily between bacteria. Genetic information can also be transferred between bacteria by bacteriophages; these bacterial viruses are stable and resistant and, protected by a shell of proteins, often transport DNA from one bacterium to another. Genetic information can be transferred within a chromosome or between a chromosome and a plasmid by transposons. This allows them to adapt to exposure to new dangers and to avail themselves of new opportunities. Hence bacteria possess a powerful armoury for altering and rearranging their genetic material. They possess, in other words, a system for both solving problems and for anticipating problems.

3.3 Hyperstructures

In the pursuit of the nature of the bacterial cell, we and others have explored the possibility of the existence of a level of organisation intermediate between macromolecules and whole cells

the level of hyperstructures [Amar et al., 2002; Guzman et al., 2002; Molina and Skarstad, 2004]. A hyperstructure is a collection of diverse molecules (genes, mRNAs, proteins, ions, lipids) that is associated with at least one function. A non-equilibrium hyperstructure is assembled into a large, spatially distinct structure to perform a function and is disassembled, wholly or partially, when no longer required [Norris et al., 2004]. Examples in E. coli of non- equilibrium hyperstructures include a nucleolar hyperstructure (analogous to the microcompartment within which ribosomes are assembled inside the eukaryotic nucleus) for synthesizing ribosomal RNA [Cabrera and

Jin, 2003], the dynamic cytoskeletal hyperstructures, and the division hyperstructure responsible for the invagination of the membrane and peptidoglycan layer [Aarsman et al., 2005]. An equilibrium hyperstructure is also a large spatially distinct structure with a function but its life is not dependent on spending energy. Examples, again in E. coli, include the highly-ordered RecA-DNA co-crystal, which forms when there is insufficient ATP to repair DNA damage, and in which the tight crystalline packaging is believed to protect the DNA by physically sequestering it [Levin-Zaidman et al., 2000]. Certain hyperstructures straddle the non-equilibrium/equilibrium divide such as the flagellar hyperstructure which has both an equilibrium part (the flagellum itself) and, during the formation of the flagellum, a non-equilibrium part comprising the transcribed genes and their products that acts as a sensor of hydration [Wang et al., 2005].

In the hyperstructure approach, hyperstructures structure membranes, cytoplasm and nucleoid and progress through the cell cycle becomes a state cycle of hyperstructures. Such hyperstructures may interact via a variety of mechanisms including the familiar processes of DNA supercoiling, coupled transcription/translation, molecular and macromolecular signalling, tensegrity and local concentrations, as well as the speculative ones of ion condensation, oscillating water structures, and intracellular streaming.

3.4 Minimal genomes

Bacteria have existed for billions of years. As the growing problem of antibiotic resistance shows, they readily adapt to and escape from human control. Computers based on bacteria are therefore likely to have a short lifespan unless the adaptability of bacteria is taken into account or indeed unless it becomes part of the computing. One approach to make bacteria more malleable is to take bacteria such as *E. coli* and *Bacillus subtilis* and to cut down the genome so as to eliminate 'unnecessary' functions [Kobayashi et al., 2003; Gil et al., 2004]. How far might this be taken? Until recently, it was believed that 250 or so genes would be needed for a minimal version of a modern cell in the most favourable conditions [Mushegian and Koonin, 1996] similar to minimal genome sizes inferred by site-directed gene disruptions and transposon-mediated mutagenesis knockouts in several bacteria (for references see [Luisi et al., 2006]). However, the symbiont *Carsonella ruddii*, which lives inside insects, has a 160 kb chromosome that encodes only 182 proteins although admittedly it does lack many genes that are thought essential for independent life outside a host [Nakabachi et al., 2006]. Hence the number of 250 genes might be reduced considerably, for example, if repair and other functions are dispensed with and if protein synthesis is imagined to be performed with a reduced set of ribosomal proteins. Attempts to generate bacteria with minimal genomes have led to engineered *E. coli* strains with

nearly 30% of the genome missing, certain of which grow more slowly than the wild type strain [Hashimoto et al., 2005]. More recently, elimination of recombinogenic sequences and mobile DNA (such as transposons and IS elements), as well as elimination of 'non-essential' and cryptic functions, have generated strains of E. coli that have increased genomic stability, maintain otherwise unstable plasmids and can be electroporated readily with DNA [Posfai et al., 2006]; moreover, these strains grow well. Whilst such strains may be less prone to discarding or perverting the constructs that scientists have inserted into them, they may be less able to follow the natural strategies of bacteria based on the generation of genetic and phenotypic diversity (see above), strategies that may be either welcome or unwelcome depending on the type of computing to be performed.

In developing bacteria for computing purposes, little use has yet been made of L-forms. These are bacteria that have been selected for the loss of their peptidoglycan wall. Despite this major change, these simplified bacteria manage to grow and divide [Onoda et al., 2000; Siddiqui et al., 2006]. Although fragile, they seem to have cytoplasmic membranes that are naked and they may be easier to manipulate and may be more amenable for computing than their parent bacteria. In a sense, they represent a step back towards earlier forms of life that could usefully undergo genome shrinking.

3.5 Directed evolution

Another approach of possible value to bactoputing is that of directed evolution. Mutators have defective DNA proof-reading and generate mutations at a high frequency. Such mutators can be grown for thousands of generations in chemostats under a constant selective pressure to drive genotype and phenotype towards those desired by the experimenter. These selective conditions can result in the bacteria adapting in ways that are not desired, for example, by increasing their capacity to stick the walls of the chemostat and so avoid being flushed out; this type of problem, which may arise in both the construction of the strain and in the operation of the bacteria-based computer, can be partly resolved with two chemostats one being used whilst the other is being sterilised [de Crecy-Lagard et al., 2001]. The mutations in such conditions occur independently of one another in individual bacteria. However, if a proportion of bacteria lyse, the possibility exists that other bacteria can take up their DNA and benefit (see above). Hence, the rapidity of directed evolution can be increased by the use of bacterial species that take up foreign DNA at a high frequency such as *Acinetobacter sp* ADP1 [Palmen et al., 1993].

3.6 Liposomes and origins of life

There are two ways in which study of the origins of life may be useful for a bacteria-based computing. One is in the investigation of what the first cell really was (assuming it really was a specific cell rather than a population or ecosystem [Hunding et al., 2006]); clearer ideas about the nature of the first cells might help us in exploiting their descendants. The second is in the experiments performed. A prime example of combining such hypotheses with experiments is in the construction of the minimal cell *de novo*. This is a bottom-up approach (as opposed to the top-down approach of deleting chunks from an existing genome described above) where the objective is to generate the simplest cell that can be considered alive [Luisi et al., 2006]. The definition of 'alive' here need not trouble those whose objective is to obtain devices for computing. One concept is that of a minimal RNA cell comprising a vesicle with two ribozymes inside, the first of which catalyses the synthesis of the components that self- assemble into membrane whilst the second replicates both itself and the first ribozyme [Szostak et al., 2001]. The properties of vesicles (liposomes when their constituents are lipids) continue to be intensively explored. Vesicles can grow using surfactant precursors and even divide to maintain the original size distribution (for references see [Luisi et al., 2006]). Heterogeneous composition and the presence of channels may help circumvent the problem of the difficult entry of materials in modern membranes based on phospholipids or similar molecules. For example, an -hemolysin pore incorporated into liposomes permits the uptake of small metabolites from the medium [Noireaux and Libchaber, 2004; Noireaux et al., 2005]. A very different way of bringing ions and indeed macromolecules into liposomes might be based on the channels formed by the simple compounds polyphosphate and poly-β- hydroxybutyrate [Das et al., 1997 Norris, 2005 #2258]. An alternative, population-based, approach would be to develop the fusion and fission of heterogeneous liposomes [Norris and Raine, 1998]. There have been numerous experiments on fusion of compartments using water-in-oil emulsions which have the advantage of allowing high local concentrations of reactants (for references, see [Luisi et al., 2006]). Often, these reconstruction approaches entail the production of proteins detectable by fluorescence, such as the Green Fluorescent Protein (GFP). A mutant form of GFP has, for example, been produced in lecithin liposomes [Yu et al., 2001]. Continued development in this area towards real minimal cells coupled to detection systems may prove useful for computing. The work in the PACE project is relevant here (`http://complex.upf.es/ ricard/PACEsite`).

The vision in the bottom-up construction of minimal cells is that they should contain a small set of macromolecules with highly specific functions: the original cells started out simple and became complex [Luisi et al., 2006]. This vision is being fleshed out experimentally [Luisi et al., 2006]. A very dif-

ferent vision is that life appeared in the form of a pre-biotic ecology in which a rich, diverse and complex world of protocells or composomes exchanged their contents [Hunding et al., 2006]. In this vision, the first cells only have meaning within the context of a population and to investigate and exploit this, the minimal cell must give way to the minimal population.

3.7 Colonies and swarming

Populations of the bacterium, *Paenibacillus dendritiformis*, make surfactants to extract fluid from the semi-solid nutrient substrate so as to create a layer within which they can swim. The problem is that the production of the surfactant requires the collective action of a dense bacterial population which the food-depleted substrate can not sustain. The solution they have adopted is to form a colony with a branching structure within each branch the bacterial density is sufficiently high, yet the average population density of the colony is sufficiently low for the nutrients to suffice. Very different patterns form at different nutrient levels. Part of the solution resides in the precise adjustment of the viscosity of the lubricant layer and the production rate of the surfactant in order to generate specific branch structures with specific widths according to the substrate hardness and nutrient levels [Kozlovsky et al., 1999; Ben Jacob and Levine, 2005]. *P. dendritiformis* growing on poor substrates can have either a branching (B) or chiral (C) morphology. On hard substrates where high densities are required to produce enough lubricating fluid, the B morphotype is selected, leading to the formation of colonies with branching, bush-like morphologies whilst on softer substrates, the C morphotype is selected, leading to curly branches that allow faster expansion while also using patches of food left behind as the branches are twisted inward. How exactly are the branches made? Cells go into a non-motile state further back from the colony front, where the nutrient levels is extremely low. They also emit quorum-sensing molecules or pheromones that represent the state of the population and its environment and that occasion changes in gene expression. One of these changes is the inhibition of cell division which leads to them elongating. Upon elongation, the cells alter their collective movement from the typical run- and-tumble of the short B cells to a coordinated forward-backward movement that leads to the branches twisting with a specified handedness (this handedness depends on cell-cell interactions together with the inherent flagella handedness). The two possible morphotypes are inheritable and can coexist for some range of growth conditions. There are also spontaneous transitions to give new patterns that maximize the rate of colony expansion.

Learning from experience has also been described in bacteria. *Paenibacillus vortex*, forms vortices that vary in size from tens to millions of bacteria, according to their location in the colony. The cells in the vortex replicate, and

the vortex expands in size and moves outward as a unit, leaving behind a trail of motile but usually non-replicating cells the vortex branch. Maintaining the integrity of the vortex while it serves as a higher-order building block of the colony requires communication: each cell in the vortex needs to be informed that its role is now more complex, being a member of both the specific vortex and the whole colony, so it can adjust its activities accordingly. This ongoing communication is particularly apparent when it comes to the birth of new vortices. New vortices emerge in the trail behind a vortex following initiation signals that cause the bacteria there to produce more lubricating fluid and to move quite rapidly as a turbulent "biofluid", until an eddy forms and becomes a new vortex. The entire process appears to proceed as a continuous dialogue: a vortex 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, that leaves home (the trail) as a new entity toward the colonization of new territory. Recent findings based on *P. vortex* and other bacteria indicate that bacteria modify their colonies in the presence of antibiotics so as to optimise bacterial survival. It also appears that these bacteria have a short-term memory which enables them to recall the structural solution they found to the antibiotic to which they were exposed most recently [Ben Jacob et al., 2004].

3.8 Signalling

Within bacterial and other cells there are numerous types of signalling pathways of relevance to computing [Bray, 1990]. These include the well-studied two component pathways [Baker and Stock, 2007; Laub and Goulian, 2007] and other systems [Grangeasse et al., 2007] that depend on phosphorylation, those that depend on alarmones such as ppGpp [Wang et al., 2007], systems that depend on poly-β-hydroxybutyrate [Das et al., 1997 Norris, 2005 #2258; Norris, 2005] and those that depend on ions (perhaps even on ion condensation [Ripoll et al., 2004]). Another conceptually very different class of signals exists, at least potentially, in cells. This is the class gemerated by those enzymes that only associate with one another when they are actively engaged in catalysing their cognate reactions (so giving rise to functioning- dependent structures); a wide variety of types of signals in the form of enzymes or metabolites can be generated [Thellier et al., 2006].

Within bacterial communities, chemical signalling occurs via molecules such as N-acyl- homoserine lactones for Gram-negative bacteria, post-translationally modified peptides for Gram-positive bacteria and furanosyl-borate diester for all species [Palkova, 2004] as well as fragments of intracellular enzymes [Kolodkin-Gal et al., 2007]. There has also been intriguing evidence for physical signalling in bacterial communication [Matsuhashi et al., 1996; Norris and Hyland, 1997]).

4 Applying bactoputing to problems

4.1 Solving the travelling salesman problem?

Quorum-sensing can be used as the basis of a population-based computing [Bulter et al., 2004; You et al., 2004]. Suppose that short peptides A and B exported from two different bacteria into the medium bind to receptors in a third bacterium to initiate signal transduction (via for example well-known sensor kinases/response regulators) in this bacterium that then activates or represses synthesis of another peptide C that is exported. In principle, one could have a limitless supply of logic gates of every conceivable type. Each bacterium becomes a swimming logic gate communicating via diffusible peptides. Bacteria that have not taken part in signalling could be eliminated (for example, the sensor kinase could also induce synthesis of a factor that protects the bacterium from an externally added or an internally produced poison). This would be the equivalent of apoptosis in the brain. The numbers of an individual species of bacterium in the population become the equivalent of weights in a neural net. Proteases added to the media could be used to remove signalling peptides and so synchronise the system. In this approach, chemotaxis, which *E. coli* uses to swim up gradients of attractants (or down those of repellents), could be used to produce a structuring of the volume in the flask such that those bacteria that are attracted to others aggregate; such structuring could result in a rapid transfer of peptides between different bacteria and could be detected if different species of bacteria were to emit light of different frequencies. Refinements that might be possible include the use of a particular peptide to activate transport systems so that whole families of gates could be switched on or off so as to construct hierarchies of gating systems. A connection with the environment could be ensured by restricting one class of peptides to be environmental signals and a second class to be the responses at the end of the line (and which could bind to biochips to trigger electrical changes). In the ideal world, learning would occur if the correct combination of response peptides were rewarded by an influx of glucose into the system.

To illustrate how peptide signalling plus differential growth might work, consider the problem of the travelling salesman who has to find the shortest route between the cities A, B, C, D, and E which he must only visit once (see above and Appendix). Suppose each city is represented by a peptide. To set the problem up, we construct a bacterium that has input A and output B (denoted by A→B), another that has A→C ... D→A, E→A etc. Suppose each bacterium can only grow if it receives A, B, C, D and E (each of which induces the expression of a different gene encoding a labile protein essential for growth) and the culture is fed in a chemostat a limiting concentration of A, B, C, D and E (plus everything else in excess needed for growth). This selects

for an autocatalytic network based on signalling of the style A→C, C→E, E→B, B→D and D→A. In addition, each bacterium is engineered at the start so as to produce its output in inverse proportion to the distance between the cities; hence, the greater the distance between the cities, the less the bacterium produces of the output peptide for a constant input peptide (there is the equivalent time delay possibility). The initial population must contain representatives of every pair of cities in both directions (e.g. A→B and B→A). The object is to obtain the most efficient autocatalytic network since this should correspond to the shortest route between the cities. Prolonged cycles of growth occur in a soft gel in which the peptides can diffuse some distance from the bacteria that produce them (so that when the members of an autocatalytic network are near one another, they benefit rather than the entire population); the temperature is then raised so that the gel becomes a sol, the bacteria can be mixed, (perhaps a proportion removed to sample), and the temperature lowered to create a gel again. In the ideal world of tractable, docile, well-behaved bacteria, this should result in the selection of efficient networks corresponding to solutions to good routes to the cities (Amar and Norris, unpublished).

4.2 Tackling the density problem?

Bacterial colonies are complex 3D structures in which the bacteria are densely packed (see above). A cubic millimetre of bacteria would contain at least 109 individual cells. In such a cube, each bacterium could act as both an element of a population-based processor and memory. Bacteria such as *E. coli* can double every 17 minutes so increases in computing capacity are not difficult to envisage. Moreover, in the case of a bacterial colony, the colony constructs itself as it grows (see above). The capacity of bacteria to diversify their phenotypes might also be exploited by creating conditions that allow the growth of those bacteria that take part in the calculation at the expense of those that do not take part.

4.3 Optimisation and constraint problems in organisations

Many bacterial species are extremely good at steering between survival and growth and in generating a huge diversity of behaviours in the individual bacteria that constitute a population. In trying to interpret the phenotypic diversity characteristic of these populations, it would extremely useful to obtain information on the activity of transcription factors at the level of the individual cell. Given that the process of chromosome replication itself is a possible source of diversity [Rocha et al., 2003; Norris et al., 2007], it would also be useful to manipulate rates of replication [Janniere et al., 2007]. Bacterial populations also generate genetic diversity and there should be a way to make use of the mutation strategy adopted by bacteria in conditions of

stress or high population densities (see above [Matic et al., 2004; Kolodkin-Gal et al., 2007]).

It can be argued that the phenotypes of bacteria are determined at the level of hyperstructures (see above) rather than at the level of individual macromolecules (such as genes or proteins or small signalling molecules). In this hypothesis, the bacterial population generates a range of phenotypes by varying the proportion of equilibrium and non-equilibrium hyperstructures present in each bacterium. The function that describes this variation in the population changes with different conditions and different species. How might information about this function be obtained? Ongoing developments in optical and analytic microscopy are making easier to determine which molecules and macromolecules constitute hyperstructures as well as the number and distribution of a given type of hyperstructure within a bacterial population. Hence the eventual problem for students of bactoputing will be to somehow code this information into a useable form.

For billions of years, bacteria have been solving the problem of reconciling coherence with the present environment and coherence with their preceding phenotype. We have proposed that this generation of a meaningful phenotype occurs via *competitive coherence* [Norris, 1998]. This concept is based on the way a bacterium must maintain both the continuity of its composition and the coherence of this composition (with respect to the inside and outside world) so as to have phenotypes that are consistent both over time with one another and at the present time with the environment. Failure to achieve such consistency is disadvantageous and, in a competitive world, punishable by extinction.

4.4 Recognition and other problems

It is well-known that bacteria communicate within colonies (see above). This communication is usually assumed to be chemical in nature but other possibilities should be considered, including sound [Matsuhashi et al., 1998]. Chemical communication occurs by diffusion through the medium and the information may be destined for distant bacteria or for the whole population (in which case, modulation of intensity and perhaps frequency is important). Communication may also be strictly local and depend for example on exchange between neighbours via conjugation pili through which DNA can be sent.

Populations of bacteria in the form of colonies behave like huge and massively interconnected networks with seemingly intelligent behaviours. As living processors they adapt, evolve and organise themselves to process efficiently their environment and, for example, extract nutrients to transform into biomass or decide that an enemy is present. The colony can also spa-

tially reorganise under the action of orienting perturbations such as sources of chemoattractive molecules. Adaptable, reconfigurable bacterial populations that can switch between different organisational modes (i.e. from single motile cells to colonies with well structured morphologies [Ben Jacob and Shapira, 2004].) do indeed possess the properties and capabilities needed for being chemical-biological brains. This switch corresponds to the dynamic transition of the processing system from being very efficient and globally interconnected, but poorly programmable, to being structurally programmable with a better interfacing capacity [Conrad, 1995; Pfaffmann and Conrad, 2000]. As in a self-adaptative loop inducing the structuring and the processing ability of a network of fibrillar agents (e.g. microtubules in [Pfaffmann and Conrad, 2000], the spatially distributed population of bacteria can reorganise according to the environment (e.g. it can respond to the addition at a specific time and place of chemoattractants or nutrients by reorganising spatially and functionally).

How might the problem-solving prowess of colony-forming bacteria such as *P. dendritiformis* be used in bactoputing? The use of such populations for bactoputing can be envisaged through (i) a strong interface by using a restricted list of instructions (e.g. chemical instructions, temperature or electrical stimuli,...) in order to induce specific behaviour in the bactoputer, or (ii) soft and poorly defined interfaces by the direct contact of the bactoputer with the problem it has to solve. The latter case might correspond to a diagnostic chip dedicated to detection of diseases with the bacterial chip acting simultaneously as a sensor, a processor analysing complex data, and an output device that it translates this information into a form intelligible to humans. Moreover its controllability could be reinforced by driving the structuring of the colony into specific, static geometries that have been engineered (e.g. network of microscopic channels, interconnected containers ...). The result would be the accomplishment of a task or macroscopically observed manifestations of the behaviour of the colonies such as the appearance of fluorescent signals or of a colony with a particular morphology.

4.5 Beyond high level instructions

If one were to devise a new programming language based on bacteria, which instructions would it contain? Some of these instructions are easy to suggest: transport (a hundred different ions and molecules); move (up and down gradients in 3D); recombine instructions (between regions of the chromosome or by making use of plasmids and transposons); exchange instructions (by conjugating or by taking up those phage that contain some chromosomal DNA); mutate (or mutate at high frequency in the case of mutator bacteria); send messages (in the form of quorum-sensing molecules and other molecules); replicate the chromosome (and pause during replication), differentiate (per-

haps as a function of a role in a colony) and sporulate (or at least form a bacterium that has increased resistance); grow (at different rates); divide (to make progeny that are smaller and that may differ from one another); lyse to release phage. Other actions are harder to exploit due to the limited state of current knowledge; these include creating a hyperstructure, maintaining or altering the ratio of equilibrium to non-equilibrium hyperstructures, and increasing the diversity of hyperstructures.

How might such instructions be given? The hundreds of factors that control transcription and translation and that mediate the above actions might be manipulated chemically by fusing the genes that control them to inducible promoters (such as the one that controls the lactose operon and that can be induced by the chemical IPTG). They might also be manipulated physically by changing the temperature or exposing the bacteria to radiation or other stresses. Rather than discrete instructions being given, the instruction to the bacterial population would be in the form of a chemical or physical gradient. Hence the instruction would be different at different places.

How would instructions be ordered? Rather than instructions simply being given in a sequence, many instructions would be given simultaneously. A metaphor for the instructions would be that of a landscape with a varied topology, different vegetation, watercourses, soil types etc. Of course, instructions could also be given sequentially (e.g. heat shock followed by cold shock) or the possibilities inherent in pausing DNA replication might be exploited by inserting sequences into the chromosome to allow proteins to bind to them and hence slow down or block replication in chosen regions [Laub et al., 2000; Possoz et al., 2006].

How would results be read out? When the population is a colony, results could be in the form of spatiotemporal patterns (see above) or in the distribution of extracellular signalling molecules. When the population is a suspension of cells, results could be in the form of the molecules and structures that constitute the individual cells.

5 Discussion

Computer scientists are interested in solving combinatorial problems of the **NP-complete** and related classes. We have suggested above that the autocatalytic growth properties of bacterial populations might be exploited to solve the travelling salesman problem. This is an illustration of a weak form of bactoputing that, arguably, is just bioputing since it is not really in the nature of bacteria to perform the task required here. Another weak form of bactoputing would entail constructing bacteria with their metabolic enzymes on the outside (an 'inside-out' metabolism) to create a heterogeneous population in

which each individual bacterium needs the activity of other bacteria to grow. A related but stronger form of bactoputing would be to make use of those bacteria like *Clostridium cellulovorans* that use cellulsomes to degrade the walls of plants and that naturally have metabolic activities on the outside [Doi and Kosugi, 2004].

Computer scientists are also interested in solving problems with hardware, and here bacterial populations offer huge densities (a human intestinal tract contains up to 10^{14} bacteria) with numerous chemical and physical connections in three dimensions [Matsuhashi et al., 1998; Palkova, 2004]. They are also cheap and grow fast as well as being robust and self-repairing. Some species can operate at high temperatures and, of course, the presence of water is not a problem.

In the world of human affairs, an organisation often has to steer between survival and growth where conflicting constraints make it hard to find good solutions. A possible approach to finding these solutions is to use one complex system to model another. Bacteria have been selected for billions of years for their capacity to explore phenotype space; this entails exploiting opportunities to grow and to survive stresses. Both opportunities and challenges come in a huge number of combinations in an evolutionary landscape that is modified by the behaviour of the bacteria themselves. Here, we have suggested that it is in bacterial solutions to the challenge of navigating phenotype space that bactoputer scientists may discover new paradigms and applications. For example, bacterial populations anticipate nutritional crashes and, in doing this, they communicate with one another and lyse [Kolodkin-Gal et al., 2007], they also increase phenotypic diversity in the rundown to stationary phase [Vohradsky and Ramsden, 2001]. The prediction here is that as oil supplies run out and global warming increases our societies will go through a period of experimentation, which if unsuccessful, will be followed by convergence on some spartan model. Maybe a bactoputer could help us do this intelligently.

The use of bacterial colonies in their native state as 'brains' to solve recognition and other problems would be a strong form of bactoputing. It would be possible to construct a transparent chip into which grooves were cut (with for example a focussed ion beam) that would have diameters similar to those of a bacterium; bacteria containing fluorescent labels could then explore a network of grooves as guided by a variety of chemoattractants and chemorepellents; a stack of such chips, with channels connecting them, might then be made into a bacterial brain that could be described as a bactoputer insofar as it would be based on a natural property of bacteria, namely, chemotaxis. Another strong form of bactoputing could directly involve the metabolism (the network of reactions, catalysed by enzymes, that creates the cell) and, given that metabolic enzymes are encoded by genes, it should be possible to

design circuits based on coupling metabolism and gene expression [Thellier et al., 2006]. Such bactoputing could be useful in studying social systems where money is both a 'nutrient' and a signal. Significantly, it has been shown that a regulatory circuit in metabolism, namely the lycopene biosynthesis pathway in *E. coli*, can be engineered to control gene expression in response to the intracellular metabolite, acetyl phosphate [Farmer and Liao, 2000].

Perhaps the most exciting aspect of bactoputing would lie in the development of a totally new high language for computing based on the language that bacteria themselves speak (divide, replicate DNA, mutate, lyse, produce phage, conjugate etc.). As our understanding of regulation in bacteria increases, our capacity to manipulate bacteria to give them instructions also increases. But if we can speak to them, can we also listen? For that, we need better access to the phenotypes of individual bacteria. Technological advances in 'omics' will one day perhaps soon give rapid access to information on the proteomes, phosphorylomes, lipidomes, interactomes, metabolomes etc. of large numbers of individual cells. Bactoputer scientists should be preparing for this day now.

Acknowledgements

We thank the Epigenomics Project for support. VN thanks Philippe Marliere for several ideas as well as A. N. Walker for the course available on `www.maths.nott.ac.uk/personal/anw/G12FCO/`

References

[1] Aarsman ME, Piette A, Fraipont C, Vinkenvleugel TM, Nguyen-Disteche M, den Blaauwen T: Maturation of the escherichia coli divisome occurs in two steps. Molecular microbiology 2005;55:1631-1645.

[2] Adamatzky A, de Lacy Costello B: Collision-free path planning in the belousov-zhabotinsky medium assisted by a cellular automaton. Naturwissenschaften 2002;89:474-478.

[3] Adleman LM: Molecular computation of solutions to combinatorial problems. Science 1994;266:1021-1024.

[4] Amar P, Ballet P, Barlovatz-Meimon G, Benecke A, Bernot G, Bouligand Y, Bourguine P, Delaplace F, Delosme J-M, Demarty M, Fishov I, Fourmentin-Guilbert J, Fralick J, Giavitto J-L, Gleyse B, Godin C, Incitti R, Képès F, Lange C, Le Sceller L, Loutellier C, Michel O, Molina F, Monnier C, Natowicz R, Norris V, Orange N,

Pollard H, Raine D, Ripoll C, Rouviere-Yaniv J, Saier jnr. M, Soler P, Tambourin P, Thellier M, Tracqui P, Ussery D, Vannier J-P, Vincent J-C, Wiggins P, Zemirline A: Hyperstructures, genome analysis and i-cell. Acta Biotheoretica 2002;50:357-373.

[5] Atkinson MR, Savageau MA, Myers JT, Ninfa AJ: Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in escherichia coli. Cell 2003;113:597607.

[6] Avery SV: Microbial cell individuality and the underlying sources of heterogeneity. Nature reviews 2006;4:577-587.

[7] Baker M: Better living through microbes. Nature biotechnology 2005;23:645-647.

[8] Baker MD, Stock JB: Signal transduction: Networks and integrated circuits in bacterial cognition. Curr Biol 2007;17:R1021-1024.

[9] Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S: Bacterial persistence as a phenotypic switch. Science 2004;305:1622-1625.

[10] Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R: A synthetic multicellular system for programmed pattern formation. Nature 2005;434:1130-1134.

[11] Becskei A, Serrano L: Engineering stability in gene networks by autoregulation. Nature 2000;405:590593.

[12] Ben-Jacob E: Bacterial self-organization: Co-enhancement of complexification and adaptability in a dynamic environment. Philosophical transactions 2003;361:1283- 1312.

[13] Ben Jacob E, Becker I, Shapira Y, Levine H: Bacterial linguistic communication and social intelligence. Trends in Microbiology 2004;12:366-372.

[14] Ben Jacob E, Levine H: Self-engineering capabilities of bacteria. Journal of the Royal Society Interface 2005:doi:10.1098/rsif.2005.0089.

[15] Ben Jacob E, Shapira Y: Meaning-based natural intelligence vs. Information-based artificial intelligence.; in Ben-Nun H, (ed): The cradle of creativity. Jerusalem, Shaarei Tzedek, 2004, vol.

[16] Benenson Y, Gil B, Ben-Dor U, Adar R, Shapiro E: An autonomous molecular computer for logical control of gene expression. Nature 2004;429:423-429.

[17] Booth IR: Stress and the single cell: Intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. International Journal of Food Microbiology 2002;78:19-30.

[18] Bray D: Intracellular signalling as a parallel distributed process. Journal of theoretical biology 1990;143:215-231.

[19] Bulter T, Lee SG, Wong WW, Fung E, Connor MR, Liao JC: Design of artificial cell-cell communication using gene and metabolic networks. Proceedings of the National Academy of Science USA 2004;101:2299-2304.

[20] Cabrera JE, Jin DJ: The distribution of rna polymerase in escherichia coli is dynamic and sensitive to environmental cues. Molecular microbiology 2003;50:1493-1505.

[21] Carbone A, Seeman NC: Circuits and programmable self-assembling DNA structures. Proceedings of the National Academy of Science USA 2002;99:12577-12582.

[22] Conrad M: Scaling of efficiency in programmable and non-programmable systems. Bio Systems 1995;35:161-166.

[23] Das S, Lengweiler UD, Seebach D, Reusch RN: Proof for a nonproteinaceous calcium- selective channel in escherichia coli by total synthesis from (r)-3-hydroxybutanoic acid and inorganic phosphate. Proceedings of the National Academy of Science USA 1997;94:9075-9079.

[24] de Crecy-Lagard VA, Bellalou J, Mutzel R, Marliere P: Long term adaptation of a microbial population to a permanent metabolic constraint: Overcoming thymineless death by experimental evolution of escherichia coli. BMC Biotechnol 2001;1:10.

[25] Demarse TB, Wagenaar DA, Blau AW, Potter SM: The neurally controlled animat : Biological brains acting with simulated bodies. Autonomous Robots 2001;11:305-310.

[26] Denning PJ, Comer DE, Gries D, Mulder MC, Tucker A, Turner A, Young PR: Computing as a discipline.; in: Communications of the ACM, 1989, vol 32, pp. 9-23.

[27] Doi RH, Kosugi A: Cellulosomes: Plant-cell-wall-degrading enzyme complexes. Nature Reviews Microbiology 2004;4:541-551.

[28] Elowitz M, B. , Levine AJ, Siggia ED, Swain PS: Stochastic gene expression in a single cell. Science 2002;297:1183-1186.

[29] Engel GS, Calhoun TR, Read EL, Ahn TK, Mancal T, Cheng YC, Blankenship RE, Fleming GR: Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. Nature 2007;446:782-786.

[30] Farmer WR, Liao JC: Improving lycopene production in escherichia coli by engineering metabolic control. Nature Biotechnology 2000;18:533-537.

[31] Gardner TS, Cantor CR, Collins JJ: Construction of a genetic toggle switch in escherichia coli. Nature 2000;403:339-342.

[32] Garey MR, Johnson DS: Computers and intractibility: A guide to the theory of np- completeness. New York, Freeman, W.H., 1979.

[33] Gil R, Silva FJ, Pereto J, Moya A: Determination of the core of a minimal bacterial gene set. Microbiology and Molecular Biology Reviews 2004;68:518-537.

[34] Grangeasse C, Cozzone AJ, Deutscher J, Mijakovic I: Tyrosine phosphorylation: An emerging regulatory device of bacterial physiology. Trends Biochem Sci 2007;32:86- 94.

[35] Grover LK: Quantum mechanics helps in searching for a needle in a haystack. Physical Review Letters 1997;79:325-328.

[36] Guzman EC, Caballero JL, Jimenez-Sanchez A: Ribonucleoside diphosphate reductase is a component of the replication hyperstructure in escherichia coli. Molecular microbiology 2002;43:487-495.

[37] Hashimoto M, Ichimura T, Mizoguchi H, Tanaka K, Fujimitsu K, Keyamura K, Ote T, Yamakawa T, Yamazaki Y, Mori H, Katayama T, Kato J: Cell size and nucleoid organization of engineered escherichia coli cells with a reduced genome. Molecular microbiology 2005;55:137-149.

[38] Hunding A, Kepes F, Lancet D, Minsky A, Norris V, Raine D, Sriram K, Root-Bernstein R: Compositional complementarity and prebiotic ecology in the origin of life. BioEssays 2006;28:399412.

[39] Janniere L, Canceill D, Suski C, Kanga S, Dalmais B, Lestini R, Monnier AF, Chapuis J, Bolotin A, Titok M, Chatelier EL, Ehrlich SD: Genetic evidence for a link between glycolysis and DNA replication. PLoS ONE 2007;2:e447.

[40] Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Debarbouille M, Dervyn E, Deuerling E, Devine K, Devine SK, Dreesen O, Errington J, Fillinger S, Foster SJ, Fujita Y, Galizzi A, Gardan R, Eschevins C, Fukushima T, Haga K, Harwood CR, Hecker M, Hosoya D, Hullo MF, Kakeshita H, Karamata D, Kasahara Y, Kawamura F, Koga K, Koski P, Kuwana R, Imamura D, Ishimaru M, Ishikawa S, Ishio I, Le Coq D, Masson A, Mauel C, Meima R, Mellado RP, Moir A, Moriya S, Nagakawa E, Nanamiya H, Nakai S, Nygaard P, Ogura M, Ohanan T, O'Reilly M, O'Rourke M, Pragai Z, Pooley HM, Rapoport G, Rawlins JP, Rivas LA, Rivolta C, Sadaie A, Sadaie Y, Sarvas M, Sato T, Saxild HH, Scanlan E, Schumann W, Seegers JF, Sekiguchi J, Sekowska A, Seror SJ, Simon M, Stragier P, Studer R, Takamatsu H, Tanaka T, Takeuchi M, Thomaides HB, Vagner V, van Dijl JM, Watabe K, Wipat A, Yamamoto H, Yamamoto M, Yamamoto Y, Yamane K, Yata K, Yoshida K, Yoshikawa H, Zuber U, Ogasawara N: Essential bacillus subtilis genes. Proceedings of the National Academy of Sciences of the United States of America 2003;100:4678-4683.

[41] Kolodkin-Gal I, Hazan R, Gaathon A, Carmeli S, Engelberg-Kulka H: A linear pentapeptide is a quorum-sensing factor required for mazef-mediated cell death in escherichia coli. Science 2007;318:652-655.

[42] Kozlovsky Y, Cohen I, Golding I, Ben-Jacob E: Lubricating bacteria model for branching growth of bacterial colonies. Physical Review E 1999;59:7025-7035.

[43] Laub MT, Goulian M: Specificity in two-component signal transduction pathways. Annu Rev Genet 2007;41:121-145.

[44] Laub MT, McAdams HH, Feldblyum T, Fraser CM, Shapiro L: Global analysis of the genetic network controlling a bacterial cell cycle. Science 2000;290:2144-2148.

[45] Levin-Zaidman S, Frenkiel-Krispin D, Shimoni E, Sabanay I, Wolf SG, Minsky A: Ordered intracellular reca-DNA assemblies: A potential site of in vivo reca-mediated activities. Proceedings of the National Academy of Science USA 2000;97:6791-6796.

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

[47] Matic I, Taddei F, Radman M: Survival versus maintenance of genetic stability: A conflict of priorities during stress. Research in Microbiology 2004;155:337-341.

[48] Matsuhashi M, Pankrushina AN, Endoh K, Watanabe H, Ohshima H, Tobi M, Endo S, Mano Y, Hyodo M, Kaneko T, Otani S, Yoshimura S: Bacillus carboniphilus cells respond to growth-promoting physical signals from cells of homologous and heterologous bacteria. Journal of General and Applied Microbiology 1996;42:315-323.

[49] Matsuhashi M, Pankrushina AN, Takeuchi S, Ohshima H, Miyoi H, Endoh K, Murayama K, Watanabe H, Endo S, Tobi M, Mano Y, Hyodo M, Kobayashi T, Kaneko T, Otani S, Yoshimura S, Harata A, Sawada T: Production of sound waves by bacterial cells and the response of bacterial cells to sound. J Gen Appl Microbiol 1998;44:49-55.

[50] Miller MB, Bassler BL: Quorum sensing in bacteria. Annual review of microbiology 2001;55:165199.

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

[52] Molina F, Skarstad K: Replication fork and seqa focus distributions in escherichia coli suggest a replication hyperstructure dependent on nucleotide metabolism. Molecular microbiology 2004;52:1597-1612.

[53] Mushegian AR, Koonin EV: A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proceedings of the National Academy of Sciences of the United States of America 1996;93:10268-10273.

[54] Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M: The 160- kilobase genome of the bacterial endosymbiont carsonella. Science 2006;314:267.

[55] Noireaux V, Bar-Ziv R, Godefroy J, Salman H, Libchaber A: Toward an artificial cell based on gene expression in vesicles. Physical biology 2005;2:P1-8.

[56] Noireaux V, Libchaber A: A vesicle bioreactor as a step toward an artificial cell assembly. Proceedings of the National Academy of Sciences of the United States of America 2004;101:17669-17674.

[57] Norris V: Modelling e. Coli: The concept of competitive coherence. Comptes Rendus de l'Academie des Sciences 1998;321:777-787.

[58] Norris V: Poly-(r)-3-hydroxybutyrate and the pioneering work of rosetta natoli reusch. Cellular and molecular biology (Noisy-le-Grand, France) 2005;51:629-634.

[59] Norris V, Amar P, Bernot G, Delaune A, Derue C, Cabin-Flaman A, Demarty M, Grondin Y, Legent G, Monnier C, Pollard H, Raine D: Questions for cell cyclists. Journal of Biological Physics and Chemistry 2004;4:124-130.

[60] Norris V, Hyland GJ: Do bacteria "Sing"? Molecular microbiology 1997;24:879-880.

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

[62] Norris V, Madsen MS: Autocatalytic gene expression occurs via transertion and membrane domain formation and underlies differentiation in bacteria: A model. Journal of molecular biology 1995;253:739-748.

[63] Norris V, Raine DJ: A fission-fusion origin for life. Origins of Life and Evolution of the Biosphere 1998;28:523-537.

[64] Onoda T, Enokizono J, Kaya H, Oshima A, Freestone P, Norris V: Effects of calcium and calcium chelators on growth and morphology of escherichia coli l-form nc-7. Journal of bacteriology 2000;182:1419-1422.

[65] Ozbudak EM, Thattai M, Lim HN, Shraiman BI, van Oudenaarden A: Multistability in the lactose utilization network of escherichia coli. Nature 2004;427:737-740.

[66] Palkova Z: Multicellular microorganisms: Laboratory versus nature. EMBO Reports 2004;5:470476.

[67] Palmen R, Vosman B, Buijsman P, Breek CK, Hellingwerf KJ: Physiological characterization of natural transformation in acinetobacter calcoaceticus. J Gen Microbiol 1993;139:295-305.

[68] Pfaffmann JO, Conrad M: Adaptive information processing in microtubule networks. Bio Systems 2000;55:47-57.

[69] Posfai G, Plunkett III G, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR: Emergent properties of reduced-genome escherichia coli. Science 2006;312:1044-1046.

[70] Possoz C, Filipe SR, Grainge I, Sherratt DJ: Tracking of controlled escherichia coli replication fork stalling and restart at repressor-bound DNA in vivo. The EMBO journal 2006;25:2596-2604.

[71] Rauch EM, Bar-Yam Y: Theory predicts the uneven distribution of genetic diversity within species. Nature 2004;431:449-452.

[72] Ripoll C, Norris V, Thellier M: Ion condensation and signal transduction. BioEssays 2004;26:549-557.

[73] Rocha E, Fralick J, Vediyappan G, Danchin A, Norris V: A strand-specific model for chromosome segregation in bacteria. Molecular microbiology 2003;49:895-903.

[74] Rothemund PW, Papadakis N, Winfree E: Algorithmic self-assembly of DNA sierpinski triangles. PLoS biology 2004;2:e424.

[75] Sato K, Ito Y, Yomo T, Kaneko K: On the relation between fluctuation and response in biological systems. Proceedings of the National Academy of Sciences of the United States of America 2003;100:14086-14090.

[76] Siddiqui RA, Hoischen C, Holst O, Heinze I, Schlott B, Gumpert J, Diekmann S, Grosse F, Platzer M: The analysis of cell division and cell wall synthesis genes reveals mutationally inactivated ftsq and mray in a protoplast-type l-form of escherichia coli. FEMS microbiology letters 2006;258:305-311.

[77] Skretas G, Wood DW: A bacterial biosensor of endocrine modulators. Journal of molecular biology 2005;349:464-474.

[78] Szostak JW, Bartel DP, Luisi PL: Synthesizing life. Nature 2001;409:387390.

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

[80] Thomas R: On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. Series in Synergies 1980;9:180-193.

[81] Tolker-Nielsen T, Holmstrom K, Boe L, Molin S: Non-genetic population heterogeneity studied by in situ polymerase chain reaction. Molecular microbiology 1998;27:1099- 1105.

[82] Tsuda S, Zauner K-P, Gunji Y-P: Robot control: From silicon circuitry to cells.; in Ijspeert AJ, Masuzawa T, Kusumoto S, (eds): Lecture notes in computer science 2006, vol 3853, pp. 2032.

[83] Vohradsky J, Ramsden JJ: Genome resource utilization during prokaryotic development. FASEB Journal 2001;15:2054-2056.

[84] Wall ME, Hlavacek WS, Savageau MA: Design of gene circuits: Lessons from bacteria. Nature Review Genetics 2004;5:3442.

[85] Wang JD, Sanders GM, Grossman AD: Nutritional control of elongation of DNA replication by (p)ppgpp. Cell 2007;128:865-875.

[86] Wang Q, Suzuki A, Mariconda S, Porwollik S, Harshey RM: Sensing wetness: A new role for the bacterial flagellum. EMBO Journal 2005;24:2034-2042.

[87] You L, Cox RSr, Weiss R, Arnold FH: Programmed population control by cell-cell communication and regulated killing. Nature 2004;428:868-871.

[88] Yu W, Sato K, Wakabayashi M, Nakaishi T, Ko-Mitamura EP, Shima Y, Urabe I, Yomo T: Synthesis of functional protein in liposome. Journal of bioscience and bioengineering 2001;92:590-593.

From metabolic hyperstructures to DNA replication complexes and back again

Anne-Francoise Monnier[1], Armelle Cabin[2,3], Guillaume Legent[2,3], Danielle Canceill[1], Emmanuelle Le Chatelier[1], Camille Ripoll[2,3], Michel Thellier[2,3], Vic Norris[2,3] and Laurent Janniere[1]

[1] Laboratoire de Genetique Microbienne, INRA, Jouy en Josas, France
[2] Laboratoire d'Assemblages moléculaires: modélisation et imagerie SIMS, FRE CNRS 2829, Faculté des Sciences de l'Université de Rouen, 76821 Mont Saint Aignan Cedex, France
[3] Epigenomics Project, Genopole®, F-91000 Evry, France

Recent discovery in *Bacillus subtilis* of a dependence of the rate of elongation of chromosome replication on the activity of the enzymes in central carbon metabolism may be interpreted in the light of hyperstructure dynamics. This dependence may be related to signals in the form of enzymes and metabolites due to the properties of functioning dependent hyperstructures. Communication between metabolism and chromosome replication is probably two-way with hyperstructure-mediated information flowing in both directions. Here, we review evidence for the involvement of metabolic enzymes in DNA replication and for the existence of glycolytic and replication hyperstructures. We suggest that obtaining an integrated picture of the cell based on hyperstructure dynamics can only be achieved via interdisciplinary collaborations.

Introduction

Since the 1960's, it has been well established that DNA replication, like many other processes in bacteria, varies according to the richness of available nutrients (Bremer & Dennis, 1987). This is part of the mechanism of growth rate control which also determines the proportion of cellular mass devoted to the transcriptional and translational machinery. Despite intensive investigations, the nature of the link between DNA replication and growth rate has remained largely unknown. This is an indication of the complexity of this link which presumably results from a co-evolution of central carbon metabolism (CCM) and replication since the emergence of the original DNA-containing cells or even earlier, that is, since the prebiotic world (Forterre, 2005; Hunding et al., 2006; Koonin, 2006).

There is increasing evidence that, in both eukaryotic and prokaryotic cells, enzymes and other macromolecules are organized into - and function as - higher order structures. These may take the form of metabolons in metabolism (Alvarez et al., 2005; Jorgensen et al., 2005; Shmelev & Serebrenikova, 1997;

Srere, 1994; Velot et al., 1997) or transducons in signaling (Trewavas & Malho, 1997)and may themselves be associated in still higher order hyperstructures (Guzman et al., 2002; Le Sceller et al., 2000; Molina & Skarstad, 2004; Norris et al., 2007a; Norris et al., 2007b). These hyperstructures include the filaments formed from enzymes such as RNaseE and enolase (Taghbalout & Rothfield, 2007) as well as EF-Tu (Mayer, 2006). Such evidence for a (quasi-) solid-state metabolism raises a number of fundamental questions. One of them is whether a language exists at the level of hyperstructures. The link between metabolism and DNA replication may be evidence of just such a language, but what form does it take?

In this review, we bring together information on the metabolism-replication link and on the possible involvement in this link of the often overlooked non-metabolic functions of metabolic enzymes and of hyperstructures. Because of the complexity of the system, its understanding requires a combination of approaches from Systems Biology, physical chemistry and microbiology. This combination is outlined below.

1 Evidence for a metabolism-replication link

In the years 1960-1980, the synthetic activities of *Escherichia coli* cells growing in steady- state conditions in different carbon sources were analysed in detail by numerous teams. From these studies, reviewed by Helmstetter (Helmstetter, 1996), and from a more recent analysis (Michelsen et al., 2003), it becomes clear that DNA synthesis varies as a function of cell metabolism. This variation is achieved in two ways. In cells grown in rich media at rates ranging from about 25 to 70 minutes, the modulation of DNA replication results from a control of the initiation frequency (from 3 to 1 initiations of replication per cell cycle). In poorer media, the time required to replicate the chromosome progressively increases from 40 to 200 minutes suggesting that regulation of replication in slow growing cells is exerted by slowing down the rate of chain elongation by polymerases from about 500 to 100 nucleotides per second. A similar correlation probably also occurs in *Bacillus subtilis* ((Sharpe et al., 1998), our unpublished data) and other bacteria. The mechanism responsible, however, remains elusive.

The tight link between metabolism and replication is also obvious when bacterial cells are abruptly depleted of amino-acids. In *E. coli*, such nutritional shifts inhibit plasmid replication (Wegrzyn & Wegrzyn, 2002) and initiation of new round of chromosomal replication (Lark et al., 1963; Levine et al., 1991). Amino acid starvation induces the stringent response and a transient accumulation of (p)ppGpp, a global regulator derived from GTP (Magnusson et al., 2005). While it was shown that (p)ppGpp is the key determinant in inhibition of chromosomal initiation in starved cells (Schreiber

et al., 1995), its mechanism of action is unknown. The stringent response also inhibits replication in *B. subtilis* (Levine et al., 1991; Wang et al., 2007). However, contrary to *E. coli*, replication inhibition in this Gram-positive bacteria occurs at the level of DNA elongation likely by direct inhibition of the essential DnaG primase (which produces RNA primers in the replisome) by (p)ppGpp (Wang et al., 2007). While (p)ppGpp is a key determinant in replication arrest during amino-acid depletion, this nucleotide cannot be so important in modulation of initiation and elongation of replication in bacteria growing in steady state as metabolic-driven modulation of replication persists in *E. coli* mutants unable to produce ppGpp (Hernandez & Bremer, 1993).

In eukaryotic cells, nutrients also modulate DNA replication (Bohnsack & Hirschi, 2004). Glucose stimulates the G1 to S transition in breast cancer cells (Okumura et al., 2002). It also stimulates initiation and elongation of DNA synthesis in SV40 and in HeLa cells grown under hypoxia (Riedinger et al., 2001). In *Saccharomyces cerevisiae*, two independent studies showed that S phase occurs during the reduction phase of an internal cycle of oxido-reduction (Klevecz et al., 2004; Tu et al., 2006). The variation in reactive oxygen species that naturally takes place during cell cycle progression controls entry in S phase in human cells (Havens et al., 2006). Finally, in yeast, a thermosensitive (Ts) mutation in MCM1, a transcriptional factor that stimulates initiation of DNA replication by binding in the vicinity of replication origins, is suppressed by mutations in glycolytic genes (Chang et al., 2003; Chang et al., 2004; Chen & Tye, 1995). From a more general point of view, a relationship between metabolism and replication is also revealed in carcinogenesis where early events include a profound alteration of glycolysis (the Warburg effect) and a major reduction in DNA replication fidelity and genome stability (Gatenby & Gillies, 2004; Loeb et al., 2003).

Factors that might be involved in the metabolism-replication link in *B. subtilis* steady-state cells have been uncovered recently (Janniere et al., 2007). These factors include the activity of the bottom part of glycolysis and three enzymes of the replisome, namely the lagging strand DNA polymerase DnaE, the helicase DnaC and the primase DnaG. Additional data suggest that the link might rely on metabolic-driven changes in the properties of replication enzymes. These changes have been proposed to be orchestrated by signals generated by glycolytic activity and sensed by replication enzymes. Whether the metabolism-replication relationship involves the same key determinants in bacteria other than *B. subtilis* is under investigation.

To our knowledge, all the data obtained so far show an influence of metabolism on DNA replication. However, the converse relationship might also exist. Indeed, by adjusting the rate of DNA synthesis to metabolism, bacterial cells ensure that a particular gene is duplicated at a particular time in the cell

cycle. In this way, gene duplication by replication might stimulate particular metabolic activities at particular times. We are currently using a stochastic automaton, HSIM (Amar et al., 2004), to explore this possibility (see below).

2 Metabolic enzymes and replication

Although the link between metabolism and replication has been known for a long time, its molecular mechanism is still elusive. This is an indication of the likely complexity of this link. Another indication of complexity is that the link presumably results from a long period of co-evolution of central carbon metabolism (CCM) and replication (Forterre, 2005; Hunding et al., 2006; Koonin, 2006). Data obtained in studies of *B. subtilis* and *S. cerevisiae* argue for a molecular mechanism based on metabolism-driven conformation changes in replication enzymes (Chen & Tye, 1995; Janniere et al., 2007). In principle, these conformational changes might be caused by processes as different as allosteric regulation, covalent modifications and protein-protein interactions. To our knowledge, no solid example of allosteric regulation of replication enzymes is available in the literature. In contrast, several studies indicate that replication enzymes might interact (directly or indirectly) with metabolic enzymes (Butland et al., 2005; Noirot-Gros et al., 2002) and an impressive number of studies reveal the importance of covalent modifications such as phosphorylation and acetylation in the regulation of the replication enzymes activities of eukaryotic cells (Bell & Dutta, 2002; Forsburg, 2004; Henneke et al., 2003). In prokaryotes, interest in protein modification continues to increase (Deutscher & Saier, 2005; Grangeasse et al., 2007; Mijakovic et al., 2005) and several studies aimed at describing the bacterial phosphoproteome have been performed (Bendt et al., 2003; Cortay et al., 1986; Eymann et al., 2007; Freestone et al., 1995; Levine et al., 2006). Recently, *B. subtilis SSB*, a single-stranded DNA-binding protein that is required for DNA repair, recombination and replication, has been shown to be phosphorylated on tyrosine and the protein kinase and phosphatase regulating this modification, identified (Mijakovic et al., 2006). Data suggest that this phosphorylation is a conserved process in distant bacteria (Mijakovic et al., 2006). Interestingly, SSB kinase inactivation alters initiation of DNA replication and cell cycle (Petranovic et al., 2007). Taken together, these data suggest that protein modification in bacteria might be of greater importance in cell cycle regulation than some have thought (Norris et al., 1994). Whether these modifications can be somehow connected to cell metabolism remains to be established.

CCM enzymes appear to be tightly linked to DNA metabolism in both prokaryotes and eukaryotes. For instance, an *in silico* analysis of fully sequenced bacterial genomes revealed a highly conserved physical linkage be-

tween two glycolytic enzymes (pyruvate kinase and phosphofructokinase) and the replicative DNA polymerase DnaE (Overbeek et al., 1999). Such conserved neighbourhood is thought to indicate related functions (Dandekar et al., 1998; Overbeek et al., 1999). The stimulation of histone H2B gene expression in human cells, which is essential for S phase progression, strictly depends on glyceraldehyde-3-phosphate dehydrogenase; in this progression, glyceraldehyde-3-phosphate dehydrogenase may regulate the activity of the transcriptional co-activator OCA-S, with which it forms a complex, by sensing the NAD+/NADH redox status (Zheng et al., 2003); this pathway of regulation of H2B gene expression may also involve lactate dehydrogenase. Aconitase, an enzyme of the tricarboxylic acid cycle, is a scaffolding protein of mitochondrial DNA. By assisting mtDNA organisation, this enzyme might couple mtDNA inheritance with cellular metabolism (Chen & Butow, 2005; Chen et al., 2005; Chen et al., 2007). Phosphoglucomutase, a glycolytic enzyme, and transketolase, a pentose phosphate pathway enzyme, modulate chromosome supercoiling in *E. coli* with transketolase being implicated in the formation of a scaffold (Hardy & Cozzarelli, 2005). In *B. subtilis*, the helicase (DnaC) and the primase (DnaG) were shown to interact in a yeast two hybrid screen with subunits of the related pyruvate dehydrogenase, branched–ketoacid dehydrogenase and acetoin dehydrogenase. The binding of the primase to the E2 subunit of the pyruvate dehydrogenase is thought to cause inhibition of replication in a membrane-based *in vitro* assay (Noirot-Gros et al., 2002; Stein & Firshein, 2000). In human cells, at least three glycolytic enzymes (glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and lactate dehydrogenase), possibly in a modified form, are present in the nucleus (in addition to the cytosol) (Berry, 2004; Laschet et al., 2004; Ronai, 1993; Zheng et al., 2003) where they may influence DNA replication; indeed, *in vitro* studies showed that these enzymes bind to single-stranded DNA and modulate the activity of the replicative polymerases PolA, PolD and PolE (Baxi & Vishwanatha, 1995; Grosse et al., 1986; Popanda et al., 1998; Ronai, 1993); in the case of PolA, the modulation of activity involves PGK binding to the polymerase possibly in association with annexin II (Jindal & Vishwanatha, 1990); *in vivo*, this binding might play a role in DNA synthesis as (i) both enzymes are associated with the nuclear matrix of Hela cells in S phase and (ii) reduction of annexin II, and to a lesser extent, PGK synthesis, reduces ongoing DNA synthesis (Kumble et al., 1992; Vishwanatha et al., 1992). Since the glycolytic enzymes may retain their ability to bind metabolic substrates in the nucleus, it was proposed that they might act as signal transducers, connecting cell metabolism to DNA replication (Popanda et al., 1998; Riedinger et al., 2001).

3 Hyperstructures

3.1 DNA replication hyperstructure

In *Bacillus subtilis*, the hyperstructure responsible for chromosome replication includes a replisome containing two different polymerases for leading (PolC) and lagging (DnaE) strand polymerization, PolC and DnaE, respectively (Dervyn et al., 2001). The replisome also contains: i/ the b clamp that encircles and slides along the DNA; this protein is put on the DNA by the proteins DnaX, d and d9, ii/ the initiation factors DnaA (or PriA), DnaB, DnaD and DnaI, iii/ the DnaC helicase which melts the DNA duplex, and iv/ the primase DnaG which generates the RNA primers used by polymerases for strand synthesis (for references see (Janniere et al., 2007)). This hyperstructure probably also includes a domain in the phospholipid membrane (for references see (Firshein, 1989; Norris et al., 2007b)). Indeed, the origin region of *B. subtilis*, like that of *E. coli*, is enriched in membrane fractions, and in *B. subtilis* it has been proposed that the loading of the DnaC helicase (see above) requires a membrane-mediated interaction between DnaB and DnaD that constitutes a spatial means to regulate initiation (Rokop et al., 2004). It also seems likely that, as proposed for *E. coli*, the replication hyperstructure contains the enzymes needed for DNA repair, recombination and synthesis of precursors (for references see (Norris et al., 2007b)). Note that ongoing replication requires feeding the hyperstructure with the four deoxyribonucleotides (dNTPs) at the rate of about 3000 nucleotides per second, yet, despite this high rate, there are only sufficient dNTPs for half a minute of synthesis. It is therefore not surprising that there is evidence in both eukaryotes and prokaryotes for the presence of ribonucleoside diphosphate reductase, which catalyses the synthesis of the dNTPs, in the hyperstructure (Guarino et al., 2007; Guzman et al., 2002; Mathews, 1988). Several mechanisms exist in *E. coli* to ensure that once initiation has occurred it does not immediately recur (Leonard & Grimwade, 2005) and, for example, the SeqA protein binds to newly replicated DNA to form extended replication/sequestration hyperstructures (Ohsumi et al., 2001) that may even contain the genes encoding the enzymes in the hyperstructure (for references see (Norris et al., 2000; Norris et al., 2007b)).

3.2 Glycolytic hyperstructures

Proteins involved in metabolic pathways are often reported as existing in the form of multi- molecular assemblies or metabolons that allow some form of channeling of intermediates to occur (Mowbray & Moses, 1976; Srere, 1987; Srivastava & Bernhard, 1986; Velot et al., 1997). Such metabolons are generally envisaged as containing only a few enzymes that act on successive intermediates in a pathway which are channelled from one enzyme

to the next. One example is the multienzyme complex of tricarboxylic acid cycle enzymes, which catalyse the consecutive reactions from fumarate to 2-oxoglutarate, in *Pseudomonas aeruginosa* (Mitchell, 1996). Another example is the huge, extracellular polycellulosomes produced to degrade plant walls by bacteria such as *Clostridium cellulovorans* (Doi & Kosugi, 2004).

Glycolysis, gluconeogenesis, the pentose phosphate pathway and the tricarboxylic acid pathway are major elements of the central carbon metabolism in which nutrients are metabolised. Glycolysis is a nine reaction pathway that, conceptually, is in two parts. In the first part, glucose 6-phosphate is converted into glyceraldehyde-3-phosphate. This set of reactions is shunted by the pentose phosphate pathway. In the second part, glyceraldehyde-3-phosphate is transformed into pyruvate. These reactions, the three-carbon part of glycolysis, are central to metabolism and cannot be shunted in most organisms. The gluconeogenesis pathway uses most of the glycolytic reactions in the opposite direction to produce glucose 6- phosphate.

In the case of glycolysis, there is reasonable evidence for an extensive glycolytic metabolon (Srere, 1994) although it must be conceded that here has been controversy over the existence of glycolytic metabolons in prokaryotes and eukaryotes (Cornish-Bowden & Cardenas, 1993; Mendes et al., 1995; Ovadi & Saks, 2004). In eukaryotic cells, interactions between sequential pairs of glycolytic enzymes have been demonstrated, with glycolytic enzymes being partitioned reversibly between cytoplasmic and cytomatrix-bound states depending on physiological conditions (for references see (Welch & Easterby, 1994)) or indeed confined to an organelle, the glycosome, in trypanosomes (Kessler & Parsons, 2005). In *E. coli*, the glycolytic pathway has been isolated as an equimolar multi-enzyme complex with compartmentation of substrates; one such complex was reported to have a molecular mass of 1.65 megadaltons, similar to that calculated for an equimolar complex of the enzymes of glycolysis, and had a particle diameter of 30-40 nm (Gorringe & Moses, 1978; Mowbray & Moses, 1976). Finally, the full enzymatic activity of the glycolytic enzymes glyceraldehyde-3- phosphate dehydrogenase, phosphoglycerate mutase and enolase results from their homo- oligomeric association, supporting the idea that a single species of enzyme can be activated to oligomerize by substrate (Torshin, 1999); such association could help nucleate and stabilize a hyperstructure.

There is also evidence that enzymes may be assembled into large hyperstructures, as in the case of cellulolytic enzymes in bacteria such as *Acetivibrio cellulolyticus* (Doi & Kosugi, 2004) or associated with the tubulin cytoskeleton, as in the case of many glycolytic enzymes (Walsh et al., 1989), or associated with sacromeres as in the case of six glycolytic enzymes in *Drosophila* (Sullivan et al., 2003). The actual advantages - or lack of them

- conferred by metabolons and hyperstructures on metabolic efficiency are unclear.

Another important aspect to consider is the extent to which enzymic hyperstructures are assembled in response to the cell's need for them. Certain transient, dynamic multi-molecular assemblies only form in an activity-dependent manner (Norris et al., 1999; Ovadi, 1988; Winkel, 2004) due, for example, to an association between enzymes that only occurs when they are engaged in transporting or transforming substrates or transducing a signal. These assemblies are *functioning-dependent structures*, FDSs; this is because an FDS assembles when functioning and disassembles when no longer functioning and thus is created and maintained by the very fact that it is in the process of accomplishing a task (Thellier et al., 2004). (In its most useful form, an FDS is a scale-free concept.) In this hypothesis, a functioning-dependent sugar transporter-glycolytic hyperstructure might be expected to have its metabolic activity maximized by active, enzyme-promoted, association of membrane and cytoplasmic constituents because: i) the multiple interactions involved in hyperstructure formation would help maintain the hyperstructure during fluctuations in substrate supply, ii) enzyme association due to substrate-induced binding might select the appropriate transporter from a competing population, during, for example, diauxic growth of a bacterium on two substrate sugars, iii) the dissipative nature of the structure would imply that when the substrate was completely exhausted, the membrane domain would disperse, and the cytoplasmic structure would dissociate to free the space for other structures (Norris et al., 1999). In eukaryotes, numerous examples of FDSs exist whilst in *E. coli* there is the example of the promotion by substrate binding of the assembly of the three components of protein-mediated transporter responsible for protein secretion (Letoffe et al., 1996).

4 A choice of mechanisms for the metabolism/replication link

In the vision of the bacterial cell as an ensemble of hyperstructures, interactions between hyperstructures create the phenotype (Norris et al., 2007a). But what is the nature of these interactions? What are the mechanisms that allow hyperstructures to communicate? There are several possibilities. As mentioned above, at the level of macromolecules, an FDS hyperstructure forms only if some of its constituent enzymes have captured their substrate and are performing their tasks. When these substrates run out, the FDS disassociates. A glycolytic hyperstructure of the FDS type might disassemble in the absence of a flow of metabolites to release enzymes that, when free or degraded (to give fragments), might inhibit replication. In fact, a wide variety of types of signals in the form of enzymes or metabolites can be generated by FDSs (Thellier et al., 2006). The many ways a glycolytic hyperstructure might

signal to a replication one include (for references see (Norris et al., 2007a)): (1) the free glycolytic enzymes might have kinase activities and phosphorylate constituents of the replication hyperstructure, (2) free or hyperstructure-bound glycolytic enzymes might compete as substrates for kinases that phosphorylate replication enzymes, (3) a changing PTS-glycolytic hyperstructure might release certain phospholipids or proteins and thereby change the microviscosity or domain distribution of the membrane, (4) in a tensegrity approach (Ingber, 1997), a glycolytic hyperstructure might be under mechanical tension and, a change in this due to a change in substrate concentration might affect the tension of a replication hyperstructure, (5) changes in the charge density of a glycolytic hyperstructure might result in ion condensation or decondensation to affect the availability of ions for a replication hyperstructure, (6) alterations and even oscillations in the dynamics of a glycolytic hyperstructure might affect water structures and hence affect a replication hyperstructure etc. In the case of the link between metabolism and replication, we focus here on one speculative form of communication between hyperstructures: that involving CCM enzymes and the metabolites on which they act.

The E2 subunit of the pyruvate dehydrogenase complex, dihydrolipoamide acetyltransferase, is a membrane-associated inhibitor of DNA replication in *B. subtilis* (Stein & Firshein, 2000). Also in *B. subtilis*, the primase and the helicase (key enzymes in replication) appear to interact directly with metabolic enzymes such as pyruvate dehydrogenase (Janniere et al., 2007; Noirot-Gros et al., 2002). Moreover, three *B. subtilis* enzymes known or proposed to act on the lagging-strand template in the replicating fork (the DNA polymerase DnaE, the helicase and the primase) are functionally connected to the five terminal reactions of glycolysis (Janniere et al., 2007; Stein & Firshein, 2000). Many metabolic enzymes can be abundant and, since the synthesis of abundant enzymes almost necessarily entails the formation of a transcription-translation hyperstructure, it is conceivable that the high concentration of enzymes newly released from such a hyperstructure might drive their assembly into an adjoining metabolic hyperstructure. Alternatively, if the substrate becomes limiting, such enzymes would fail to be incorporated into a metabolic hyperstructure of the FDS type (where they might be safe from proteases) and could then be degraded. Degradation of glucose-6- phosphate dehydrogenase, an enzyme in the pentose phosphate pathway, gives a pentapeptide that acts as an important extracellular signal (Kolodkin-Gal et al., 2007) and establishes a precedent for the role of degradation of metabolic enzymes in intracellular signalling.

In the context of signalling by metabolites or small molecules, note that when a translating ribosome meets an uncharged tRNA in a transcription-translation hyperstructure, (p)ppGpp is produced (see above), which is a major inhibitor of chromosome replication in both *E. coli* and *Bacillus subtilis*

(Schreiber et al., 1995) (Levine et al., 1991; Wang et al., 2007). Amino acids provide a precedent for metabolites involved in signalling. For example, leucine, which is the most abundant amino acid in proteins (constituting 9% of amino acids) acts via the leucine- responsive regulatory protein to affects transcription of at least 10% of all *E. coli* genes including most of those expressed on entry into stationary phase (Brinkman et al., 2003).

5 A reciprocal link from replication to metabolism?

It would be very surprising if information were only to flow in one direction, from metabolism to replication. There are a number of circumstantial arguments that can be made that the metabolism-replication communication is likely to be a dialogue than a monologue. Firstly, the existence of two identical chromosomes within the same cell in which genes and higher order structures compete for limited resources is a symmetry-breaking situation previously proposed to lead to differentiation (Norris & Madsen, 1995). In this overall sense, replication affects phenotype. Secondly, as discussed above, recent experiments are consistent with an intimate relationship between metabolism and the rate of chromosome replication in bacteria (Janniere et al., 2007; Wang et al., 2007). In *E. coli*, the velocity of the replication fork may vary from about 1000 to 200 nt/s as a function of the energy contained in the nutrients (Helmstetter, 1991; Michelsen et al., 2003). Thirdly, the process of chromosome replication progressively changes the copy number of genes and sites in a linear order. This raises the possibility that slowing or even pausing replication for different times at different sites in the chromosome might be combined with various mechanisms leading to local cooperation and global competition. If so, such replication-phenotype coupling could produce different patterns of gene expression (Norris et al., 2007c). Indeed, replication-phenotype coupling may constitute a powerful and fundamental way of generating coherent phenotypes. Fourthly, studies of 'combed' chromosomes from a mutant of *E. coli* synchronised for replication reveal a heterogeneity in the pattern of replication, consistent with different rates of replication in different regions (Breier et al., 2005). Fifthly, sequences that slow or halt replication have been found in both *E. coli* and *B. subtilis*. In *E. coli*, a polar DNA replication barrier is formed when the DNA-binding protein Tus forms a complex with any of the four 23-base-pair terminator (ter) sites found in the terminus region of the chromosome (in addition to other systems (Guijo et al., 2001)). In *B. subtilis*, a replication barrier exists near the origin of replication and arrest is dependent upon the RelA protein, the action of which is correlated with high levels of the alarmone, ppGpp (Autret et al., 1999) but see (Mulugu et al., 2001).

6 Investigative tools from Systems Biology

Simulation with stochastic automata and multi-agent systems is an attractive alternative to differential equations for studying the diffusion and interaction of the many different enzymes and metabolites in cells (Ballet et al., 2004; Kier et al., 2005). A stochastic automaton, HSIM (for Hyperstructure Simulator), has been developed used to determine quantitatively the effects of metabolon and hyperstructure formation on a modified version of the PTS and glycolysis (Amar et al., 2008). This 3-D virtual bacterium showed that assembly of metabolons into a hyperstructure could generate gradients of metabolites that could serve as signals.

Using Ordinary Differential Equations (ODEs), it can be shown how the assembly of enzymes into higher order structures might confer new properties (Marmillot et al., 1992). By taking the simplest of cases, the functioning-dependent association of two sequential monomeric enzymes, we found that kinetics can be generated by such association that cannot be generated by the individual enzymes alone. These kinetic characteristics include the full range of signals found in electronic circuits such as linearity, invariance, pulsing and switching (Thellier et al., 2006).

7 Discussion

The establishment of a link between the central carbon metabolism and the elongation step in DNA replication opens up new perspectives in prokaryotic and eukaryotic biology.

Firstly, the widespread existence of non-metabolic functions of CCM enzymes becomes easier to understand.
Ambiquitous enzymes can occupy two different positions in the cell where they may, perhaps, have different functions (Nemat-Gorgani & Wilson, 1980). Such enzymes may be able to generate waves of metabolites that could act as signals (Marmillot et al., 1992). Conceptually related, functioning-dependent structures, FDSs, only assemble when their constituents are performing a task and disassemble when the task is over (Thellier et al., 2004). An example of an FDS, in the case of the PTS and glycolysis, would be a metabolite-induced metabolon (Norris et al., 1999).The non-metabolic functions of CCM enzymes may correspond (Thellier, 2003; Thellier et al., 2004) to their being free rather than being in a metabolon or a larger hyperstructure; hence, the fact of a CCM enzyme being free may give information about a reduction in carbon flow. Given a higher probability of degradation for free enzymes, fragments of CCM enzymes might constitute signals to the replication machinery.

Secondly, the relationship between metabolism and DNA replication may afford new insights at the level of interaction between hyperstructures.
Evidence continues to accumulate that, in eukaryotes (Hartwell et al., 1999) as well as in prokaryotes (Norris et al., 2007a), the phenotype of the cell is determined at the level of dynamic multi-molecular assemblies alias hyperstructures. These hyperstructures include those comprising EF-Tu (Mayer, 2003; Mayer, 2006), RNaseE and enolase (Taghbalout & Rothfield, 2007), glycolytic (Gorringe & Moses, 1978; Mowbray & Moses, 1976), PTS (Patel et al., 2004) and TCA cycle enzymes (Velot et al., 1997), RNA polymerase and ribosomes (Cabrera & Jin, 2003), SecA (Campo et al., 2004), recombination enzymes (Kidane et al., 2004; Kidane & Graumann, 2005) and enzymes responsible for amino acid biosynthesis (for references see (Bailly et al., 2007; Norris et al., 2007b)). In the perspective of a metabolism (Norris et al., 1996; Srere, 1994) and a signal transduction system based on hyperstructures, these hyperstructures, particularly functioning-dependent hyperstructures, might signal to another. Insofar as the metabolism-replication link is based on glycolytic and replication hyperstructures, it gives an opportunity to investigate such signalling.

Thirdly, discovery of the physical nature of the signal transduction responsible for the metabolism-replication link may provide a deeper understanding of signal transduction.
It has been customary to separate metabolism from signal transduction. Our findings suggest that, in principle, almost any metabolic pathway in which the enzymes can associate in a functioning-dependent manner can generate signals. Previously, we have shown how, in steady state conditions, a model system of two sequential monomeric enzymes that can associate into FDSs can generate signals of metabolites in the form of linearity pulsing and switching (Thellier et al., 2006). More recently, in non-steady state conditions when the supply of substrate can vary, we have shown how an FDS can encode the initial concentration of the first substrate as the length of time a pulse of a final product stays above a threshold (Legent et al., submitted). The lesson here is that the assembly and disassembly of metabolic enzymes is a potentially powerful way of generating signals that, in the case of CCM enzymes, may have been available to cells early in their evolution. To paraphrase McLuhan, 'the signalling structure is the message'.

Fourthly, we may expect the relationship between metabolism and replication to be a two-way relationship. This would help create a new vision of the coherent exploration of phenotype space by variations in DNA replication.
A fundamental problem in biology is how cells integrate gene expression and environmental conditions to steer their phenotypes in a coherent, reproducible way through a vast space of apparent possibilities. One class of solutions lies

in the fact that two chemically identical chromosomes in the same cytoplasm may spontaneously adopt different patterns of gene expression (Norris & Madsen, 1995). The basic idea is that if a gene attracts an RNA polymerase it has a greater chance of attracting a second one and hence, if two identical copies of a gene compete for a limited number of RNA polymerases, one copy is expressed whilst the other falls silent. The idea can be taken much further: by slowing or accelerating the elongation step of DNA replication in different regions of the genome, a bacterium could generate different patterns of copy numbers of different genetic elements and, given local synergistic relationships and global competitive ones, this might then result in different, coherent, phenotypes in daughter cells (Norris et al., 2007c). In this context, the link from CCM to replication becomes particularly attractive if, reciprocally, changing the rate of replication feeds back into expression of the genes encoding the CCM enzymes. This would lead to convergence on, and stabilisation of, certain phenotypes.

Finally, opening up these new perspectives requires interdisciplinarity. Just how does the cell manage to function as a coherent, integrated whole? This can be illustrated from the hyperstructure perspective in the following scenario. Imagine a bacterium growing on a medium containing both amino acids and glucose. Large coupled transcription- translation-insertion (transertion) hyperstructures produce proteins for the membrane(s) thereby dynamically coupling DNA to the cell envelope. Coupled transcription-translation hyperstructures also produce CCM enzymes that may then be incorporated into neighbouring, functioning-dependent, PTS-glycolytic hyperstructures which metabolise glucose. Concentrations of available amino acids and glucose (along with concentrations of signalling molecules from competitors) are constantly fluctuating. Imagine now the amino acid levels drop. At the hyperstructure level, the bacterial responses might take the form of (1) the transertion hyperstructures fail to deliver sufficient CCM enzymes to maintain previous levels of PTS-glycolytic hyperstructures, (2) this change in the dynamics of PTS-glycolytic hyperstructures leads to alterations in metabolic intermediates and free enzymes, (3) these alterations constitute signals that are sensed by the replication hyperstructure and alter elongation rates, and (4) altering replication rates alters the pattern of hyperstructures distributed to daughter cells so that the bacterial population adapts to the fluctuations in its environment. There is much more to this putative hyperstructure scenario than this. Many hyperstructures are tensegrity structures (Ingber, 1998) and respond to changes in the tension generated by alterations in the activity of polymerases and ribosomes; these responses include changes to frequencies of rhythms of condensation and decondensation of ions accompanied by variations in water structure (for references see (Norris et al., 2007a)). Understanding which of these changes makes physical - and biological - sense

depends on the existence of research environments to bring together the right mix of specialists in an atmosphere that is both friendly and demanding. The Epigenomics Project has been a model in this area.

Acknowledgements

For support, we thank the Epigenomics Project, Genopole , INRA and the Normandy Region.

References

[1] Alvarez, B. V., Vilas, G. L. & Casey, J. R. (2005).
Metabolon disruption: a mechanism that regulates bicarbonate transport. *The EMBO journal* **24**, 2499-2511.

[2] Amar, P., Bernot, G. & Norris, V. (2004).
HSIM: a simulation programme to study large assemblies of proteins. *Journal of Biological Physics and Chemistry* **4**, 79-84 http://www.lri.fr/ pa/Hsim/JBPC_04.pdf.

[3] Amar, P., Legent, G., Thellier, M., Ripoll, C., Bernot, G., Nystrom, T., Saier, M. H. J. & Norris, V. (2008).
A stochastic automaton shows how enzyme assemblies may contribute to metabolic efficiency. *BMC systems biology*, in press.

[4] Autret, S., Levine, A., Vannier, F., Fujita, Y. & Seror, S. J. (1999).
The replication checkpoint control in Bacillus subtilis: identification of a novel RTP-binding sequence essential for the replication fork arrest after induction of the stringent response. *Molecular microbiology* **31**, 1665-1679.

[5] Bailly, M., Blaise, M., Lorber, B., Becker, H. D. & Kern, D. (2007).
The transamidosome: a dynamic ribonucleoprotein particle dedicated to prokaryotic tRNA-dependent asparagine biosynthesis. *Molecular cell* **28**, 228-239.

[6] Ballet, P., Zemirline, A. & Marcé, L. (2004).
The BioDyn language and simulator. Application to an immune response and *E. coli* and phage interaction. *Journal of Biological Physics and Chemistry* **4**, 93-101.

[7] Baxi, M. D. & Vishwanatha, J. K. (1995).
Uracil DNA-glycosylase/glyceraldehyde-3-phosphate dehydrogenase is an Ap4A binding protein. *Biochemistry* **34**, 9700-9707.

[8] Bell, S. P. & Dutta, A. (2002).
DNA replication in eukaryotic cells. *Annual review of biochemistry* **71**, 333-374.

[9] Bendt, A. K., Burkovski, A., Schaffer, S., Bott, M., Farwick, M. & Hermann, T. (2003).
Towards a phosphoproteome map of Corynebacterium glutamicum. *Proteomics* **3**, 1637-1646.

[10] Berry, M. D. (2004).
Glyceraldehyde-3-phosphate dehydrogenase as a target for small-molecule disease-modifying therapies in human neurodegenerative disorders. *J Psychiatry Neurosci* **29**, 337-345.

[11] Bohnsack, B. L. & Hirschi, K. K. (2004).
Nutrient regulation of cell cycle progression. *Annual review of nutrition* **24**, 433-453.

[12] Bremer, H. & Dennis, P. P. (1987).
Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella*, pp. 1527-1542. Edited by F. C. Neidhart, R. Curtis III & J. L. Ingraham. Washington, DC: ASM Press.

[13] Brinkman, A. B., Ettema, T. J., de Vos, W. M. & van der Oost, J. (2003).
The Lrp family of transcriptional regulators. *Molecular microbiology* **48**, 287-294.

[14] Butland, G., Peregrin-Alvarez, J. M., Li, J. & other authors (2005).
Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531-537.

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

[16] Campo, N., Tjalsma, H., Buist, G. & other authors (2004).
Subcellular sites for bacterial protein export. *Molecular microbiology* **53**, 1583-1599.

[17] Chang, V. K., Fitch, M. J., Donato, J. J., Christensen, T. W., Merchant, A. M. & Tye, B. K. (2003).
Mcm1 binds replication origins. *The Journal of biological chemistry* **278**, 6093-6100.

[18] Chang, V. K., Donato, J. J., Chan, C. S. & Tye, B. K. (2004).
Mcm1 promotes replication initiation by binding specific elements at replication origins. *Mol Cell Biol* **24**, 6514-6524.

[19] Chen, X. J. & Butow, R. A. (2005).
The organization and inheritance of the mitochondrial genome. *Nat Rev Genet* **6**, 815-825.

[20] Chen, X. J., Wang, X., Kaufman, B. A. & Butow, R. A. (2005).
Aconitase couples metabolic regulation to mitochondrial DNA maintenance. *Science* **307**, 714-717.

[21] Chen, X. J., Wang, X. & Butow, R. A. (2007).
Yeast aconitase binds and provides metabolically coupled protection to mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 13738-13743.

[22] Chen, Y. & Tye, B. K. (1995).
The yeast Mcm1 protein is regulated posttranscriptionally by the flux of glycolysis. *Mol Cell Biol* **15**, 4631-4639.

[23] Cornish-Bowden, A. & Cardenas, M. L. (1993).
Channelling can affect concentrations of metabolic intermediates at constant net flux: artefact or reality? *European Journal of Biochemistry* **213**, 87-92.

[24] Cortay, J.-C., Rieul, C., Duclos, B. & Cozzone, A. J. (1986).
Characterization of the phosphoproteins of *Escherichia coli* by electrophoretic analysis. *European Journal of Biochemistry* **159**, 227-237.

[25] Dandekar, T., Snel, B., Huynen, M. & Bork, P. (1998).
Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* **23**, 324-328.

[26] Dervyn, E., Suski, C., Daniel, R., Bruand, C., Chapuis, J., Errington, J., Janniere, L. & Ehrlich, S. D. (2001).
Two essential DNA polymerases at the bacterial replication fork. *Science* **294**, 1716-1719.

[27] Deutscher, J. & Saier, M. H., Jr. (2005).
Ser/Thr/Tyr protein phosphorylation in bacteria - for long time neglected, now well established. *Journal of molecular microbiology and biotechnology* **9**, 125-131.

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

[29] Eymann, C., Becher, D., Bernhardt, J., Gronau, K., Klutzny, A. & Hecker, M. (2007).
Dynamics of protein phosphorylation on Ser/Thr/Tyr in Bacillus subtilis. *Proteomics* **7**, 3509-3526.

[30] Firshein, W. (1989).
Role of the DNA/membrane complex in prokaryotic DNA replication. *Annual review of microbiology* **43**, 89-120.

[31] Forsburg, S. L. (2004).
Eukaryotic MCM proteins: beyond replication initiation. *Microbiol Mol Biol Rev* **68**, 109-131.

[32] Forterre, P. (2005).
The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie* **87**, 793-803.

[33] Freestone, P., Grant, S., Toth, I. & Norris, V. (1995).
Identification of phosphoproteins in *Escherichia coli*. *Molecular microbiology* **15**, 573-580.

[34] Gatenby, R. A. & Gillies, R. J. (2004).
Why do cancers have high aerobic glycolysis? *Nat Rev Cancer* **4**, 891-899.

[35] Gorringe, D. M. & Moses, V. (1978).
A multienzyme aggregate with glycolytic activity from *Escherichia coli*. *Biochemical Society transactions* **6**, 167-169.

[36] Grangeasse, C., Cozzone, A. J., Deutscher, J. & Mijakovic, I. (2007).
Tyrosine phosphorylation: an emerging regulatory device of bacterial physiology. *Trends Biochem Sci* **32**, 86-94.

[37] Grosse, F., Nasheuer, H. P., Scholtissek, S. & Schomburg, U. (1986).
Lactate dehydrogenase and glyceraldehyde-phosphate dehydrogenase are single-stranded DNA-binding proteins that affect the DNA-polymerase-alpha-primase complex. *European journal of biochemistry / FEBS* **160**, 459-467.

[38] Guarino, E., Jimenez-Sanchez, A. & Guzman, E. C. (2007).
Defective Ribonucleoside Diphosphate Reductase Impairs Replication Fork Progression in Escherichia coli. *Journal of bacteriology* **189**, 3496-3501.

[39] Guijo, M. I., Patte, J., del Mar Campos, M., Louarn, J. M. & Rebollo, J. E. (2001).
Localized remodeling of the Escherichia coli chromosome: the patchwork of segments refractory and tolerant to inversion near the replication terminus. *Genetics* **157**, 1413-1423.

[40] Guzman, E. C., Caballero, J. L. & Jimenez-Sanchez, A. (2002).
Ribonucleoside diphosphate reductase is a component of the replication hyperstructure in *Escherichia coli*. *Molecular microbiology* **43**, 487-495.

[41] Hardy, C. D. & Cozzarelli, N. R. (2005).
A genetic selection for supercoiling mutants of Escherichia coli reveals proteins implicated in chromosome structure. *Molecular microbiology* **57**, 1636-1652.

[42] Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. (1999).
From molecular to modular cell biology. *Nature* **402(6761 Suppl)**, C47-52.

[43] Havens, C. G., Ho, A., Yoshioka, N. & Dowdy, S. F. (2006).
Regulation of late G1/S phase transition and APC Cdh1 by reactive oxygen species. *Mol Cell Biol* **26**, 4701-4711.

[44] Helmstetter, C. E. (1991).
Timing of synthetic activities in the cell cycle. In *Escherichia coli and Salmonella*, pp. 1594-1605. Edited by F. C. Neidhardt, R. Curtiss III, J. L. Ingraham & other authors. Washington D.C.: American Society for Microbiology.

[45] Helmstetter, C. E. (1996).
Timing of synthetic activities in the cell cycle. In *Escherichia coli and Salmonella*, pp. 1627-1639. Edited by F. C. Neidhardt, R. Curtiss III, J. L. Ingraham & other authors. Washington D.C.: American Society for Microbiology.

[46] Henneke, G., Koundrioukoff, S. & Hubscher, U. (2003).
Multiple roles for kinases in DNA replication. *EMBO reports* **4**, 252-256.

[47] Hernandez, V. J. & Bremer, H. (1993).
Characterization of RNA and DNA synthesis in Escherichia coli strains devoid of ppGpp. *J Biol Chem* **268**, 10851-10862.

[48] Hunding, A., Kepes, F., Lancet, D., Minsky, A., Norris, V., Raine, D., Sriram, K. & Root-Bernstein, R. (2006).
Compositional complementarity and prebiotic ecology in the origin of life. *Bioessays* **28**, 399-412.

[49] Ingber, D. E. (1997).
Tensegrity: the architectural basis of cellular mechanotransduction. *Annual Review of Physiology* **59**, 575-599.

[50] Ingber, D. E. (1998).
The architecture of life. *Scientific American* **278 January**, 30-39.

[51] Janniere, L., Canceill, D., Suski, C. & other authors (2007).
Genetic evidence for a link between glycolysis and DNA replication. *PLoS ONE* **2**, e447.

[52] Jindal, H. K. & Vishwanatha, J. K. (1990).
Purification and characterization of primer recognition proteins from HeLa cells. *Biochemistry* **29**, 4767-4773.

[53] Jorgensen, K., Rasmussen, A. V., Morant, M., Nielsen, A. H., Bjarnholt, N., Zagrobelny, M., Bak, S. & Moller, B. L. (2005).
Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Current opinion in plant biology* **8**, 280-291.

[54] Kessler, P. S. & Parsons, M. (2005).
Probing the role of compartmentation of glycolysis in procyclic form Trypanosoma brucei: RNA interference studies of PEX14, hexokinase, and phosphofructokinase. *Journal of Biological Chemistry* **280**, 9030-9036.

[55] Kidane, D., Sanchez, H., Alonso, J. C. & Graumann, P. L. (2004).
Visualization of DNA double-strand break repair in live bacteria reveals dynamic recruitment of *Bacillus subtilis* RecF, RecO and RecN proteins to distinct sites on the nucleoids. *Molecular microbiology* **52**, 1627-1639.

[56] Kidane, D. & Graumann, P. L. (2005).
Intracellular protein and DNA dynamics in competent *Bacillus subtilis* cells. *Cell* **122**, 73-84.

[57] Kier, L. B., Bonchev, D. & Buck, G. A. (2005).
Modeling biochemical networks: a cellular-automata approach. *Chemistry & biodiversity* **2**, 233-243.

[58] Klevecz, R. R., Bolen, J., Forrest, G. & Murray, D. B. (2004).
A genomewide oscillation in transcription gates DNA replication and cell cycle. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 1200-1205.

[59] Kolodkin-Gal, I., Hazan, R., Gaathon, A., Carmeli, S. & Engelberg-Kulka, H. (2007).
A linear pentapeptide is a quorum-sensing factor required for mazEF-mediated cell death in Escherichia coli. *Science* **318**, 652-655.

[60] Koonin, E. V. (2006).
Temporal order of evolution of DNA replication systems inferred by comparison of cellular and viral DNA polymerases. *Biology direct* **1**, 39.

[61] Kumble, K. D., Iversen, P. L. & Vishwanatha, J. K. (1992).
The role of primer recognition proteins in DNA replication: inhibition of cellular proliferation by antisense oligodeoxyribonucleotides. *Journal of cell science* **101 (Pt 1)**, 35-41.

[62] Lark, K. G., Repko, T. & Hoffman, E. J. (1963).
The Effect of Amino Acid Deprivation on Subsequent Deoxyribonucleic Acid Replication. *Biochim Biophys Acta* **76**, 9-24.

[63] Laschet, J. J., Minier, F., Kurcewicz, I. & other authors (2004).
Glyceraldehyde-3-phosphate dehydrogenase is a GABAA receptor kinase linking glycolysis to neuronal inhibition. *J Neurosci* **24**, 7614-7622.

[64] Le Sceller, L., Ripoll, C., Demarty, M., Cabin-Flaman, A., NystrÃ¶m, T., Saier Jnr., M. & Norris, V. (2000).
Modelling bacterial hyperstructures with cellular automata. *Interjournal of Complex Systems* **Paper 366**, http://www.lri.fr/ pa/Hsim/InterJournal.pdf.

[65] Leonard, A. C. & Grimwade, J. E. (2005).
Building a bacterial orisome: emergence of new regulatory features for replication origin unwinding. *Molecular microbiology* **55**, 978-985.

[66] Letoffe, S., Delepelaire, P. & Wandersman, C. (1996).
Protein secretion in gram-negative bacteria: assembly of the three components of ABC protein mediated exporters is ordered and promoted by substrate binding. *EMBO Journal* **15**, 5804-5811.

[67] Levine, A., Vannier, F., Dehbi, M., Henckes, G. & Seror, S. J. (1991).
The stringent response blocks DNA replication outside the ori region in Bacillus subtilis and at the origin in Escherichia coli. *J Mol Biol* **219**, 605-613.

[68] Levine, A., Vannier, F., Absalon, C. & other authors (2006).
Analysis of the dynamic Bacillus subtilis Ser/Thr/Tyr phosphoproteome

implicated in a wide variety of cellular processes. *Proteomics* **6**, 2157-2173.

[69] Loeb, L. A., Loeb, K. R. & Anderson, J. P. (2003).
Multiple mutations and cancer. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 776-781.

[70] Magnusson, L. U., Farewell, A. & Nystrom, T. (2005).
ppGpp: a global regulator in Escherichia coli. *Trends Microbiol* **13**, 236-242.

[71] Marmillot, P., Hervagault, J. F. & Welch, G. R. (1992).
Patterns of spatiotemporal organization in an "ambiquitous" enzyme model. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 12103-12107.

[72] Mathews, C. K. (1988).
Microcompartmentation of DNA precursors. In *Microcompartmentation*, pp. 155-169. Edited by D. P. Jones. Boca Raton: CRC Press Inc.

[73] Mayer, F. (2003).
Cytoskeletons in prokaryotes. *Cell biology international* **27**, 429-438.

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

[75] Mendes, P., Kell, D. B. & Welch, G. R. (1995).
Metabolic channelling in organized enzyme systems: experiments and models. *Advances in Molecular and Cellular Biology* **11**, 1-19.

[76] Michelsen, O., Teixeira de Mattos, M. J., Jensen, P. R. & Hansen, F. G. (2003).
Precise determinations of C and D periods by flow cytometry in *Escherichia coli* K-12 and B/r. *Microbiology* **149**, 1001-1010.

[77] Mijakovic, I., Petranovic, D., Bottini, N., Deutscher, J. & Jensen, P. R. (2005).
Protein-tyrosine phosphorylation in Bacillus subtilis. *Journal of molecular microbiology and biotechnology* **9**, 189-197.

[78] Mijakovic, I., Petranovic, D., Macek, B., Cepo, T., Mann, M., Davies, J., Jensen, P. R. & Vujaklija, D. (2006).
Bacterial single-stranded DNA-binding proteins are phosphorylated on tyrosine. *Nucleic acids research* **34**, 1588-1596.

[79] Mitchell, C. G. (1996).
Identification of a multienzyme complex of the tricarboxylic acid cycle enzymes containing citrate synthase isoenzymes from *Pseudomonas aeruginosa*. *Biochemical Journal* **313**, 769-774.

[80] Molina, F. & Skarstad, K. (2004).
Replication fork and SeqA focus distributions in *Escherichia coli* suggest a replication hyperstructure dependent on nucleotide metabolism. *Molecular microbiology* **52**, 1597-1612.

[81] Mowbray, J. & Moses, V. (1976).
The tentative identification in *Escherichia coli* of a multi-enzyme complex with glycolytic activity. *European Journal of Biochemistry* **66**, 25-36.

[82] Mulugu, S., Potnis, A., Shamsuzzaman, Taylor, J., Alexander, K. & Bastia, D. (2001).
Mechanism of termination of DNA replication of Escherichia coli involves helicase-contrahelicase interaction. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 9569-9574.

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

[84] Noirot-Gros, M. F., Dervyn, E., Wu, L. J., Mervelet, P., Errington, J., Ehrlich, S. D. & Noirot, P. (2002).
An expanded view of bacterial DNA replication. *Proceedings of the National Academy of Science USA* **99**, 8342-8347.

[85] Norris, V., Ayala, J. A., Begg, K. & other authors (1994).
Cell cycle control : prokaryotic solutions to eukaryotic problems? *Journal of theoretical biology* **168**, 227-230.

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

[87] Norris, V., Turnock, G. & Sigee, D. (1996).
The *Escherichia coli* enzoskeleton. *Molecular microbiology* **19**, 197-204.

[88] Norris, V., Gascuel, P., Guespin-Michel, J., Ripoll, C. & Saier Jr., M. H. (1999).
Metabolite-induced metabolons: the activation of transporter-enzyme

complexes by substrate binding. *Molecular microbiology* **31**, 1592-1595.

[89] Norris, V., Fralick, J. & Danchin, A. (2000).
A SeqA hyperstructure and its interactions direct the replication and sequestration of DNA. *Molecular microbiology* **37**, 696-702.

[90] Norris, V., Blaauwen, T. D., Doi, R. H. & other authors (2007a).
Toward a Hyperstructure Taxonomy. *Annual review of microbiology* **61**, 309-329.

[91] Norris, V., den Blaauwen, T., Cabin-Flaman, A. & other authors (2007b).
Functional taxonomy of bacterial hyperstructures. *Microbiology and Molecular Biology Reviews* **71**, 230-253.

[92] Norris, V., Janniere, L. & Amar, P. (2007c).
ypothesis: Variations in the rate of DNA replication determine the phenotype of daughter cells. In *Modelling complex biological systems in the context of genomics*, pp. 71-81. Edited by P. Amar, F. KÃ©pÃ¨s, V. Norris & G. Bernot. Evry, France: EDP Sciences.

[93] Ohsumi, K., Yamazoe, M. & Hiraga, S. (2001).
Different localization of SeqA-bound nascent DNA clusters and MukF-MukE-MukB complex in *Escherichia coli* cells. *Molecular microbiology* **40**, 835-845.

[94] Okumura, M., Yamamoto, M., Sakuma, H., Kojima, T., Maruyama, T., Jamali, M., Cooper, D. R. & Yasuda, K. (2002).
Leptin and high glucose stimulate cell proliferation in MCF-7 human breast cancer cells: reciprocal involvement of PKC-alpha and PPAR expression. *Biochimica et biophysica acta* **1592**, 107-116.

[95] Ovadi, J. (1988).
Old pathway-new concept: control of glycolysis by metabolite-modulated dynamic enzyme associations. *Trends in Biochemical Sciences* **13**, 486-490.

[96] Ovadi, J. & Saks, V. (2004).
On the origin of intracellular compartmentation and organized metabolic systems. *Molecular and Cellular Biochemistry* **256-257**, 5-12.

[97] Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. (1999).
The use of gene clusters to infer functional coupling. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 2896-2901.

[98] Patel, H. V., Vyas, K. A., Li, X., Savtchenko, R. & Roseman, S. (2004). Subcellular distribution of enzyme I of the *Escherichia coli* phosphoenolpyruvate:glycose phosphotransferase system depends on growth conditions. *Proceedings of the National Academy of Science USA* **101**, 17486-17491.

[99] Petranovic, D., Michelsen, O., Zahradka, K., Silva, C., Petranovic, M., Jensen, P. R. & Mijakovic, I. (2007).
Bacillus subtilis strain deficient for the protein-tyrosine kinase PtkA exhibits impaired DNA replication. *Molecular microbiology* **63**, 1797-1805.

[100] Popanda, O., Fox, G. & Thielmann, H. W. (1998).
Modulation of DNA polymerases alpha, delta and epsilon by lactate dehydrogenase and 3-phosphoglycerate kinase. *Biochimica et biophysica acta* **1397**, 102-117.

[101] Riedinger, H. J., van Betteraey-Nikoleit, M., Hilfrich, U., Eisele, K. H. & Probst, H. (2001).
Oxygen-dependent regulation of in vivo replication of simian virus 40 DNA is modulated by glucose. *The Journal of biological chemistry* **276**, 47122-47130.

[102] Rokop, M. E., Auchtung, J. M. & Grossman, A. D. (2004).
Control of DNA replication initiation by recruitment of an essential initiation protein to the membrane of *Bacillus subtilis*. *Molecular microbiology* **52**, 1757-1767.

[103] Ronai, Z. (1993).
Glycolytic enzymes as DNA binding proteins. *The International journal of biochemistry* **25**, 1073-1076.

[104] Schreiber, G., Ron, E. Z. & Glaser, G. (1995).
ppGpp-mediated regulation of DNA replication and cell division in Escherichia coli. *Curr Microbiol* **30**, 27-32.

[105] Sharpe, M. E., Hauser, P. M., Sharpe, R. G. & Errington, J. (1998).
Bacillus subtilis cell cycle as studied by fluorescence microscopy: constancy of cell length at initiation of DNA replication and evidence for active nucleoid partitioning. *Journal of bacteriology* **180**, 547-555.

[106] Shmelev, V. K. & Serebrenikova, T. P. (1997).
A study of supramolecular organization of glycogenolytic enzymes in vertebrate muscle tissue. *Biochemistry and molecular biology international* **43**, 867-872.

[107] Srere, P. (1994).
Complexities of metabolic regulation. *Trends in Biochemical Sciences* **19**, 519-520.

[108] Srere, P. A. (1987).
Complexes of sequential metabolic enzymes. *Annual Review of Biochemistry* **56**, 89-124.

[109] Srivastava, D. K. & Bernhard, S. A. (1986).
Metabolite transfer via enzyme-enzyme complexes. *Science* **234**, 1081-1086.

[110] Stein, A. & Firshein, W. (2000).
Probable identification of a membrane-associated repressor of *Bacillus subtilis* DNA replication as the E2 subunit of the pyruvate dehydrogenase complex. *Journal of bacteriology* **182**, 2119-2124.

[111] Sullivan, D. T., MacIntyre, R., Fuda, N., Fiori, J., Barrilla, J. & Ramizel, L. (2003).
Analysis of glycolytic enzyme co-localization in Drosophila flight muscle. *The Journal of experimental biology* **206**, 2031-2038.

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

[113] Thellier, M. (2003).
From a static to a dynamic description of living systems: the framework. *Nova Acta Leopoldina NF* **88**, 11-15.

[114] Thellier, M., Legent, G., Norris, V., Baron, C. & Ripoll, C. (2004).
Introduction to the concept of "functioning-dependent structures" in living cells. *Comptes rendus biologies* **327**, 1017-1024.

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

[116] Torshin, I. (1999).
Activating oligomerization as intermediate level of signal transduction: analysis of protein-protein contacts and active sites in several glycolytic enzymes. *Front Biosci* **4**, D557-570.

[117] Trewavas, A. J. & Malho, R. (1997).
Signal Perception and Transduction: The Origin of the Phenotype. *Plant Cell* **9**, 1181-1195.

[118] Tu, B. P., Kudlicki, A., Rowicka, M. & McKnight, S. L. (2006).
Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science* **310**, 1152-1158.

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

[120] Vishwanatha, J. K., Jindal, H. K. & Davis, R. G. (1992).
The role of primer recognition proteins in DNA replication: association with nuclear matrix in HeLa cells. *Journal of cell science* **101 (Pt 1)**, 25-34.

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

[122] Wang, J. D., Sanders, G. M. & Grossman, A. D. (2007).
Nutritional control of elongation of DNA replication by (p)ppGpp. *Cell* **128**, 865-875.

[123] Wegrzyn, G. & Wegrzyn, A. (2002).
Stress responses and replication of plasmids in bacterial cells. *Microb Cell Fact* **1**, 2.

[124] Welch, G. R. & Easterby, J. S. (1994).
Metabolic channeling versus free diffusion: transition-time analysis. *Trends in Biochemical Sciences* **19**, 193-196.

[125] Winkel, B. S. J. (2004).
Metabolic channeling in plants. *Annual Review of Plant Biology* **55**, 85-107.

[126] Zheng, L., Roeder, R. G. & Luo, Y. (2003).
S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. *Cell* **114**, 255-266.

PART III POSTERS

Multidimensional Scaling: Unrevealing relational geometric patterns in transcriptome biology

Christophe Bécavin[1] and Arndt Benecke[1]

[1]Institut des Hautes Études Scientifiques & Institut de Recherche Interdisciplinaire

Abstract

Multidimensional scaling (MDS) is a well-known technique of data analysis, often used in psychology, sociology and economy [1], but forsaken in genomics. This method allows construction of a geometric graph in N-dimensions, based on a measure of "distance", such as correlation, between data. The only information we need for this technique is a distance matrix containing information about the relative distances between data points. The construction quality of the N-dimensional space is evaluated by recalculating the differences between distances in the geometric space, and the original distances in the input data. This difference is called the stress; a good MDS algorithm will tend to minimize this stress.

We have developed two MDS algorithm, both use distances between biological conditions as measured by micro-array experiments. We calculate input distances using Euclidean distance in the ˜34000-dimensional gene space. Thus with this type of data we expect to have a geometric visualization of bonds between biological conditions. We present here the two MDS techniques we have developed. The first one is a 3D multidimensional scaling method. To minimize the stress we use a physical modelling of our data, each pair of points is connected by a spring. This physical system of springs converges to an equilibrium state with minimal stress. We demonstrate usage of this technique on a problem of gene signature definition for regulatory T-cells and we directly observe and compare, in 3D, patterns between different biological conditions. The second MDS algorithm we develop uses an N-1-dimensional space, with N being the number of data points. It also uses spring modelling of the data minimizing the stress, but in order to create a unique geometric graph the construction of those objects is bijective. Applicability of this latter approach will be demonstrated using transcriptome measures from a large HIV patient study (ANRS 118) which we analyze using this technology.

We begun our research on MDS techniques recently but preliminary results already indicate its potential as an essential data analysis tool for transcriptome data. Because MDS combines techniques of ordination and classification it appears to be an optimal tool for non-supervised data mining.

[1] Borg, I., and Groenen, P. (1997) Modern Multidimensional Scaling, Springer, New York.

The deterministic dynamics of Gate-based Gene Networks

Claudiu Giuraniuc[1] and Ralf Blossey[1]

[1]Interdisciplinary Research Institute, FRE 2963 CNRS, Lille France

Abstract

The notion of a gene gate, as employed in the compositional approach to the dynamics of gene regulatory networks based on the stochastic pi-calculus [1], is considered in the context of the nonlinear dynamics description of gene regulatory networks in terms of continuous concentration variables. The kinetic reactions of the gate are translated into ordinary differential equations for both protein and gene concentrations. Basic gene networks are built from the gates and their dynamics is discussed. It is shown that for the determination of the fixed-points of the gate-based networks the gene dynamics is irrelevant. The stability of the fixed-points, however, may be affected by gene fluctuations, as we show for the limit cycle of the repressilator [2].

References

[1] Ralf Blossey, Luca Cardelli, Andrew Phillips, A Compositional Approach to the Stochastc Dynamics of Gene Networks, *Trans. Comp. Sys. Biol.* 3939, 99-122 (2006)

[2] M. Elowitz, S. Leibler, A Synthetic Oscillatory Network of Transcriptional Regulators. *Nature* 403, 335-8 (2000)

DNA melting models and genomic applications

Daniel Jost and Ralf Everaers

Université de Lyon, Laboratoire de Physique de l'ENS Lyon, CNRS UMR 5672,
46 Allée d'Italie 69364 Lyon Cedex 07, France

Key biological and nano-technological processes require the partial or complete association and dissociation of complementary DNA strands. We present a variant of the Poland-Scheraga model for DNA melting with a local, sequence-dependent salt correction of the Nearest-Neighbor parameters and end as well as interfacial energies for helical and coiled sections of the chain. We show that the model reproduces experimental data for melting temperatures over the full experimental range of strand length, strand concentration and ionic strength of the solution. In particular, we reproduce a phenomenological relation by Frank-Kamenteskii for very long chains using a parametrization based on melting curves for short oligomers. However, we also show that the parameters of the Poland-Scheraga model are still not known with sufficient precision to quantitatively predict the fine structure of melting curves. The present formulation of the Poland-Scheraga model opens the possibility to overcome this limitation by optimizing parameters with respect to an extended base of experimental data for short, medium and long chain melting. We argue that the often discarded melting data for longer oligomers exhibiting non-two-state transitions could play a particularly important role.

Moreover, recent studies have tried to link melting properties and biologically functional domains in a genome using the PS-model. Melting profiles define precisely domains in the genome which can be directly compared to functional annotations. Nevertheless, treating an entire genome ($\sim$ Mbp) is computationally demanding with the PS-model. That's why, to study very long chains, we propose to simplify the PS-model. We use a simple Ising-model, replacing the cooperative entropic loop free energy by a mean-field approximation over the loop size by introducing a typical bubble length. We show that this approximation is very efficient, with only weak losses in precision and a significant gain in model solving time. While previous studies discussed only parts of genomes, now, we are able to work on entire genomes of several species. We observe that some correlations can exist between coding (exons) and thermodynamically stable domains but discrepancies are numerous and depend notably on difference in GC-content between coding and non-coding parts of the genome. We also show that thermodynamic boundaries are regularly spaced at about $300-350$ bp in the DNA coding parts (cDNA) of genomes and that significant correlations may appear between exon-exon and thermodynamic boundaries in cDNA.

Identifying independent sub-networks of biological regulatory networks for ensuring preservation of observations issued from biological experiments

Mbarka Mabrouki[1], Jean-Paul Comet[2], Pascale Le Gall[1] and Sandrine Vial[3]

[1] Programme d'Épigénomique, F-91025 Evry, France
[2] Laboratory I3S, UMR 6070 CNRS/UNSA, F-06903 Sophia-Antipolis, France
[3] PRiSM, UMR 8144 CNRS & Univervité de Versailles, F-78035 Versailles, France

Abstract

In order to better understand biological regulatory networks (BRN), scientists usually focused on subparts of the global system by using qualitative modeling frameworks. In the multivalued discrete approach developed by R. Thomas and co-workers [1], the concentrations of constituents are abstracted by integers to denote thresholds from which they can act on other constituents in the network. An interaction graph is given by a set of interacting biological entities (genes or proteins) and by a set of interactions between these entities, represented as edges in the interaction graph labeled by a threshold and a sign. An edge between x and y labeled by (n,+) (resp. by (n,-)) is called an activation (resp. inhibition) of x over y : when the concentration level of x is greater than or equal to n (resp. less than n), then the concentration level of y tends to reach a new concentration level, which is most of times unknown. A huge set of state transition graphs defines all the possible dynamics of the system. However, only a few dynamic models meet the set of biological experiment observations. When these observations are expressed according to a temporal logic as CTL, then suitable dynamic models can be selected with model checking technics [2, 3]. Because of the classical state explosion problem, biologists start by studying small BRN that they believe to be of particular importance to represent a biological function. The interactions of this BRN with the external genes, are studied only afterwards even if these external genes potentially could influence the functioning of the studied part. Considering a BRN in which a variable can take only two abstract concentration levels (off or on), embedding it into another one can introduce for this variable other intermediate abstract concentration levels: the off-level corresponds to the level 0 and the on-level to the level n. Thus whereas in the initial BRN a path along which the variable is switched on, can take only one step, in the larger one, such paths are of length greater than n. Consequently, the classical CTL logic is not well suited for studying preservation of properties along BRN embeddings. We then use the temporal logic CTL-X which is obtained from CTL by removing the next-state connector. Since the embedding modifies the constraints dues to interactions, it can also

modify the dynamics of the subsystem. We have proven that CTL-X temporal formulas verified by a sub-BRN are preserved in any BRN embedding it provided that there is no entering interaction edge from a biological entity external to the sub-BRN towards an entity of the sub-BRN. Such sub-BRN without entering edges are called independent sub-BRN. On the contrary, without such a restrictive condition, some temporal properties (i.e. biological experimental observations) can be questioned.

In order to identify sub-BRN whose associated temporal properties are preserved by construction, we have built an algorithm searching for independent sub-BRNs. They are constituting a lattice whose minimal elements are the smallest independent sub-BRNs. Our algorithm is in two phases. In the first phase, it detects all the strongly connected components of the BRN [4]. Then, we can construct the graphs of all these components. In the second phase, we are looking for the smallest independent sub-BRNs in the lattice of all partitions of the connected components.

References

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

[2] Bernot, G., Comet, J.-P., Richard, A. and Guespin, J., (2004) Application of Formal Methods to Biological Regulatory Networks: Extending Thomas' Asynchronous Logical Approach with Temporal Logic. *Journal of Theoretical Biology*.

[3] Mateus, D. and Gallois, J.-P. and Comet, J.-P. and Le Gall, P., (2007) Symbolic modeling of Genetic Regulatory Networks. *Journal of Bioinformatics and Computational Biology*.

[4] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein, (2001) Introduction to Algorithms. *The MIT Press*.

Biological Information and Self-Organizing Dynamics

Walter Riofrio[1]

[1]Universidad Peruana Cayetano Heredia

Abstract

In this presentation, we propose to use an evolutionary approach to help understanding some of the questions related with certain issues in biological information. The special character of biological systems, tell us that its distinctive capacities could have been developed in pre-biotic times. In other words, the basic properties of life would be better comprehended if we think that they were much more likely early in time.

Recent findings such as the molecular mechanisms of the differential regulation of gene activity, and in the genomics, postgenomics and proteomics control mechanisms, suppose a richer source of information transmission inside living systems: How living systems gather information about the environment and respond adequately to it?

We will present an argument showing how it is possible to shed new light on the emergence of a dynamic self-organization which conveys the capacity to react adequately, to some extent, due to interactions with signals inciding on the system.

It is our line of reasoning that the nature of biological information does not stop at the mere idea of genetic information but is rooted in the foundations of biological phenomena. We propose that information arises in the physical world as 'information-with- biological-meaning' or **bio-meaning**. In this way, information will always be 'meaningful information for the system'.

LIST OF ATTENDEES

(March 1st, 2008)

ABCHICHE-MIMOUNI	Nadia (nadia.abchiche@ibisc.univ-evry.fr)
AIGUIER	Marc (marc.aiguier@ecp.fr)
AMAR	Patrick (pa@lri.fr)
BALLET	Pascal (pascal.ballet@univ-brest.fr)
BECAVIN	Christophe (becavin@ihes.fr)
BERNOT	Gilles (bernot@unice.fr)
BEURTON-AIMAR	Marie (beurton@labri.fr)
BIDAUT	Ghislain (ghislain.bidaut@marseille.inserm.fr)
BOEVA	Valentina (vboeva@lix.polytechnique.fr)
CHAN	Raymond (rchan@mis.mpg.de)
CHRISTOPHE	soulé (soule@ihes.fr)
DESPOIS	Didier (despois@obs.u-bordeaux1.fr)
DEVAUCHELLE	Claudine (devauchelle@genopole.cnrs.fr)
GOLIB	Dzib Jose Felipe (golib@ihes.fr)
GROSSMANN	Alexander (agrossmann@genopole.cnrs.fr)
GUESPIN	Janine (janine.guespin@wanadoo.fr)
HÉRISSON	Joan (joan.herisson@epigenomique.genopole.fr)
HUE	Isabelle (isabelle.hue@jouy.inra.fr)
JANNIERE	Laurent (laurent.janniere@jouy.inra.fr)
JOST	Daniel (daniel.jost@ens-lyon.fr)
JUNIER	Ivan (ivan.junier@epigenomique.genopole.fr)
KÉPÈS	François (Francois.Kepes@epigenomique.genopole.fr)
LARIVIERE	Damien (damien.lariviere@fourmentinguilbert.org)

LE GALL	Pascale (pascale.legall@epigenomique.genopole.fr)
LHOUSSAINE	Cedric (cedric.lhoussaine@lifl.fr)
MABROUKI	Mbarka (mabrouki@epigenomique.genopole.fr)
MARTINE	Vaxillaire (martine.vaxillaire@good.ibl.fr)
MAZAT	Jean-Pierre (jpm@u-bordeaux2.fr)
MICHEL	Olivier (michel@ibisc.univ-evry.fr)
NEBUT	Mirabelle (mirabelle.nebut@lifl.fr)
NORRIS	vic (victor.norris@univ-rouen.fr)
PARISEY	Nicolas (parisey@labri.fr)
PEREZ SILVA	Carlos Armando (cperez@ivic.ve)
PIR	Pinar (pp305@mole.bio.cam.ac.uk)
POLLET	Nicolas (Nicolas.Pollet@epigenomique.genopole.fr)
RIOFRIO	Walter (Walter.Riofrio@terra.com.pe)
RODIN	Vincent (vincent.rodin@univ-brest.fr)
ROMANEL	Alessandro (romanel@cosbi.eu)
RUCKEBUSCH	cyril (cyril.ruckebusch@univ-lille1.fr)
SOPHIE	Gallina (s.gallina@good.ibl.fr)
TAHI	Fariza (fariza.tahi@ibisc.univ-evry.fr)
THERET	Nathalie (nathalie.theret@univ-rennes1.fr)
THORREZ	Lieven (lieven.thorrez@esat.kuleuven.be)
VANDENBUNDER	Bernard (Bernard.Vandenbunder@ibl.fr)
WANG	Juhui (Juhui.Wang@jouy.inra.fr)
ZEMIRLINE	Abdallah (zemirlin@univ-brest.fr)
ZHANG	Runxuan (rzhang@pasteur.fr)