Bioinformatique : de la cellule à la puce. Alchimistes de l'informatique, entretien avec Hugues Berry et Olivier Temam, propos recueillis par Dominique Chouchan.

Hugues Berry, Olivier Temam

▶ **To cite this version:**

Hugues Berry, Olivier Temam. Bioinformatique : de la cellule à la puce. Alchimistes de l'informatique, entretien avec Hugues Berry et Olivier Temam, propos recueillis par Dominique Chouchan.. Les Cahiers de l'INRIA - La Recherche, 2008, Homo Erectus Sort D'afrique, 419 mai 2008. inria-00546805

HAL Id: inria-00546805
https://inria.hal.science/inria-00546805

Submitted on 14 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

les cahiers de l'INRIA

Rendez-vous avec les sciences et technologies de l'information et de la communication

BIOINFORMATIQUE

De la cellule à la puce

Les systèmes de régulation chez les bactéries peuvent être analysés comme des circuits intégrés. Seront-ils une source d'inspiration pour les ordinateurs de demain ?

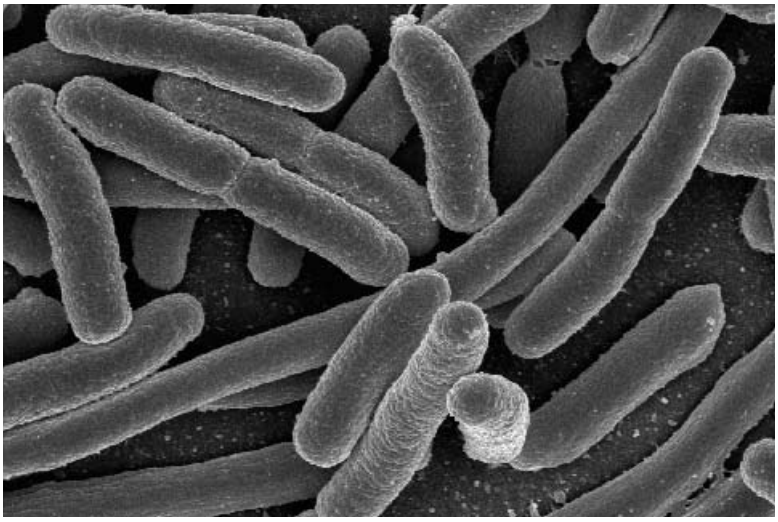
On compte environ dix fois plus de bactéries que de cellules composant les tissus dans le corps humain, des bénéfiques et des pathogènes. L'une des espèces, *Escherichia coli* (*E. coli*), identifiée depuis les années 1880, est la plus étudiée à ce jour (fig. 1). Quelques souches sont connues pour provoquer des gastro-entérites, diarrhées, infections urinaires, etc. Comment se débrouille ce micro-organisme unicellulaire pour s'adapter à l'environnement intestinal très changeant qu'il colonise ? Cette question intéresse bien sûr les biologistes, mais les informaticiens également : les systèmes qui régulent l'adaptation de la cellule bactérienne ressemblent étrangement à des circuits intégrés (puces électroniques). D'où l'idée de les appréhender avec des outils conceptuels communs.

Les bactéries disposent de capteurs moléculaires qui leur permettent de percevoir une variété de signaux provenant de l'environnement, par exemple l'absence d'un nutriment essentiel ou la densité de leur population. A ces signaux de stress, les bactéries répondent de diverses manières : par une augmentation de leur mobilité, une modification structurelle de la membrane pour rendre la cellule plus résistante, la synthèse d'acides animés qui ne se trouvent plus dans le milieu. Comment ces signaux sont-ils intégrés par la cellule et transformés en réponses appropriées ?

Les systèmes de régulation responsables de cette adaptation peuvent être représentés par des réseaux interconnectés et très complexes de réactions biochimiques : ils sont constitués de gènes, enzymes et protéines de stress (s'exprimant en cas de stress), mais aussi de métabolites et d'autres petites molécules (fig. 2). La biologie moléculaire a permis d'identifier une grande partie des composants moléculaires de ces réseaux et les réactions auxquelles ils participent[1]. Mais la compréhension du fonctionnement de l'ensemble exige de nouveaux outils.

A cet égard, l'informatique est une source d'inspiration. En effet, les puces électroniques sont elles aussi des réseaux composés d'éléments hétérogènes densément interconnectés (fig. 2). Pour les analyser, les informaticiens disposent d'outils de modélisation. Ces derniers sont surtout utilisés pour s'assurer que les propriétés de bon fonctionnement souhaitées comme la sûreté (le système ne fait pas de faute) et la vivacité (le calcul progresse) sont satisfaites par le modèle du circuit, avant de passer à la phase coûteuse de réalisation.

Parmi ces outils, ceux fondés sur les méthodes dites de vérification formelle se sont imposés, en particulier le *model checking*[2]. Ces méthodes, proposées dans les années 1980, consistent à vérifier

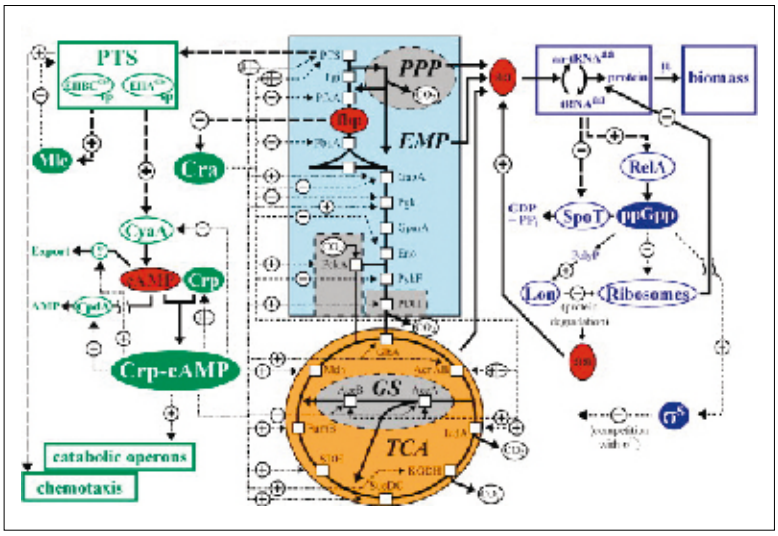


© ROCKY MOUNTAIN LABORATORIES, NIAID, NIH

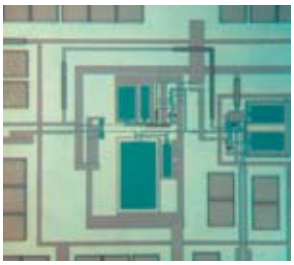
Fig. 1 : Les bactéries E. coli croissent et se multiplient activement tant qu'elles ne manquent pas des nutriments nécessaires.

[1] F. Rechenmann et H. de Jong, Le vivant en équations, *La Recherche, 383,* 2005. Voir aussi : http://interstices.info/jcms/c_6474/bioinformatique.

[2] E.M. Clarke, O. Grumberg, D.A. Peled, *Model Checking,* MIT Press, 1999.



© T. HARDIMAN ET AL., J. BIOTECHNOL., 132, 2007.



© UQAM

Fig. 2 : ci-dessus, le réseau de régulation contrôlant le métabolisme central chez *E. coli* dans une situation de stress lié à un manque de carbone. Ci-contre, une puce électronique dont la taille originale est de 1,5 mm sur 1,5 mm.

une propriété par l'analyse exhaustive de tous les comportements possibles engendrés par le modèle. Des algorithmes efficaces ont été développés depuis qui n'imposent plus cette exhaustivité (un obstacle évident pour l'analyse d'un grand système), mais qui manipulent des descriptions implicites très compactes de la dynamique du système. Le *model checking* s'est aujourd'hui largement répandu dans l'industrie, de la microélectronique à l'avionique[3].

Or si ces méthodes de vérification formelle marchent si bien pour l'analyse des systèmes critiques d'un Airbus, pourquoi ne pas les appliquer au réseau complexe d'*E. coli* ? C'est cette idée qui a donné naissance au projet européen EC-MOAN (voir l'encadré). L'objectif est de mieux comprendre comment la bactérie adapte son fonctionnement à la disponibilité dans son milieu des sources de carbone et d'ammonium, deux nutriments essentiels pour sa croissance.

Nous développons donc des modèles mathématiques classiques, à base d'équations différentielles, pour décrire le fonctionnement des réseaux impliqués dans le contrôle de l'assimilation du carbone et de l'ammonium. Nous transformons ces modèles en automates, un type de modèles utilisés pour décrire les circuits intégrés et vérifiables par *model checking*. Nous pouvons alors vérifier que ces modèles informatiques (automates) représentent bien les comportements de *E. coli* observés en laboratoire par cette technique de vérification formelle.

Au terme d'une étude préliminaire, nous avons pu identifier des propriétés intéressantes et révélatrices du fonctionnement du système, dont la validité peut être testée en laboratoire. En particulier, l'analyse de sous-réseaux de taille modeste a permis de mieux caractériser la réponse d'*E. coli* à un apport soudain de glucose (source de carbone) [4]. Cette étude et des travaux similaires ont toutefois mis en évidence l'importance d'une adaptation des méthodes de vérification formelle, compte tenu de la taille importante et de la complexité des réseaux de régulation de la bactérie [5][6]. Parmi les solutions explorées, la parallélisation des algorithmes de *model checking* : cela permettrait de vérifier les comportements de la bactérie sur des grilles de calcul.

Au-delà d'une meilleure compréhension de l'adaptation des bactéries à leur environnement, le projet EC-MOAN nous invite à nous poser d'autres questions, plus ambitieuses encore. Peut-on redessiner les réseaux de régulation pour changer le fonctionnement des bactéries, par exemple pour leur faire produire des médicaments ? Peut-on à l'inverse utiliser notre compréhension de ces réseaux, dont la robustesse et la décentralisation sont les points forts, pour repenser l'architecture de nos ordinateurs ? Cette dernière question montre que la perspective qui s'ouvrira, au terme de ce projet, ne sera pas seulement d'appliquer les méthodes informatiques à la biologie, mais aussi de s'inspirer du fonctionnement des systèmes biologiques pour contribuer à l'informatique. **H. de J., R. M., D. R., J. G.**

Hidde de Jong et Delphine Ropers sont chercheurs en bioinformatique à l'INRIA Grenoble-Rhône-Alpes (équipe IBIS).

Radu Mateescu est chercheur en vérification formelle dans le même institut (équipe VASY).

Johannes Geiselmann, professeur de microbiologie à l'université Joseph-Fourier (UJF) à Grenoble, dirige l'équipe Contrôle de l'expression génique au Laboratoire Adaptation et Pathogénie des Microorganismes (CNRS-UJF).

EC-MOAN EN BREF

Le projet EC-MOAN (6e PCRD), ou *Scalable modeling and analysis techniques to study emergent cell behavior : Understanding the E. coli stress response*, a démarré en 2007 et durera jusqu'en 2010. Il regroupe des biologistes de l'université Joseph-Fourier (Grenoble) et de l'université Libre d'Amsterdam, ainsi que des mathématiciens et des informaticiens du Centre de recherche en informatique (CWI) d'Amsterdam, de l'INRIA Grenoble-Rhône-Alpes, de l'université de Twente (Pays-Bas) et de l'université Masaryk (République Tchèque).

Pour en savoir plus : http://www.ec-moan.org/.

[3] H. Garavel *et al.*, CADP 2006 : A toolbox for the construction and analysis of distributed processes. *CAV'2007*, LNCS 4590, Springer-Verlag, 2007.

[4] G. Batt *et al.*, Validation of qualitative models of genetic regulatory networks by model checking, *Bioinformatics*, 21 (Suppl 1), 2005.

[5] G. Bernot *et al.*, Application of formal methods to biological regulatory networks, *Journal of Theoretical Biology*, 229, 2004

[6] N. Chabrier-Rivier *et al.*, Modeling and querying biomolecular interaction networks, *Theoretical Computer Science, 325 (1)*, 2004.

D.R.

Hugues Berry (à droite) est chercheur à l'Inria-Saclay-Ile-de-France (équipe Alchemy). Biochimiste, il travaille sur la modélisation des cellules et son application à l'architecture des ordinateurs.
Olivier Temam (à gauche) est chercheur à l'Inria-Saclay-Ile-de-France (responsable de l'équipe Alchemy). Informaticien, il est spécialiste de l'architecture des ordinateurs.

Entretien avec Hugues Berry et Olivier Temam

Alchimistes de l'informatique

Du réel à l'artificiel, et réciproquement : l'ordinateur du futur sera-t-il le fruit d'un astucieux mélange de l'un et de l'autre ?

Pourquoi imaginer de nouvelles architectures d'ordinateurs ?
Olivier Temam : L'une des questions, pour le futur, est celle de savoir comment augmenter la puissance des machines. La loi de Moore* n'a pas été démentie à ce jour. Une puce de taille inférieure au cm² contient actuellement plus d'un milliard de transistors et la densité des transistors devrait continuer à augmenter. Mais il devient difficile d'exploiter cette miniaturisation, autrement dit d'augmenter la vitesse de fonctionnement des processeurs conventionnels (à un seul nœud de calcul), du fait des effets thermiques : plus la dimension des transistors diminue, plus leur vitesse de commutation peut être élevée et plus il y a dissipation d'énergie par unité de surface. C'est pourquoi des firmes comme Intel conçoivent désormais des processeurs multi-cœurs, c'est-à-dire équipés de plusieurs nœuds (unités de calcul) calculant en parallèle. Reste qu'à multiplier le nombre d'unités de calcul par processeur, nous risquons de rencontrer un nouvel obstacle, lié cette fois aux difficultés de programmation : il serait souvent difficile d'extraire un grand nombre de tâches qui puissent s'exécuter en parallèle. D'où notre volonté d'explorer d'autres paradigmes.

Vous cherchez donc des sources d'inspiration dans le vivant…
Hugues Berry : L'observation du monde vivant au niveau cellulaire révèle en effet l'existence de modes d'organisation très similaires à ce que l'on aimerait obtenir des futurs ordinateurs. Exemple : nos capacités visuelles, qui nous permettent, entre autres, de reconnaître sans effort les objets complexes d'une image. Dans le cerveau, le nombre de cellules (neurones) pour effectuer une telle tâche est énorme. Chacune est relativement lente et loin d'être infaillible (risque d'erreur ou de panne). Enfin, le cerveau ne s'appuie sur aucune centralisation des opérations.
O. T. : On commence en outre à comprendre en détail comment sont atteintes de telles performances, notamment grâce aux travaux de l'Américain Tomaso Poggio (MIT). Celui-ci a construit des fonctions visuelles à partir du neurone, un peu à la manière dont on construit un processeur à partir du transistor. On imagine donc des systèmes composés d'un nombre énorme de composants très petits (des nanotubes par exemple), éventuellement très lents et susceptibles de défauts ou d'erreurs. Pour l'instant, Poggio assemble, « à la main » en quelque sorte, des modèles de neurones. Pour notre part, nous travaillons sur l'émergence de telles fonctions en partant d'un grand nombre de neurones, non structurés et obéissant seulement à quelques règles élémentaires et locales.

Un raisonnement similaire vaut-il pour d'autres systèmes cellulaires ?
H. B. : Absolument. D'une part on peut s'inspirer des communications chimiques entre cellules. D'autre part, à l'intérieur même d'une cellule, donc à plus petite échelle, il existe des réseaux de régulation moléculaires aux propriétés intéressantes également (voir « De la cellule à la puce » p. I).
O. T. : Nous menons en fait deux approches en parallèle. Nous travaillons sur de nouvelles architectures en nous inspirant des réseaux de cellules neuronales. Hugues, pour sa part, s'intéresse également à la possibilité d'utiliser les cellules comme hardware, en « programmant » les réseaux métaboliques et génétiques au sein même de ces cellules pour faire des calculs.

Propos recueillis par Dominique Chouchan

*La loi de Moore doit son nom à l'Américain Gordon Moore (cofondateur d'Intel) qui énonça que tous les 18 à 24 mois, la dimension des transistors serait divisée par deux, donc que leurs densité et vitesse de commutation doubleraient.

D.R.

De gauche à droite et de haut en bas : Thomas Clozel, Thomas Landrain, David Bikard, David Guegan, Aurélien Rizk, Gilles Vieira, Nicolas Chiaruttini, Fanny Caffin, Eimad Shotar

et un étudiant en médecine. Il a très vite attiré d'autres compétences : des étudiants en informatique, un ingénieur de l'Ecole centrale, un physicien de l'Ecole normale supérieure (ENS)... C'est dans ce contexte que nous nous sommes mis d'accord sur un projet pour concourir : la réalisation d'une « bactérie multicellulaire » synthétique. Notre idée, modifier génétiquement une bactérie (unicellulaire) de telle sorte qu'elle se différencie en deux types cellulaires interdépendants, un premier pas vers la fabrication d'un organisme multicellulaire.

Il s'agissait d'insérer, dans le génome de la bactérie unicellulaire (*Escherichia coli*), un fragment d'ADN permettant d'induire une différentiation de la cellule en deux types différents : l'un ayant besoin pour survivre et se diviser à nouveau d'un métabolite fourni par l'autre (qui ne se divise plus). Le travail a consisté à modéliser le système sur ordinateur et à faire des expériences en laboratoire. Pour la modélisation, nous avons utilisé des outils originaux comme Biocham, un environnement de programmation de systèmes biochimiques, et MGS, un langage de programmation développé à l'Université d'Evry pour la simulation de systèmes dynamiques.

« Un projet de biologie synthétique en trois mois d'été »

Réaliser une bactérie multicellulaire synthétique, tel est le projet qui a valu le prix de la recherche fondamentale du concours international iGEM* à une équipe d'étudiants en master. Une recherche originale en vue de la conception d'organismes vivants capables de répondre à des spécifications données.

« Le concours iGEM s'adresse aux étudiants de niveau inférieur à bac + 5. La règle du jeu : réaliser un projet de biologie synthétique pendant les trois mois d'été. Pour notre part, nous nous sommes rencontrés au Centre de recherche interdisciplinaire (CRI, faculté de médecine de Paris Descartes). Ce centre associe des scientifiques de la région parisienne et des étudiants issus notamment du master Approches interdisciplinaires du vivant (AIV) et de l'Ecole doctorale Frontières du vivant (FdV). Ces derniers sont incités à créer des clubs scientifiques au sein desquels ils échangent connaissances et expériences, mais aussi invitent des chercheurs d'horizons divers.

Le club de biologie synthétique a ainsi été créé par deux d'entre nous, un ingénieur agronome

Ce type de démarche vise à modifier des organismes, unicellulaires dans un premier temps, afin qu'ils produisent des substances utiles (thérapeutiques,...) : c'est l'un des buts de la biologie synthétique, un autre étant par exemple de faire adopter par la cellule des comportements particuliers. Au plan fondamental, l'idée est d'élaborer des outils pour étudier les mécanismes de la multicellularité (conditions nécessaires à son apparition...), mais aussi de comprendre suffisamment un organisme pour pouvoir éventuellement le reconstruire de toutes pièces.

Nous avons été soutenus par des scientifiques de différentes structures (INSERM, CNRS, INRIA, CEA, université d'Evry...), en particulier par Ariel Lindner (Paris Descartes, INSERM) et Samuel Bottani (physicien à Paris VII). Pour les financements, nous avons nous-mêmes démarché auprès de sponsors (notamment la fondation Bettencourt Schueller). Cette expérience nous a permis de nouer des relations fortes que nous espérons entretenir au plan professionnel, à commencer par les dernières vérifications encore à effectuer pour finaliser notre projet. » **Les lauréats du prix iGEM**

* Le concours iGEM (International Genetically Engineered Machine Competition) est organisé chaque année par le MIT depuis 2004. Pour en savoir plus sur le travail des lauréats : http : //parts.mit.edu/igem07/index.php/Paris

Réalisation rédactionnelle : Dominique Chouchan - Réalisation graphique : E-Graphics