

Introduction à la biologie synthétique et au développement d'outils de conception dédiés

Morgan MADEC¹, Yves GENDRAULT¹, Christophe LALLEMENT¹, Jacques HAIECH²

¹Université de Strasbourg : Telecom Physique Strasbourg, Boulevard Sébastien Brandt, 67412 ILLKIRCH. Laboratoire : ICube UMR 7357 (CNRS / Unistra), Equipe Système et Microsystèmes Hétérogènes, 300 Bd Sébastien Brandt, 67412 ILLKIRCH CEDEX 02

²Université de Strasbourg : Ecole Supérieure de Biotechnologie de Strasbourg, Boulevard Sébastien Brandt, 67412 ILLKIRCH CEDEX 02 / Laboratoire : Laboratoire d'Innovation Thérapeutique (LIT – UMR 7200), Faculté de Pharmacie, 74 rte du Rhin, 67401 ILLKIRCH

Résumé — L'article décrit les travaux de recherche d'une équipe d'enseignants-chercheurs Strasbourgeois (laboratoire ICube, équipe SMH, <http://icube-smh.unistra.fr/index.php/Accueil>) dans le domaine du développement d'outils d'aide à la conception pour la biologie synthétique. L'idée-force de notre approche est d'exploiter les similitudes qu'il existe entre la microélectronique et le génie génétique, tant au niveau des fonctions que des méthodes utilisées, afin d'adapter les outils, méthodes et langages puissants et éprouvés de la microélectronique à la biologie synthétique. Le développement de cette nouvelle thématique s'est fait en synergie avec l'école d'ingénieur Telecom Physique Strasbourg (<http://www.telecom-physique.fr/>). Les étudiants, de formation généraliste, suivent des enseignements d'ouverture dans le domaine des biotechnologies et réinvestissent leurs compétences dans des projets autour du développement de ces outils de conception.

1. INTRODUCTION

La biologie synthétique est une nouvelle branche des biotechnologies à l'interface avec les sciences pour l'ingénieur. Elle a pour objectif le développement de nouveaux micro-organismes génétiquement programmés pour réaliser une fonction donnée (ex : synthétiser de l'insuline en cas de taux de glucose trop élevé, détecter la formation de tumeurs par hypoxie, fluorescer en cas de présence de polluant en milieu aqueux, détecter une signature génétique spécifique ...). La médecine et l'environnement sont les deux principaux champs d'application de cette nouvelle technoscience [1][2][3][4].

La biologie synthétique a été pointée par l'Europe et la France comme une technologie à fort potentiel applicatif, comme en témoignent deux rapports récents, l'un publié par l'EMBO (European Molecular Biology Organization, www.embo.org) [5] et l'autre par le Ministère de l'Enseignement et de la Recherche [6]. Les spécialistes du domaine prédisent que la biologie synthétique connaîtra dans un avenir proche le même essor que la microélectronique à la fin du siècle dernier à condition que les innovations technologiques s'accompagnent de rigueur et de structuration dans la démarche de conception des biosystèmes [7][8].

Dans ce contexte, l'expérience de la microélectronique peut s'avérer être un atout majeur pour atteindre cet objectif. C'est dans ce cadre que l'équipe Système et Microsystèmes Hétérogènes du laboratoire ICube (Strasbourg, France, <http://icube-smh.unistra.fr/index.php/Accueil>) développe depuis quatre ans, en partenariat avec le Laboratoire d'Innovations Thérapeutiques (LIT, Strasbourg, France, <http://medchem.u-strasbg.fr/>), une thématique de recherche dans le but traiter cette question.

Notre objectif est de réinvestir nos compétences dans le domaine de la modélisation des systèmes hétérogènes ainsi que dans le domaine de la technologie de la conception en microélectronique pour élaborer de nouveaux outils de conception (ou réadapter des outils existants) pour la biologie synthétique. Des études préliminaires ont montré le potentiel de cette nouvelle approche et nous ont donc encouragés à démarrer cette élaboration d'outils [9]. Outre le volet recherche, nous avons également mis en place des modules d'ouverture et de formation à ces problématiques dans le cadre de la formation d'ingénieur généraliste et de la nouvelle formation TIC-Santé de Telecom Physique Strasbourg (<http://telecom-physique.fr>).

L'objectif de ce document est d'introduire la biologie synthétique à la communauté de la microélectronique et de démontrer, par l'intermédiaire de quelques exemples de réalisation, la synergie qui peut exister entre ces deux disciplines. Le document démarre par une présentation de la biologie synthétique et de ses principes fondamentaux. Ensuite, nous présenterons les similitudes qui peuvent exister d'un point de vue des méthodologies de conception entre cette science et la microélectronique. Enfin, dans la dernière partie du texte, nous détaillerons la formation mise en place à Telecom Physique Strasbourg autour de ces problématiques et donnerons quelques exemples de réalisation d'étudiants.

2. LA BIOLOGIE SYNTHETIQUE

Pour bien appréhender le sujet, commençons par exposer les principes de base de la biologie synthétique.

2.1. Les concepts de la biologie synthétique

En paraphrasant le terme « biologie synthétique », on arrive rapidement à l'idée de « vie artificielle ». C'est en effet de cela qu'il s'agit, même si les concepts de la biologie synthétique restent encore très éloignés de ceux des scénaristes de science-fiction. Aujourd'hui encore, la biologie traditionnelle reste avant tout une science descriptive et empirique visant à expliquer le vivant et comprendre les mécanismes de la nature. Pourtant, dès le début du XXe siècle, certains biologistes mettent en avant l'intérêt de réinvestir les connaissances acquises afin de concevoir et de prédire la vie. Néanmoins, les technologies de l'époque n'étaient pas assez développées pour concrétiser ces idées.

Ce principe de la biologie synthétique est revisité à la fin du XXe siècle par des scientifiques issus des sciences pour l'ingénieur. Au-delà du concept de « vie artificielle », ils proposent également de nouvelles méthodologies pour remplacer celles qui existent alors et qu'ils jugent jusque-là insuffisantes dans le domaine. En 2003, Tom Knight introduit le concept de BioBriques qui se veulent être les composants de base de la Biologie Synthétique [10]. Knight imagine que la conception de nouveaux organismes passera par un assemblage rationnel de ces BioBriques. En 2008, Drew Endy va plus loin en décrivant le concept de bibliothèque de conception. Il associe un modèle à la BioBrique, des règles d'assemblage, des règles de compatibilités avec le microorganisme dans lequel le gène va être implanté, etc. [8].

Le troisième concept découle naturellement des deux premiers, puisqu'il s'agit de créer de nouvelles fonctions biologiques avec une approche rationnelle et méthodique. Le recours à des outils informatiques de modélisation, de simulation, et d'aide à la conception s'est donc très rapidement imposé. A l'image de ce qu'il existe dans d'autres domaines, l'idée de pouvoir concevoir de la vie en s'appuyant sur des logiciels capables d'automatiser certaines étapes du flot de conception performant fait rêver. Actuellement, les logiciels sont développés au cas par cas pour les besoins spécifiques de chaque application. Bien que le besoin d'un logiciel générique soit clairement exprimé, la communauté reste globalement sceptique à l'idée de *Genetic Design Automation* (GDA) [11].

Aujourd'hui, bien que ces concepts soient en avance par rapport à l'état de l'art de ce qu'il est possible de réaliser avec les technologies actuelles, ils apparaissent néanmoins comme primordiaux pour le développement à moyen terme de cette science.

2.2. Ce que l'on sait faire ...

Depuis le début du XXI^e siècle, de plus en plus de travaux sont publiés dans le domaine de la biologie synthétique. Si les premiers travaux étaient plus des preuves de concepts, montrant la possibilité de reprogrammer une bactérie pour lui donner des fonctions nouvelles (ex : synthétiser des molécules fluorescentes à intervalle régulier) [2], les premières applications concrètes apparaissent à la fin des années 2010. C'est le cas par exemple d'un traitement innovant contre la malaria, développé par l'équipe de Keasling [3]. Le concept de ce médicament est de programmer la chaîne métabolique conduisant à la synthèse de la protéine active au sein d'une levure au lieu de la faire « en usine ». Dans le même ordre d'idée, nous pouvons également citer les travaux récents de Fussengger sur l'insémination artificielle des bovins [4]. Les spermatozoïdes du bœuf sont encapsulés avec des bactéries dans une enveloppe de cellulose. Les bactéries sont reprogrammées pour synthétiser de la cellulase (détruisant l'enveloppe) lorsqu'elles sont en contact avec une hormone présente chez la vache au moment de l'ovulation. Ainsi, les spermatozoïdes sont libérés exactement au bon moment, ce qui permet de tripler le taux de réussite de l'insémination artificielle. Ces exemples constituent les premiers micro-organismes reprogrammés fonctionnels. Ils ne sont cependant constitués que de quelques gènes artificiels.

2.3. ... et ce que l'on voudrait savoir faire.

De par leur relative simplicité (en termes de fonctionnalité), les systèmes présentés ci-dessus ont été conçus « à la main » et sont pour le moment loin de justifier le développement et l'utilisation d'outils d'aide à la conception générique. Néanmoins, les prévisions d'évolution de ce domaine sont très optimistes : les technologies nécessaires à la réalisation et à l'implantation des gènes artificiels sont mieux maîtrisées, les coûts de productions sont réduits et la reproductibilité des systèmes réalisés s'améliore également. Pour optimiser encore la mise au point de nouveaux biosystèmes synthétiques, l'idée serait de pouvoir découpler les métiers entre la conception et la réalisation. Le concept de BioBrick serait alors prépondérant, puisqu'il serait l'interface entre ces deux métiers : le concepteur assemblerait virtuellement des modèles prédictifs de BioBricks jusqu'à obtenir la fonction et les performances souhaitées puis confierait le « plan de construction » aux industriels capable de réaliser et d'intégrer ces mêmes BioBriques. Ceux-ci deviendront en quelques sortes des « bio-fondeurs ».

3. LIENS ENTRE LA BIOLOGIE SYNTHETIQUE ET LA MICROELECTRONIQUE

3.1. Démarche de conception – Approche digitale.

Lorsque nous relisons le paragraphe précédent avec un peu de recul, nous y retrouvons les concepts clefs de la microélectronique actuelle : les BioBriques qui sont les composants élémentaires, et les notions de modélisation, de simulation et de conception virtuelle, jusqu'à la notion de « design kit » qui est reprise par Endy dans son idée de bibliothèque de conception [8]. Il semble donc pertinent de vouloir adapter les outils et les méthodes utilisés en microélectronique pour l'automatisation de la conception en biologie synthétique.

3.2. Modélisation – Approche digitale.

En allant plus loin dans l'étude du fonctionnement des réseaux génétiques, d'autres analogies avec la microélectronique apparaissent. Lorsque l'on considère un gène unique, de la même manière que le courant dans un transistor MOSFET peut être modulé par sa tension de grille. L'expression génétique X peut être modulée par la quantité de certaines espèces chimiques dans la cellule. En pratique, pour chaque gène, il existe deux type d'espèces régulatrices, les activateurs A ou les répresseurs R . Le fonctionnement d'un gène peut donc être modélisé de manière abstraite par l'équation logique : $X = A \text{ AND } (\text{NOT } R)$.

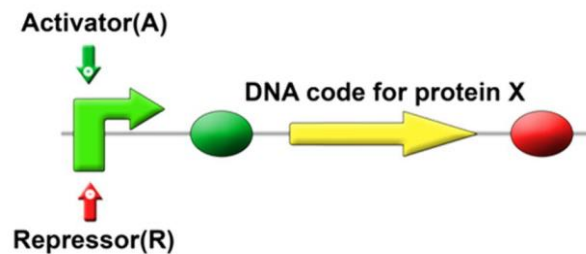


Fig. 1 – Mécanisme élémentaire de la régulation génétique : la quantité de X synthétisé par le gène dépend directement de la quantité d'activateur et de répresseur au sein de la cellule

A partir de cette fonction élémentaire, nous pouvons alors créer d'autres fonctions logiques, comme par exemple une porte AND imaginée par Anderson en 2007 [12]. La présence de salicylate dans la cellule induit la synthèse d'une protéine artificielle ($supD$). De la même manière, la présence d'arabinose induit la synthèse d'une autre protéine artificielle ($T7ptag$). Lorsque ces deux protéines artificielles sont mises en contact, elles réagissent pour former le complexe T7, lui-même activateur d'un troisième gène synthétisant une protéine fluorescente verte (GFP, Fig. 2).

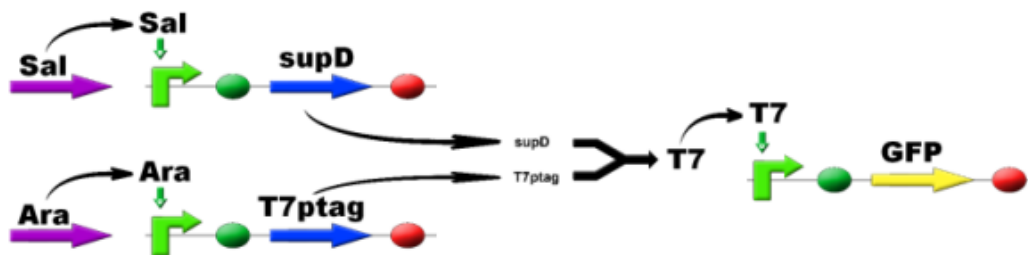


Fig. 2 – Porte AND biologique réalisée par Anderson.

Compte-tenu des similitudes dans l'approche de conception et des liens forts qui existent en terme de comportement entre les deux domaines, il doit être possible de réadapter une partie de flot de conception (et des outils associés) de la microélectronique à la biologie synthétique.

3.3. Modélisation – Approche « analogique ».

La vision d'un réseau génétique décrite précédemment n'est qu'une abstraction du fonctionnement réel des systèmes. C'est déjà vrai en électronique, ce l'est encore plus en biologie. A l'heure actuelle, il n'existe pas encore les bases théoriques nécessaires permettant d'établir des modèles biologiques à bas niveau (comme cela peut être le cas par exemple en microélectronique où la physique du semi-conducteur permet d'établir le modèle des transistors utilisés dans un circuit). Pour la conception de biosystèmes, on s'en remet donc le plus souvent à des modèles empiriques. De manière générale, les modèles d'un réseau génétique s'écrit toujours sous la forme d'un système d'équations différentielles couplées de la forme suivante :

$$\frac{dX_k}{dt} = S(X_1, \dots, X_N) - C(X_1, \dots, X_N) - d_k X_k$$

où X_k est la concentration de l'espèce k , S et C sont respectivement deux fonctions de synthèse et de consommation dépendant des autres espèces présentes dans la cellule et d_k est un coefficient d'auto-dégradation.

Dans le cas d'un phénomène de régulation génétique (Fig. 1), la synthèse de la protéine X s'effectue en deux étapes : la transcription de l'ADN en ARN messenger (on note mX l'ARN messenger de la protéine X) et la traduction de l'ARN message en protéine. Ces deux étapes sont modélisées par deux équations différentielles, l'équation de Hill (tenant compte respectivement les effets d'activation de la protéine A et des effets de répression de la protéine R) pour la traduction et une équation du premier ordre pour la transcription.

$$\frac{dmX}{dt} = \frac{K_{tl}}{\left[1 + \left(\frac{KA}{A}\right)^{n_A}\right] \left[1 + \left(\frac{R}{KR}\right)^{n_R}\right]} - d_{mX} mX$$

$$\frac{dX}{dt} = K_{tr} mX - d_X X$$

où K_{tl} et K_{tr} sont respectivement les coefficients de traduction et de transcription, K_A et K_R les forces des activateurs et des répresseurs et n_A et n_B les coefficients de Hill. La résolution de ces équations aux limites et en régime permanent conduit à l'approche digitale expliquée en partie 3.2.

Dans le cas d'une réaction chimique simple (ex : liaison entre un ligand et un polymère), on utilise un modèle cinétique du premier ordre de type. Le terme de synthèse pour les produits P_k , ainsi que le terme de consommation des réactifs R_k , sont donc donnés par

$$S_{Pk}(P_1, \dots, P_n, R_1, \dots, R_n) = C_{Rk}(P_1, \dots, P_n, R_1, \dots, R_n) = K_{on} \prod R_k - K_{off} \prod P_k$$

où K_{ON} est le coefficient d'activation de la réaction et K_{OFF} est le coefficient de la réaction inverse.

Dans tous les cas, il s'agit d'équations différentielles ordinaires couplées conservatives (les espèces chimiques synthétisées sont soit dégradées, soit stockées dans la cellule). Le jeu d'équation est donc parfaitement adapté pour être décrit et résolu par l'intermédiaire de simulateurs électroniques, comme par exemple SPICE ou VHDL-AMS.

3.4. Vers un EDA biologique ...

Dans ce contexte, notre objectif est de réaliser un outil dédié à la biologie synthétique qui se base sur nos connaissances en microélectronique. Le GDA auquel nous souhaitons aboutir à terme est décrit sur la Fig. 3. Il reprend l'approche de conception top-down avec au départ des spécifications biologiques, une synthèse automatisée et une recherche de BioBriques dans des bases de données pour aboutir à des schémas du même type que celui représenté sur la Fig. 1, une série de simulations à plusieurs niveaux pour valider les propositions du synthétiseur, des étapes d'optimisation et une formalisation de l'assemblage de BioBriques à synthétiser en vue de la réalisation du biosystème [13].

Cette première analyse a permis d'identifier un certain nombre de verrous à lever dans l'optique du développement d'un tel outil :

- Un synthétiseur logique modifié afin de prendre en compte les contraintes de la biologie.
- Un compilateur de BioBriques, capable de sélectionner dans une bibliothèque les BioBriques optimales pour remplir les fonctions élémentaires identifiées à la synthèse

- Un outil permettant d'établir des modèles des BioBriques compatibles avec les langages de conception de la microélectronique (VHDL/VHDL-AMS, Verilog, SystemC, etc.)
- Des modèles et un outil de simulation bas niveau, équivalent de Silvaco en électronique, permettant d'étudier finement les mécanismes biologiques mis en jeu.
- Une normalisation des BioBriques intégrant à la fois leur fonction, leur modèle ainsi que les paramètres et toutes les règles de compatibilité (équivalent à des règles de dessins en microélectronique) permettant d'assurer que la BioBrique fonctionne de manière nominale.

Ces points constituent les défis que devront relever ensemble les chercheurs et les ingénieurs.

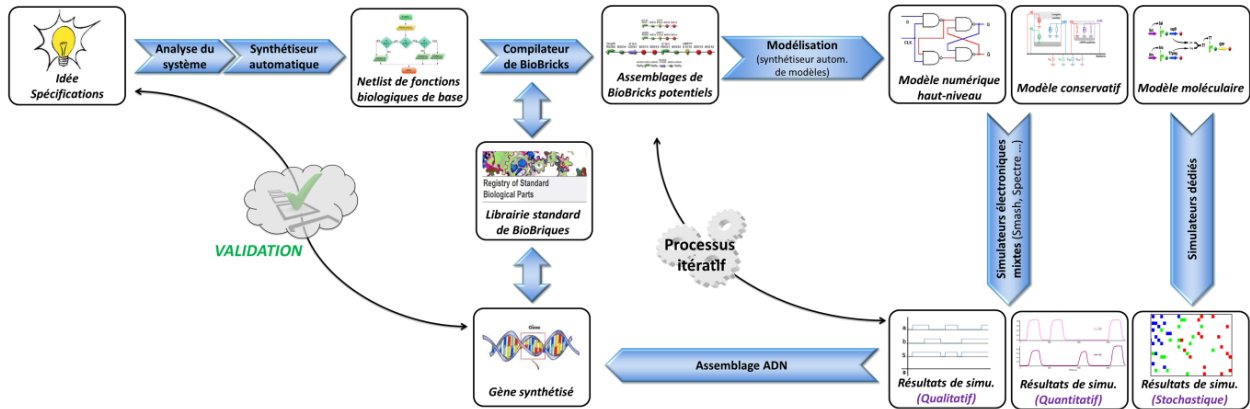


Fig. 3 – Flot de conception proposé pour la biologie synthétique par analogie à la microélectronique.

4. FORMATIONS ASSOCIEES

Le LEEM (Les Entreprises du Médicament) et l'ARIIS (Alliance Recherche et Innovation des Industries de la Santé) identifient, dans un rapport publié en 2010, la biologie synthétique comme une technologie clé dans le développement de l'industrie pharmaceutique et des industries de santé. Elle prévoit notamment l'émergence de nouveaux métiers en lien avec le développement de nouvelles approches de R&D telles que celle décrite dans le paragraphe précédent [14]. Dans ce même rapport, intitulé « Plan Compétence BioTech 2015 », le LEEM décrit la nécessité de mettre en place des formations à ces nouveaux métiers, en se basant sur des socles de formation aussi bien du domaine des biotechnologies que de celui des sciences pour l'ingénieur. A ce titre, Telecom Physique Strasbourg (anciennement Ecole Nationale Supérieure de Physique de Strasbourg) est la première école d'ingénieur à proposer à ses étudiants des UE d'ouverture vers ce nouveau domaine. L'école offre aujourd'hui deux formations distinctes : une formation d'ingénieur généraliste et une nouvelle formation, inédite en France, spécialisée dans le domaine des STIC appliqués au domaine de la santé. La biologie synthétique est présente dans les deux formations.

En formation généraliste, après trois semestres de tronc commun, les étudiants sont amenés à choisir l'un des 7 parcours proposés par l'école. Dans le parcours Micro- et Nano-Electronique, les étudiants peuvent choisir lors du 4^e semestre une UE optionnelle de 50 heures qui s'intitule « Biosystèmes » visant à découvrir les interfaces possibles entre la microélectronique et la biologie : biologie synthétique, bien sûr, mais également tout ce qui concerne les biosystèmes, biocapteurs intégrés, micro-fluidique etc. Cette UE peut être complétée lors du 5^e semestre par une UE « Biologie Synthétique » de 70 heures qui permet aux étudiants de travailler sous la forme de projets sur les problématiques présentées dans le paragraphe précédent.

En formation TIC-Santé, les étudiants suivent un tronc commun de 4 semestres dans lequel ils ont des cours d'ouverture vers la biologie et les biotechnologies, les biosystèmes et les biocapteurs, les notions fondamentales de la biologie synthétique ainsi que les techniques utilisées pour la mesure et d'observation de phénomènes biologiques (micro-fluidique, fluorescence etc.). A l'issue de ce tronc commun, les étudiants choisissent entre deux parcours, l'un orienté vers le diagnostic et la robotique chirurgicale et l'autre orienté vers les thérapies innovantes, dont fait bien entendu partie la biologie synthétique. Lors du 5^e semestre, ils suivent la même UE que les formations généralistes et travaillent sur les mêmes projets en apportant un regard différent sur les problématiques abordées.

En plus de ces UE, un certain nombre de sujets de stage et de projets sont proposés aux étudiants en lien avec nos activités de recherche. Cette forme de travail permet un bon équilibre entre les recherches amont développées par les enseignants-chercheurs au sein de l'Institut d'Electronique du Solide et des Système (InESS) et le développement des outils associés mené par les étudiants. De plus, une certaine synergie se crée entre les étudiants de Telecom Physique Strasbourg et leurs voisins de l'Ecole Supérieure de Biotechnologie de Strasbourg (ESBS) qui apportent également un regard extérieur pertinent à nos travaux. Ensemble, ils ont déjà participé à trois reprises au concours iGEM (concours international inter-universités organisé annuellement par le MIT autour des différentes problématiques liées à la biologie synthétique). En 2009 et 2010, les étudiants de Telecom Physique Strasbourg sont venus en support des travaux des étudiants de l'ESBS sur la mise au point de nouvelles BioBriques, notamment sur les aspects liés à la modélisation. En 2011, Telecom Physique Strasbourg a participé pour la première fois en nom propre à ce concours, dans la session « software » et y a décroché une médaille de bronze. Une nouvelle participation est envisagée en 2014, toujours dans la session « software », avec une équipe mixte composée d'étudiants des deux écoles. De plus, afin de pérenniser cette collaboration, une association « TPS Biologie Synthétique » a été créée en juin 2012 [14].

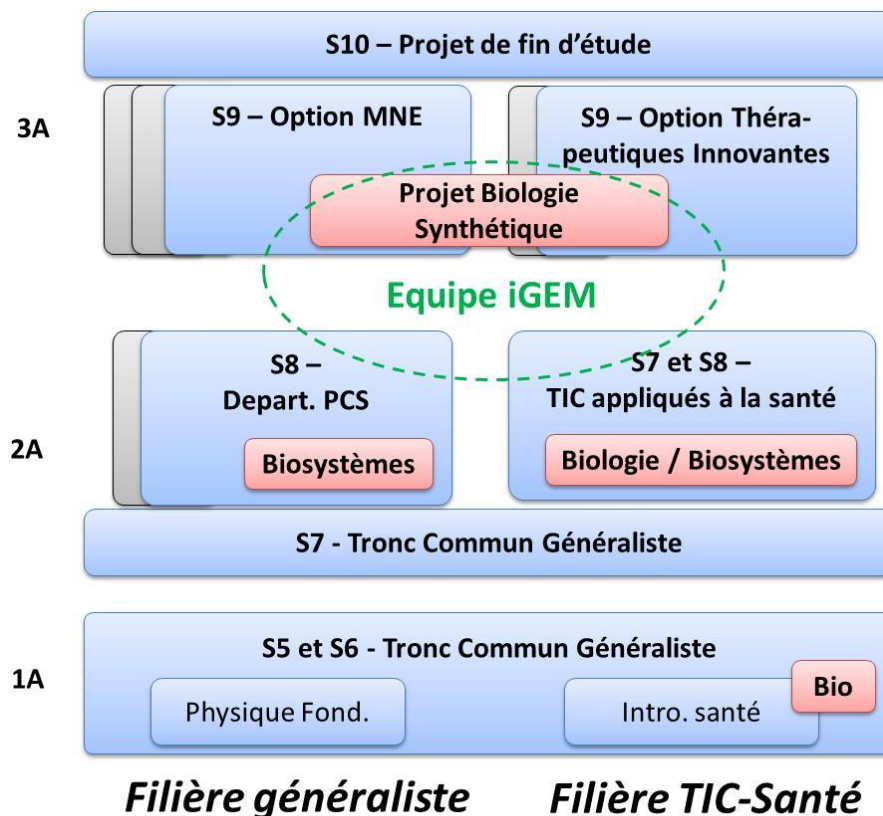


Fig. 3 – Organisation de la formation autour de la Biologie Synthétique à Télécom Physique Strasbourg.

5. EXEMPLES DE REALISATION

Dans la dernière partie de ce document, nous vous présenterons quelques travaux réalisés par les étudiants dans le cadre de ces formations.

5.1. Synthétiseur automatique de modèles VHDL-AMS pour les BioBriques.

Si nous voulons pouvoir réutiliser tout ou partie du flot de conception de la microélectronique, il faut, dans un premier temps, pouvoir s'assurer que les modèles des comportements biologiques d'une BioBrique puissent être décrits par les langages de description de la microélectronique, comme par exemple VHDL-AMS. L'équipe SMH d'ICube a donc exploré cette voie et a mis en avant plusieurs manières d'écrire ces modèles en exploitant au maximum le potentiel des langages. L'approche conservatrice multi-domaines semble être la plus proche de la réalité biologique [9]. Contrairement à la microélectronique, où les électrons sont les seuls à porter l'information, en biologie, plusieurs molécules interagissent entre elles et il y a donc plusieurs vecteurs d'information. Notre approche consiste à différencier chaque type de molécule et à lui attribuer une nature (au sens du VHDL-AMS). Les interactions entre les molécules sont modélisées par des boîtes faisant intervenir des terminaux de natures différentes, tous connectés à un même bus multi-domaine représentant la cellule (Fig. 3).

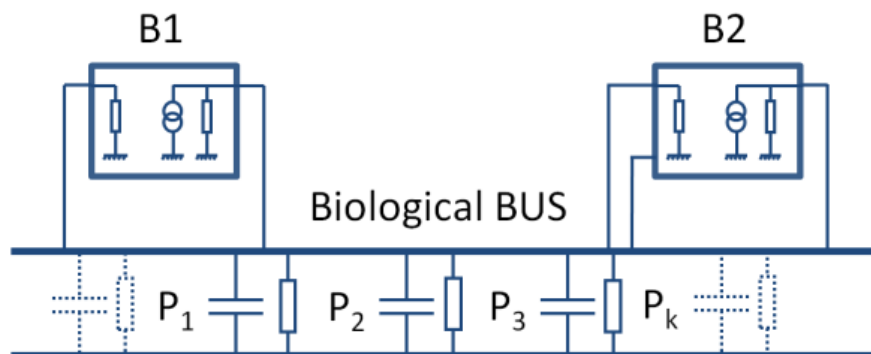


Fig. 4 – Formalisme proposé pour la description des BioBriques

Par exemple, dans le cas du circuit de la Fig. 1, nous aurions 6 natures (salicylate, arabinose, subD, T7ptag, T7 et GFP) et quatre boîtes contenant respectivement : le modèle de synthèse de subD régulé par la quantité de salicylate, le modèle de synthèse de T7ptag régulé par la quantité de arabinose, le modèle de la réaction de complexation entre subD et T7ptagb et le modèle de synthèse de GFP régulé par la quantité de T7). L'implémentation de ces modèles en VHDL-AMS est donc fastidieuse car, d'une part, il faut créer les fichiers de définition de chaque nature, puis recréer les composants élémentaires pour chaque nature. De plus, comme les terminaux des modèles sont de nature différente, il est impossible de créer un modèle générique que l'on appellerait plusieurs fois : il faut donc créer un fichier *.vhd* par boîte.

Pour simplifier et automatiser cette tâche, l'idée serait de construire un logiciel capable de prendre, en entrée, une description du type de la Fig. 2 et de générer automatiquement l'ensemble des fichiers VHDL-AMS permettant la simulation du système. Pour cela, les étudiants ont tout d'abord créé une interface graphique permettant à l'utilisateur de définir les différentes espèces intervenant dans la BioBrique ainsi que les interactions possibles entre ces espèces. Le moteur du logiciel, développé en C++, compile alors toutes ces informations pour créer un jeu de fichiers *.vhd* (libraires, modèles

élémentaires, fonctions, configurations) à partir de *templates* vierges. L'interface graphique permet également de définir certains scénarii de test et de synthétiser les test-bench associés. Enfin, les modèles étant simulés ensuite sous SMASH, les fichiers *.nsx* et *.pat* nécessaires à la compilation et à la simulation sont également générés automatiquement.

Ce travail est illustré sur un cas concret. Il s'agit d'un biosystème synthétique, mis au point par Xie *et al.* [16] qui constitue aujourd'hui l'état de l'art de ce qui est réalisable avec les technologies à disposition. Le système, composé d'une dizaine de BioBriques, est conçu pour détecter une cellule cancéreuse en synthétisant une protéine fluorescente rouge (DsRed) par analyse de sa « signature biologique ». Une cellule saine contient en effet un certain nombre d'espèces, appelés micro ARN, qu'une cellule tumorale ne possède pas. A l'inverse, en fonction de la pathologie, celles-ci possèdent d'autres ANR qui ne sont pas présentes dans les cellules saines. La description logique du système conduit à une équation Booléenne :

$$DsRed = miR21 \cdot (miR17 + miR30a) \cdot \overline{miR141} \cdot \overline{miR142} \cdot \overline{miR146a}$$

Elle n'est cependant pas suffisante pour obtenir une analyse fine de la robustesse du système. L'outil développé est donc utilisé pour générer un modèle bas-niveau du système. Celui-ci est constitué de pas moins de 60 fichiers VHDM-AMS, impliquant une dizaine de natures et plus de 300 quantités. Les résultats des simulations, présentés sur la Fig. 4, sont en accord avec les résultats expérimentaux publiés dans [16].

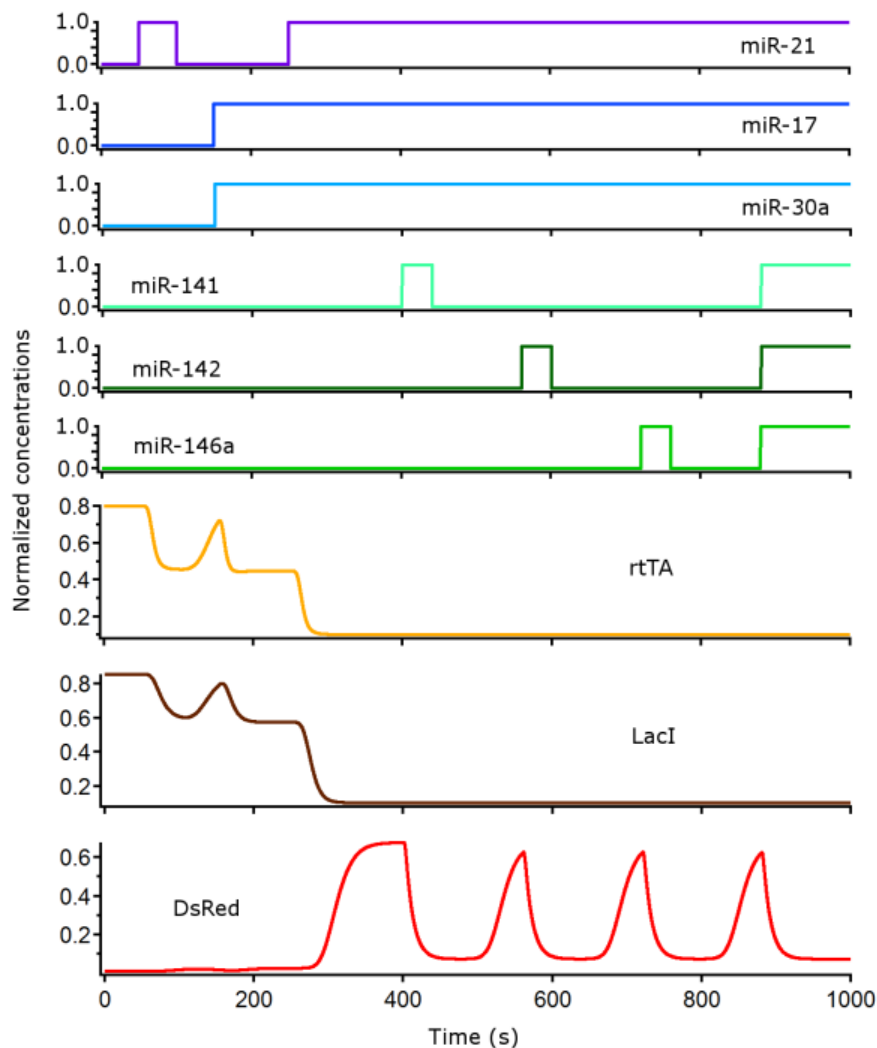


Fig. 5 – Résultats de simulation du détecteur de cellules cancéreuses

Ces travaux ont été publiés lors de la conférence IEEE Engineering in Medicine & Biology [13] et présentés par l'équipe Telecom Physique Strasbourg lors du concours iGEM 2011 [17].

5.2. Simulateur biologique bas-niveau s'inspirant du jeu de la vie.

Le simulateur biologique bas-niveau est un outil qui a pour vocation de rendre les mêmes services que les outils de T-CAD en microélectronique. Il s'agit de simuler le comportement réel d'une BioBrique, au niveau le plus élémentaire, c'est-à-dire la molécule. Cet outil a un intérêt double dans le flot de conception que nous envisageons. Le premier est de pouvoir prédire, de manière fidèle, le comportement d'une BioBrique en fonction de différents paramètres. Cela doit permettre une étude approfondie, *in silico*, des BioBriques et l'établissement de modèles simplifiés pour la conception. La deuxième est de pouvoir simuler très précisément, en bout de ligne, les systèmes synthétisés pour vérifier s'ils ont bien le comportement attendu. L'un des postulats de base, dans l'approche BioBrique, est que les briques soient indépendantes les unes des autres (comme c'est le cas pour les portes logiques en microélectronique). Les modèles haut-niveau sont construits sur ce postulat. La simulation bas-niveau permet de le vérifier.

Dans sa première version, le simulateur moléculaire est basé sur le principe d'un automate cellulaire. La cellule est divisée par un maillage en un certain nombre de cases. Les différentes espèces sont réparties aléatoirement au sein de ce maillage et peuvent évoluer selon plusieurs règles. En premier lieu, nous définissons des règles de déplacement de chaque espèce (aléatoire ou contrôlé, avec ou sans variation de mobilité ...). Ensuite, nous définissons des règles d'interaction entre les espèces : lorsque deux espèces A et B se retrouvent à partager la même maille, elles ont une certaine probabilité d'interagir pour former un complexe [AB]. Enfin, il y a des règles liées à la synthèse (apparition d'espèces dans une maille sous certaines conditions) et à la dégradation (probabilité de disparition d'une espèce à chaque pas de temps).

Sur ces principes, une première version du logiciel a été réalisée sous MATLAB. Il est composé d'un moteur de simulation et d'une interface graphique permettant de spécifier le problème ainsi que les différents paramètres des espèces (mobilité, probabilité de réaction, etc.). Cette première version a été validée sur un exemple simple et classique de la biologie : le cas d'un tétramère (molécule susceptible d'interagir avec au maximum 4 ligands). Dans ce type d'étude, pour obtenir un modèle comportemental du mécanisme, nous nous intéressons au nombre moyen de ligands liés sur chaque tétramère de la cellule θ_A en fonction de la concentration initiale de ligand $[B]$. Pour décrire ce comportement, nous utilisons le plus souvent un modèle mathématique basé sur l'équation de Hill :

$$\theta_A = \frac{\theta_{max}}{1 + \left(\frac{K}{[B]}\right)^n}$$

où θ_{max} est le nombre maximal de ligand (4 dans le cas d'un tétramère), K est la concentration pour laquelle la moitié des sites sont occupés et n est le coefficient de Hill. Ce coefficient a la particularité de varier en fonction de la coopérativité de liaison du tétramère. Si le fait d'avoir un ou plusieurs sites occupés favorise la liaison avec d'autres ligands, nous parlons de coopérativité positive et n tend vers θ_{max} . Dans le cas contraire, nous parlons de coopérativité négative et $n < 1$. Enfin, si la configuration du tétramère n'influe pas sur la probabilité de liaison, nous parlons d'un mécanisme non-coopératif et le coefficient de Hill vaut 1. Sur cet exemple, le simulateur moléculaire permet de vérifier ces propriétés ainsi que la validité du modèle de Hill.

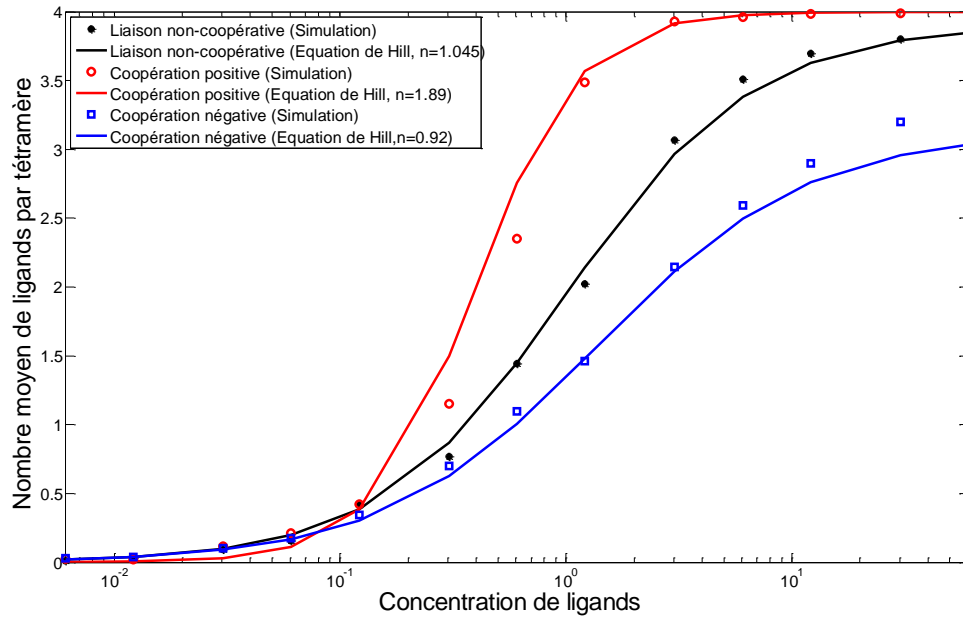


Fig. 6 – Résultat de simulations obtenues avec le simulateur moléculaire pour le cas d'un tétramère. Pour les différents cas, les résultats de simulations sont les marqueurs tandis que les courbes en trait plein sont les modèles mathématiques correspondants.

Le logiciel a également été utilisé pour simuler quelques systèmes plus complexes et en déduire leurs caractéristiques macroscopiques à partir du comportement microscopique. La Fig. 7 montre un résultat de simulation de la porte ET représenté sur la Fig. 2. Par soucis de simplicité, les deux premiers gènes ne sont pas implémentés. Les règles sont donc les suivantes : les points verts et jaunes représentent les deux protéines supD et T7Ptag. Lorsque ces deux protéines se retrouvent sur la même case, elles peuvent se lier pour donner la protéine orange. La zone grise représente la zone où la traduction et la transcription se produisent. Lorsque les molécules orange s'y lient, il y a synthèse de la GFP, représentée ici par les molécules bleues.

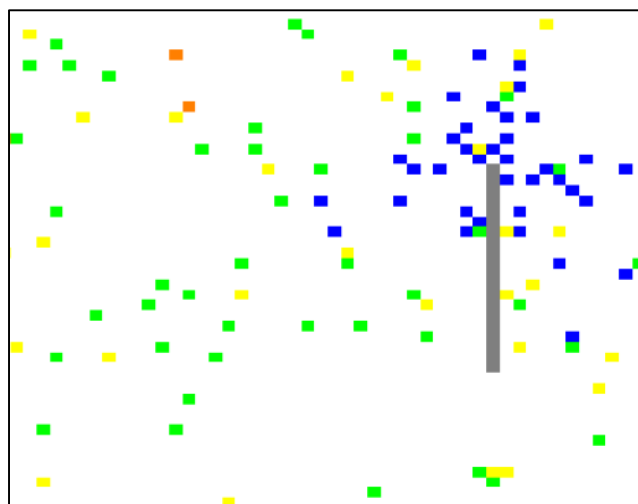


Fig. 7 – Résultats de simulation de la porte ET d'Anderson (Fig. 2). Les points verts, jaunes, orange et bleus représentent respectivement supD, T7Ptag, T7 et GFP. La zone grise correspond à un espace localisé dans la cellule où les mécanismes de traduction et de transcriptions peuvent se produire. Au final, on retrouve bien le comportement de la porte ET : bleu = vert AND green.

La première version du simulateur a également été présentée lors de la conférence *Engineering in Medicine & Biology* [18] et est actuellement utilisée pour illustrer des travaux pratiques durant la première année de la formation TIC-Santé à Telecom Physique Strasbourg ainsi que pour certaines applications au sein de notre équipe de recherche. Une deuxième version du simulateur est en cours de préparation. Celle-ci devrait être optimisée et permettre de simuler des mécanismes plus complexes. Elle devrait servir de support au projet qui sera présenté par les étudiants lors du concours iGEM 2014.

5.3. Modélisation des mécanismes biologiques en utilisant la logique floue.

Les deux niveaux de modélisation présentés précédemment (digital et analogique) possèdent chacun des propriétés intéressantes. Le premier est simple à manipuler ce qui permet d'envisager la réalisation d'outils de synthèse automatisés comme en électronique, l'autre est précis et permet une analyse fine des systèmes biologiques. Néanmoins, ces deux représentations des mécanismes biologiques restent très lointaines. Dans ce contexte, la logique floue peut constituer un niveau de modélisation très intéressant entre le niveau digital et le niveau analogique [19]. Il permet de combiner les avantages des deux approches : le système est décrit par l'intermédiaire de proposition logique mais permet de donner une estimation quantitative de la valeur de sortie en fonction des entrées.

Plusieurs projets étudiants ont été menés autour de l'utilisation de la logique floue :

- L'implémentation de la logique floue sous MATLAB puis le portage vers les langages de description de matériel de type VHDL ou SystemC-AMS ;
- La comparaison de plusieurs méthodes de calcul propres à l'algorithme de logique floue ;
- L'influence du nombre de « niveaux » utilisés pour la description des systèmes ainsi que de la manière dont sont décrites les règles de fonctionnement ;

L'intérêt de ce niveau de modélisation est illustré sur le cas du détecteur de cellule cancéreuse de la partie 5.3. Comme nous l'avons déjà évoqué, la description logique est simple mais insuffisante pour étudier finement le système. A l'inverse, la description analogique est complexe et le temps de simulation, pour une configuration donnée, excède largement la seconde. La logique floue permet quant à elle une simulation rapide et relativement précise. A titre de comparaison, nous avons simulé le détecteur pour 1000 combinaisons d'entrées différentes en logique floue et avec la description analogique. Les résultats sont cartographiés sur la figure 8, les croix bleues représentant les cellules saines et les croix rouges les cellules cancéreuses. La zone noire correspond à la zone à risque évaluée à partir des paramètres du système équations analogiques et des paramètres extraits de [16]. Les résultats des simulations en logique floue sont fiables à 99% et ont été obtenus en moins d'une seconde (pour les 1000 combinaisons), soit plus de 1000 fois plus vite que si on avait utilisé le modèle analogique. Les résultats dépendent clairement du nombre de pas utilisés pour la logique floue, mais la complexité du système, et donc le temps de simulation, augmente de manière quadratique avec celui-ci. Le nombre de 5, choisi ici, semble être un bon compromis.

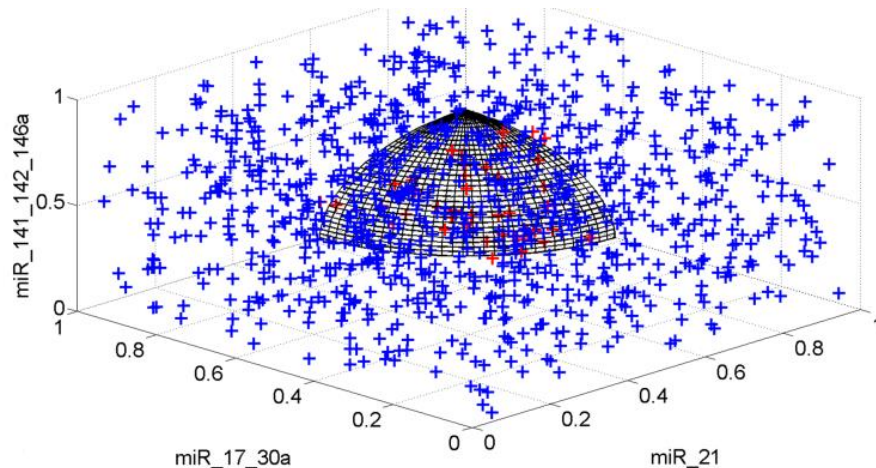


Fig. 8 – Cartographie des résultats de simulation du détecteur de cellule cancéreuse de Xie pour 1000 combinaisons d'entrée différentes. Les croix bleues et rouges représentent respectivement les cellules détectées comme saines et comme cancéreuses par le modèle en logique floue. La zone en noir correspond à l'espace des combinaisons d'entrée conduisant à une cellule cancéreuse (l'équation de l'espace est déterminée à partir du modèle analogique)

6. CONCLUSION

La biologie synthétique est une nouvelle science dont le fort potentiel est reconnu par l'ensemble de la communauté du domaine de la santé. De par leurs approches respectives, la biologie synthétique et la microélectronique sont très proches et de nombreux sujets d'études et projets industriels vont émerger à l'interface entre ces deux domaines. Pour être rapidement opérationnel, nous travaillons depuis 2008 sur des sujets de recherche en amont autour de cette thématique et sensibilisons des ingénieurs, micro-électroniciens de formation, à ces nouvelles problématiques. L'enseignement se fait majoritairement sous la forme de projets et en lien avec nos travaux de recherche, ce qui est bénéfique à la fois pour les étudiants et les encadrant de ces projets.

REMERCIEMENTS

L'équipe tient à remercier les étudiants de Telecom Physique Strasbourg ayant participé aux différents projets présentés dans cet article : Martin Andraud et Vincent Wlotzko (promo 2012), Camille Alliaga, Loïc Bauer, Alexandre Benoit, Pierre-Antoine Doisneau et Nicolas Kuntz (promo 2013), Paul Bonnet, Romain Cherpantier, Alizée Costard, Thibault Desbordes, Mathieu Dufour, Pierre Guinet, Pierre Leprovost, Hélène Moulet, Alexandre Phé, Martin Lemaire, Florian Rault, Jean-Pascal Caussé, Fabien Boyer et Clément Zidée (promo 2014) et Christophe Augusto, Jean Gazaniol, Thomas Baudier et Jérémy Bonhomme (promo 2015).

BIBLIOGRAPHIE

- [1] Endy R., "Foundations for engineering biology" in Nature, vol. 438 (24), pp. 449-453, 2005.
- [2] Stricker J. et al., "A fast, robust and tunable synthetic gene oscillator", in Nature, vol. 456, pp. 516-519, 2008.
- [3] Martin V. et al., "Engineering a mevalonate pathway in Escherichia coli for production of terpenoids" in Nature Biotechnology, vol. 21, pp. 796-802, 2003.
- [4] Weber W. et al., "Design of high-throughput-compatible protocols for microencapsulation, cryopreservation and release of bovine spermatozoa" in J. of Biotechnology, vol. 123, pp. 155-163, 2006.
- [5] Gaisser S. et al., "Making the most of synthetic biology: strategies for synthetic biology development in Europe", EMBO reports, 2009.
- [6] Ministère de la Recherche. Rapport du groupe de travail « sciences du vivant » sur les stratégies nationales de recherche et d'innovation 2009.
- [7] Shetty R. et al., "Engineering BioBrick vectors from BioBrick parts" in Journal of Biological Engineering, vol. 2 (2008).
- [8] R. Carlson, "The Changing Economics of DNA Synthesis", Nature Biotechnology, vol. 27, pp. 1091 - 1094, 2009.

- [9] Gendrault Y. et al., “Synthetic biology methodology and model refinement based on microelectronic modeling tools and languages” in *Biotechnology Journal*, vol. 6, pp. 796-806, 2011.
- [10] Knight T., “Idempotent Vector Design for Standard Assembly of BioBricks”, MIT Synthetic Biology Working Group TR, 2003.
- [11] Lux M. et al., “Genetic design automation: engineering fantasy or scientific renewal?”, in *Trends in Biotechnology*, vol. 30, pp. 120–126, 2011.
- [12] Anderson J. “Environmental signal integration by a modular AND gate”, in *Molecular Systems Biology*, vol. 3, 2007.
- [13] Gendrault Y. et al., “Using Digital Electronic Design Flow to Create a Genetic Design Automation Tool” in *Proc. of EMBC’2012*, San Diego (CA, USA), 28 aug. - 1 sept. 2012.
- [14] Synthèse de l’étude sur les besoins en compétences dans les Biotechnologies Santé, Plan Compétence Biotech 2015.
- [15] Site web de l’association TPS Biosynthétique : http://ticsante.u-strasbg.fr/doku.php?id=association_t.p.s.biologie_synthetique.
- [16] Xie Z. et al., “Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells,” in *Science*, vol. 333, pp. 1307-1311, 2011.
- [17] iGEM 2011 wiki for team ENSPS Strasbourg. <http://2011.igem.org/Team:ENSPS-Strasbourg>
- [18] Madec M. et al., “A game-of-life like simulator for design-oriented modeling of BioBricks in synthetic biology” in *Proc. of EMBC’2012*, San Diego (CA, USA), 28 aug. - 1 sept. 2012.
- [19] L.A. Zadeh, “Fuzzy Sets”, *Information and Control*, vol. 8, pp. 338-353, 1965.