

# L'ingénierie combinatoire des génomes

## Une clé pour la création de voies biosynthétiques artificielles

Denis Pompon, Thomas Lautier, Gilles Truan et Philippe Urban

**Résumé** Les méthodes combinatoires d'ingénierie métabolique, contrairement aux approches classiques de reconstruction et d'optimisation par essai-erreur, visent à générer en une seule étape une large gamme de combinaisons alternatives de biocatalyseurs, tant au niveau de leurs natures que de leurs niveaux d'expression. Ces combinaisons sont alors sélectionnées pour la production par le micro-organisme de la molécule d'intérêt. Cet article décrit comment de nouvelles approches de reconstruction et d'ingénierie des génomes rendent maintenant possible la mise en place de procédures d'ingénierie combinatoire permettant de créer soit une diversité moléculaire autour d'une plateforme structurale, soit d'accélérer considérablement la construction de voies de biosynthèses artificielles ciblant des molécules d'intérêt industriel.

**Mots-clés** Ingénierie des génomes, approches combinatoires, ingénierie métabolique, biologie de synthèse, biotechnologies.

**Abstract** **Combinatorial genome engineering: a key stone to build artificial biosynthetic pathways** Combinatorial approaches of metabolic engineering, in contrast to conventional methods of reconstruction and optimization by trial and error, aim to generate single-pot combination of a wide range of alternative biocatalysts, differing both in their natures as in their expression levels. These combinations are then selected for production by the microorganism of the product of interest. This article describes how the development of new approaches of reconstruction and genome engineering now makes possible the implementation of combinatorial engineering procedures to create molecular diversity around a structural platform or greatly accelerate the construction of artificial biosynthetic pathways targeting molecule of industrial interest.

**Keywords** Genome engineering, combinatorial methods, metabolic engineering, synthetic biology, biotechnology.

### Glossaire

Les termes suivis d'un astérisque\* sont définis ci-dessous.

**Exonucléase T5** : enzyme capable de dégrader un ADN de manière processive à partir de ses extrémités. Il existe grand nombre d'exonucléases qui se distinguent par leurs propriétés (sens de dégradation 3'-5' ou inverse, simple ou double brin, etc.).

**Hybridation** : réaction par laquelle deux brins d'ADN complémentaires s'apparient de manière antiparallèle par une série de liaisons hydrogène entre bases.

**Ligases** : enzymes capables de suturer deux fragments d'ADN en formant une liaison covalente entre l'extrémité 5'-phosphate d'un brin d'ADN et l'extrémité 3'-(désoxyribose) d'un autre brin.

**Métagénomes** : ensemble des séquences d'ADN contenu dans un échantillon naturel de matière hétérogène provenant d'un environnement spécifique. L'ADN provient des organismes contenus dans l'échantillon sans que l'on sache *a priori* associer une séquence de la collection à un organisme spécifique.

**Polymérases** : enzymes capables de synthétiser un ADN à partir de nucléotides triphosphates en utilisant comme matrice un ADN ou un ARN complémentaire. Les polymérases utilisées pour la PCR (réaction de polymérisation en chaîne) sont thermostables.

**Recombinaison** : mécanisme biologique complexe capable de fusionner ensemble deux ADN double brin au niveau de séquences identiques ou très semblables. Un double événement de recombinaison est impliqué dans les processus de réparation de l'ADN à la manière dont on réparerait un trou dans du tissu par une pièce en respectant la concordance des motifs sur les bords.

**Rétrosynthèse** : approche pour créer un schéma de synthèse d'une molécule complexe consistant à partir de la molécule cible à la décomposer de manière virtuelle en sous-ensembles de plus en plus petits jusqu'à rejoindre des molécules disponibles. La décomposition est effectuée de manière à ce que la recombinaison soit possible par des mécanismes chimiques ou des activités enzymatiques connues.

**XNA** : abréviation formée par analogie avec DNA (ADN) et caractérisant une structure moléculaire comportant de l'information génétique codée par des bases non naturelles. La lettre X rappelle le caractère non naturel.

L'ingénierie métabolique des micro-organismes représente une alternative innovante et durable pour la synthèse d'une large gamme de molécules d'intérêt industriel. Une telle ingénierie implique l'expression simultanée d'une série de biocatalyseurs, souvent issus de la biodiversité naturelle, voire d'ingénierie associée à la surexpression ou, à l'inverse, à l'atténuation de gènes endogènes.

L'ingénierie des biocatalyseurs s'est considérablement développée au cours des vingt dernières années, tant à un niveau conceptuel, au travers des approches *in silico*, qu'à un niveau de l'accès à la biodiversité au travers de l'accumulation des séquences de génomes et de métagénomes\* [1]. L'ingénierie par évolution dirigée des protéines a aussi joué un rôle majeur en permettant la génération d'activités non naturelles ou l'adaptation, souvent décisive, de biocatalyseurs à des conditions industrielles. La variable d'ajustement est alors la nature des acides aminés au sein de la séquence des biocatalyseurs, le principe étant de générer de la diversité, puis de sélectionner et recombiner les changements les plus favorables à une activité donnée [2]. Cet enjeu a donné lieu à de très nombreuses solutions, tant autour de la manière rationnelle ou semi-rationnelle de choisir la combinatoire initiale, que des solutions technologiques pour construire les banques ou redistribuer les solutions gagnantes entre les séquences. Le concept est maintenant de remplacer l'unité « acide aminé » par l'unité « gène » et l'optimisation d'une activité enzymatique isolée par la recherche d'une voie biosynthétique optimale intégrée au

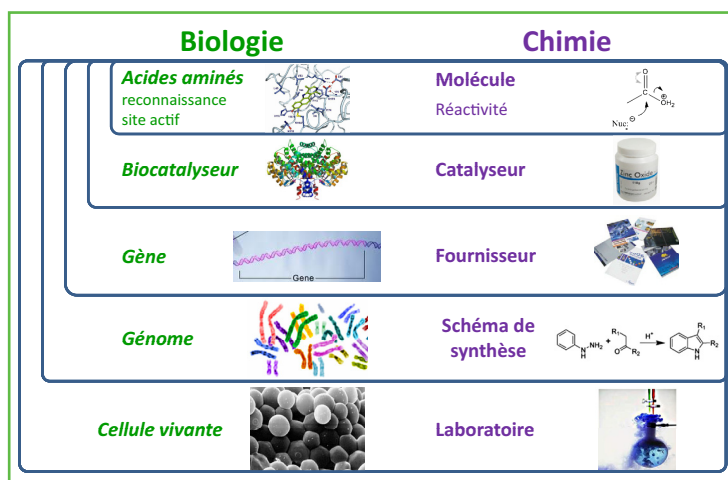


Figure 1 - Biologie et chimie : un parallèle aux différentes échelles des approches synthétiques.

#### Encadré 1

##### Notion d'organismes donneurs et accepteurs

Lors de la reconstruction d'une voie métabolique, on distingue « l'organisme accepteur », qui est celui dans lequel la voie métabolique artificielle va être reconstruite, du ou des « organismes donneurs », qui sont les espèces qui vont servir de sources pour les gènes à assembler. Ces sources peuvent être quelconques puisque le transfert d'un gène entre le donneur et l'accepteur n'est en général pas direct et fait l'objet d'un recodage qui permet de rendre compatibles les informations transférées ou d'en modifier certaines parties (niveau d'expression ou régulation par exemple). Il est donc possible d'assembler dans un même « organisme accepteur » des gènes d'origines très différentes issus de génomes, de métagénomes, voire partiellement ou totalement artificiels. Le choix de l'organisme accepteur est critique en fonction de l'application ciblée. En effet, celui-ci va fournir tout l'environnement de la biosynthèse, y compris les différentes molécules énergétiques nécessaires. Il faudra aussi éviter que le métabolisme propre de « l'accepteur » perturbe de manière incontrôlée la voie biosynthétique artificielle.

sein d'un organisme (*figure 1*). En d'autres termes, il s'agit de transposer les techniques éprouvées d'évolution dirigée des protéines à la création, par des approches combinatoires, de voies métaboliques artificielles au sein d'un organisme hôte.

## Comment concevoir une voie biosynthétique artificielle

La première approche à laquelle on pense est de copier la Nature – le « magasin du Bon Dieu » comme aimait à dire Pierre Potier – qui met à notre disposition un nombre incroyable d'exemples au travers des voies de biosynthèse naturelles. La taille accessible de ce « magasin » a littéralement explosé au cours des dernières années avec le développement des approches de séquençage à très haut débit des génomes et des métagénomes (génomes de mélanges d'organismes non définis) associées à la mise en place de plateformes performantes d'analyse des fonctions enzymatiques correspondantes [3]. Dans le schéma le plus simple, une voie artificielle de biosynthèse va constituer la copie presque conforme d'une voie naturelle, et n'en différer que par les optimisations permettant de cibler la molécule d'intérêt avec des rendements en matière et en énergie optimisés. Souvent, il va s'agir d'une transposition entre un organisme « donneur » qui sait faire, mais est mal adapté à un procédé industriel, et un organisme « accepteur », qui ne sait pas faire

naturellement ou fait mal, mais qui est bien adapté aux contraintes industrielles (voir *encadré 1*). Cette transposition peut néanmoins poser de nombreux problèmes, en particulier de connexion et d'interférences entre le métabolisme artificiel et le métabolisme naturel de « l'hôte » d'accueil.

Un second cas plus complexe est celui où une espèce « source » produisant la molécule d'intérêt est connue, mais où la voie de biosynthèse sous-jacente est partiellement voire totalement inconnue. Le cas extrême est celui où aucun organisme connu ne produit la molécule d'intérêt. S'ouvre alors une démarche centrale à la biologie de synthèse : tenter de copier la voie métabolique ou la reconcevoir. En effet la logique de la Nature n'est pas celle de l'ingénieur. Les voies naturelles de biosynthèse font partie d'un « système » qui ne vise qu'un objectif, la survie et l'avantage compétitif de l'espèce. À l'inverse, une voie biosynthétique artificielle cible une molécule ou une famille de molécules *a priori* sans intérêt spécifique pour l'organisme hôte. En d'autres termes, le biotechnologiste peut avoir tout intérêt à s'éloigner de la logique de la Nature, dans une démarche se rapprochant alors de celle du chimiste. La première étape est de constituer un catalogue des réactions possibles au travers d'un ensemble de bases de données [4] ; la seconde étape est de mettre en place une approche de rétrosynthèse\* partant de la molécule cible pour rejoindre, aux travers de biocatalyseurs connus (ou possibles à construire), une molécule du métabolisme d'un organisme hôte utilisable pour un bioprocédé. Explorer les chemins possibles est maintenant réalisable grâce à différentes approches bioinformatiques basées sur l'analyse de flux métaboliques et de bilans matière et énergie [5]. Une telle approche dégage deux types de contraintes :

- l'une relative à la collection de biocatalyseurs nécessaires (spécifiques et non plus seulement génériques), qu'il faudra alors identifier dans la biodiversité ou créer ;
- l'autre au niveau du couplage avec le métabolisme de l'hôte, qui doit bien sûr être rendu non limitant.

Une constante de ce type d'approche est malheureusement de générer la plupart du temps de nombreuses solutions potentielles difficiles à documenter sans aller à un niveau expérimental. À ce stade, deux solutions : soit faire un pari en privilégiant une solution et la tester puis tenter de l'optimiser, soit s'en remettre au hasard en mettant en place une stratégie combinatoire capable d'explorer en parallèle un grand nombre de solutions alternatives (*figure 2*).

## Des gènes synthétiques : la boîte à outils

La construction d'une voie métabolique synthétique passe par l'assemblage des gènes qui codent pour les différentes étapes enzymatiques impliquées (*figure 3*). Chacun de ces gènes est caractérisé par deux éléments principaux : un élément dit « promoteur », qui contrôle dans quelles conditions et à quel niveau un gène va s'exprimer, et un élément dit « codant », qui définit la séquence en acides aminés et, par conséquent, l'activité du biocatalyseur et d'autres aspects annexes comme sa localisation dans la cellule. L'ingénierie de promoteurs synthétiques [6] permet de générer une large gamme de niveaux d'expression dont l'ajustement (voire la régulation) est critique pour l'équilibre de la voie de biosynthèse artificielle. La plupart du temps, pour chaque étape de la biosynthèse, plusieurs choix d'enzymes naturels ou modifiés sont aussi possibles. Le développement commercial de la synthèse chimique d'ADN permet actuellement d'accéder rapidement et à bas coût (~ 0,30 € la paire

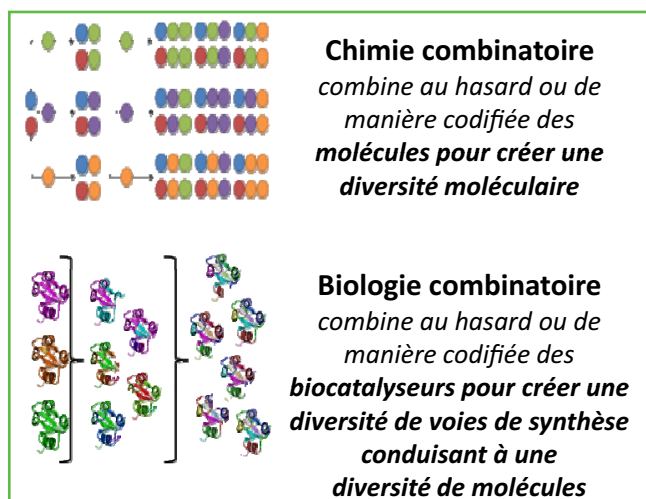


Figure 2 - La combinatoire des molécules à l'organisme.

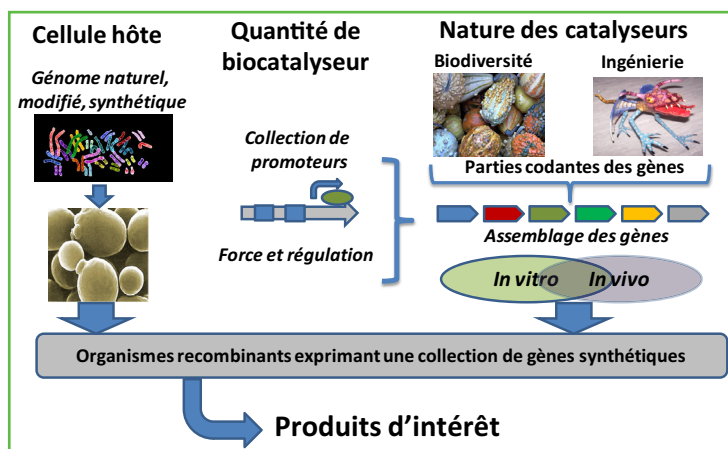


Figure 3 - Nature, quantité et environnement des biocatalyseurs s'assemblent pour former l'usine biologique.

de base) à n'importe quelle partie codante d'ADN à partir de sa simple séquence informatique, qu'elle soit naturelle ou artificielle. Les fragments d'ADN promoteurs, parties codantes et annexes sont alors facilement assemblés par ligation ou recombinaison\* enzymatique (voir encadré 2). Du fait des relations non linéaires entre le niveau de transcription, le niveau de la synthèse des protéines et l'activité enzymatique dans le contexte d'un micro-organisme, seule l'expérience peut à ce point valider les constructions optimales qui dépendent largement de la protéine considérée, même pour des montages génétiques comparables.

### De l'ingénierie individuelle à l'ingénierie collective des gènes

Dans l'approche classique de reconstruction métabolique (figure 4), celle-ci est menée étape par étape, les paramètres de chaque étape étant ajustés par essai-erreur. Le problème est que les mécanismes de couplage entre les étapes et les interactions parasites entre le métabolisme synthétique et le métabolisme de l'hôte sont souvent peu prévisibles, conduisant fréquemment à la nécessité d'importants retours en arrière pour aboutir à une optimisation globale. L'ensemble du processus est donc particulièrement long et mène souvent à une solution optimale en condition de laboratoire, mais qui doit être largement retravaillée lorsque le transfert en conditions industrielles est considéré. Les approches combinatoires, qui font l'objet de travaux récents [7-8], consistent au

#### Encadré 2

#### Quelques techniques innovantes pour l'assemblage de gènes

##### Une synthèse en phase solide pour l'assemblage de modules

Cette approche [18] est plus adaptée à l'association de petits fragments d'ADN et consiste à élonger séquentiellement un ADN amorce attaché par une extrémité sur un support solide. L'élongation est effectuée au travers de cycles d'hybridation et de ligation de petits oligonucléotides chevauchants. L'astuce consiste à débiter chaque étape d'élongation n par une étape de blocage par un excès d'un oligonucléotide spécifique de l'étape n-1. Ce dernier est choisi pour bloquer toute possibilité d'élongation ultérieure en cas d'échec d'une des étapes de couplage au cycle précédent. L'opération peut être facilement automatisée sur un robot. Une étape finale d'amplification par PCR (« polymerase chain reaction », réaction de polymérisation en chaîne) permet de n'amplifier que les fragments totalement élongés et de s'affranchir du support. Les fragments obtenus sont alors fusionnés entre eux par d'autres méthodes. De fait, cette approche est une transposition à l'échelle de l'assemblage de fragments d'ADN, de la méthode de synthèse en phase solide par voie chimique des oligonucléotides à partir de bases protégées. Par cette approche, des banques combinatoires de gènes ont pu être créées en quelques heures.

##### La méthode de Gibson

Cette approche [13], applicable à des fragments d'ADN de  $10^2$  à  $10^4$  paires de base, est fondée sur la génération d'extrémités cohésives entre fragments et permet d'assembler de manière spécifique plusieurs segments d'ADN en une seule étape. Elle consiste à dégrader partiellement les extrémités 5' de chaque fragment par l'exonucléase T5\*. Les brins complémentaires des segments d'ADN ainsi dénudés vont alors se reconnaître spécifiquement par hybridation\*. L'astuce consiste à introduire dans le même mélange une polymérase\* et une ligase\* thermostables et à faire la réaction à une température où seule l'exonucléase est instable. Dès qu'un hybride est formé entre fragments, la polymérase étend les extrémités 3' libres, alors que l'exonucléase cesse progressivement de dégrader les côtés 5-. Les trous initiaux se rebouchent donc et la ligase achève alors le travail de réparation en recollant les extrémités 3' et 5' libres, faisant disparaître toute trace de la fusion. Par cette approche, de nombreux fragments d'ADN peuvent être fusionnés dans un ordre parfaitement défini jusqu'à des tailles de l'ordre de  $10^4$  paires de base, voire au-delà. Ces fragments longs peuvent être alors associés entre eux par des méthodes de recombinaison *in vivo*.

##### La méthode GoldenBraid

Cette approche [19] est plus spécifiquement dédiée à l'assemblage contrôlé ou combinatoire de gènes complets. Dans une première étape, les gènes individuels sont assemblés sous forme de cassettes standardisées dont les extrémités comportent des sites convergents pour l'enzyme de restriction *BsaI*. Ces sites de coupure double brin sont particuliers car, bien que la séquence reconnue par l'enzyme soit présente aux extrémités de tous les gènes à assembler, la coupure s'effectue à une distance fixe du site de reconnaissance dans une zone où la séquence peut être quelconque. Les extrémités « simple brin » résultant de la coupure ne seront donc hybridables que pour les fragments présentant des extrémités aux séquences complémentaires. Il en résulte des règles d'association spécifiques qui permettent de contrôler à volonté, après réparation par ligation, le mode d'assemblage des différents gènes ou combinaison de gènes.

contraire à générer dès le départ une bibliothèque explorant le maximum de facteurs prévisibles (notamment la nature et les niveaux d'expression des enzymes) susceptibles d'influencer la biosynthèse ciblée. Chaque micro-organisme dans la bibliothèque obtenue va exprimer une combinaison différente de variants des gènes d'intérêt, voire des jeux de gènes partiellement différents si l'on recherche la génération de biodiversité moléculaire. Une étape de sélection fonctionnelle permet d'extraire de la population un sous-ensemble de gagnants (figure 4). Néanmoins, une solution optimale est rarement obtenue à ce stade pour deux raisons :

- le caractère généralement incomplet de la combinatoire initiale ;
- la présence d'interférences non optimales (souvent délétères) avec le métabolisme propre de l'hôte.

Afin de corriger le premier point, les solutions gagnantes sont alors mélangées au niveau génétique, générant une nouvelle bibliothèque (voir encadré 3). Par ailleurs, de nouveaux gènes « nourriciers » de l'hôte peuvent être surexprimés, détruits ou atténués à ce stade. Le processus de tri fonctionnel est alors recommencé jusqu'à convergence vers une ou plusieurs solutions optimales. Un intérêt de cette



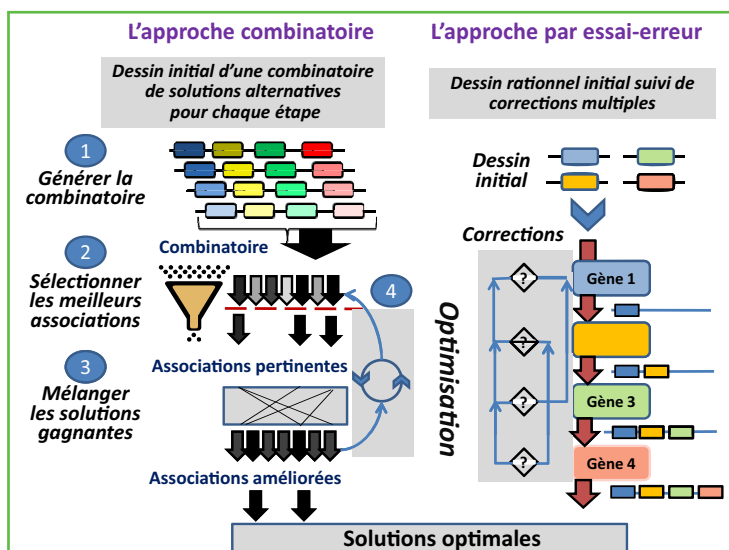


Figure 4 - Approches combinatoires et séquentielles de l'ingénierie métabolique.

### Encadré 3

#### Quand un parasite de plantes vient au secours de l'ingénieur

La réparation des cassures et des dommages de l'ADN est un mécanisme critique pour la vie. Lorsque cette cassure n'implique qu'un seul des brins de l'ADN, la réparation peut facilement être effectuée par les enzymes de maintien tels que les exonucléases\*, les polymérase\* et les ligases\* en utilisant l'information de l'autre brin pour reconstruire les séquences détériorées. Si la cassure et les dégâts impliquent les deux brins simultanément, ce mécanisme est inopérant et une « réparation d'urgence » par simple ligation conduit souvent à de la perte ou à des erreurs d'information. Un second mécanisme, la recombinaison homologue, peut alors prendre le relais. Dans ce mécanisme, la cellule va chercher « quelque part » une séquence ressemblant à celles présentes sur les deux bords de la cassure et tenter de copier ailleurs un fragment de séquence permettant de réparer, non seulement la cassure, mais aussi de reconstruire l'information qui aurait pu être perdue autour. Ce mécanisme avait été de longue date détourné par les biotechnologues comme outil d'assemblage de fragments multiples d'ADN ou pour modifier les génomes en créant des insertions et des délétions. Néanmoins, dans ce dernier cas, le processus naturel restait peu efficace, rendant son utilisation difficilement automatisable. Le problème restait de déclencher à volonté ce mécanisme. L'idée simple est de contrôler son facteur déclenchant : la formation d'une cassure double brin de l'ADN au bon moment et au bon endroit. Mais alors, comment faire dans un génome comportant des millions, voire des milliards de bases ? Une solution extrêmement performante est venue d'une bactérie parasite des plantes (*Xanthomonas*) qui fabrique des domaines protéiques modulaires (appelés TALE pour « transcription activator-like effector ») capables de reconnaître spécifiquement n'importe quelle séquence d'ADN d'un génome selon la nature et l'ordre d'assemblage de modules structuraux reconnaissant chacun une seule base d'ADN cible selon des règles tout à fait prévisibles. De nombreuses méthodes d'ingénierie des génomes basées sur les TALE [20] ou d'autres domaines synthétiques de reconnaissance [21] se sont donc développées autour de ce principe, mettant en œuvre une fusion entre de tels domaines et un domaine de nucléase rendue ainsi spécifique. Cette approche est applicable à l'ADN génomique et a donné lieu à plusieurs méthodes d'assemblage récurrentes de gènes facilement automatisables.

approche, outre la rapidité et le gain en ressource liés à l'analyse en parallèle des solutions, est que les contraintes industrielles peuvent facilement être introduites dans les derniers cycles combinatoires d'optimisation. Néanmoins, les difficultés techniques de ces approches combinatoires limitent encore à quelques démonstrateurs les exemples pratiques de leur mise en œuvre sur des problèmes industriels.

### Du gène au marché : quelques exemples

Un des exemples historiques de reconstruction réussie de voie complexe de biosynthèse à un niveau compétitif

avec la chimie a été le transfert, chez la levure de boulangerie, de la biosynthèse de l'hydrocortisone, un anti-inflammatoire (ou précurseur) important dont le marché mondial se chiffre autour de la centaine de tonnes/an. Ce projet, fruit d'une collaboration entre le CNRS et des industries pharmaceutiques, a débuté au début des années 1990 à une époque « pré-génomique » pour conduire au début des années 2000 à une preuve de concept préindustrielle [9]. Les modifications génomiques introduites, environ une quinzaine, correspondaient à l'introduction de gènes de mammifères, de plantes ou de micro-organismes et à la suppression de certains gènes de levure. Les optimisations réalisées (variation des niveaux d'expression des gènes, adressage des activités enzymatiques dans des compartiments cellulaires différents de ceux de leur organisme d'origine, et délétion des activités endogènes perturbatrices) ont été réalisées par une approche séquentielle d'essai-erreur couvrant une large gamme de solutions alternatives. Il est à noter que, dans ce cas, la gestion des relations entre métabolisme synthétique et métabolisme de l'hôte et les problèmes de compartimentation des réactions ont constitué les difficultés majeures. Il est très probable que dans le contexte post-génomique actuel et avec la mise en œuvre d'approches combinatoires, une réduction de deux à trois fois du temps nécessaire serait possible pour une ingénierie de même type.

Un second exemple concerne l'introduction réussie en 2006 chez la levure de la voie de biosynthèse de l'acide artémisinique, un précurseur de l'artémisinine (l'un des principaux médicaments antimalarial actuel) [10]. L'extraction de l'artémisinine à partir de sa source naturelle, une plante, était relativement coûteuse et sa synthèse chimique publiée en 1985 trop complexe pour être rentable. Après un travail remarquable de *design* génétique pour optimiser le taux de production du précurseur terpénique naturel, une souche de levure recombinante, qui produit suffisamment d'acide artémisinique pour envisager une production industrielle, a été obtenue. Les modifications introduites dans la levure, une quinzaine dans ce cas aussi, correspondaient à la fois en l'introduction de trois gènes provenant de la plante produisant l'artémisinine (*Artemisia annua*) et à la modification dirigée des niveaux d'expression d'une douzaine de gènes de la levure. Cette fois, grâce aux progrès réalisés dans les domaines de la génomique, ce travail d'optimisation a duré un peu plus de cinq années. La demande mondiale en artémisinine prévue pour 2013 est aussi de l'ordre de la centaine de tonnes, dont 10 % pourraient provenir de la voie recombinante.

Un autre exemple concerne la biosynthèse de caroténoïdes, une famille comptant plus de sept cents molécules différentes à la base d'un nombre important d'applications en nutrition animale ou humaine et en pharmacologie. La reconstitution des voies de biosynthèse de certains caroténoïdes dans des micro-organismes (levures et bactéries) a été largement explorée et est à l'heure actuelle une alternative extrêmement sérieuse à la chimie de synthèse. Différentes stratégies de changement de niveaux d'expression des voies synthétiques ont été développées chez la levure et *E. coli* pour améliorer la production de certains caroténoïdes jusqu'à des niveaux comparables aux meilleurs sources naturelles [11]. L'approche s'étend maintenant aux dérivés du  $\beta$ -carotène, notamment la synthèse de zéaxanthine et d'astaxanthine (figure 5). Par ailleurs, bien que la combinatoire naturelle soit déjà extrêmement riche, une combinatoire synthétique a permis à des chercheurs du groupe de F. Arnold [12] de produire des molécules contenant 35 carbones, alors que les caroténoïdes naturels sont plutôt des C30 ou C40. Ces

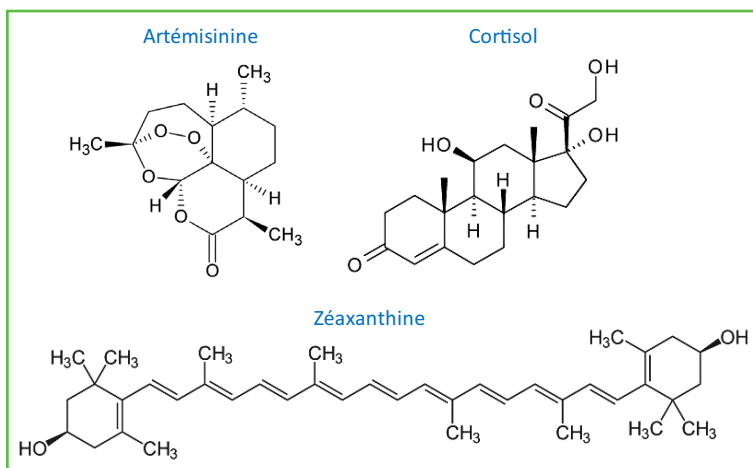


Figure 5 - Quelques formules.

premiers résultats montrent que la juxtaposition d'approches combinatoires en ingénierie des protéines et ingénierie génomique permettra non seulement de reconstruire des voies de biosynthèse, mais également de générer toute une gamme de molécules non naturel les d'intérêt.

## De nouveaux outils pour l'ingénierie des génomes

Les techniques d'ingénierie de séquences et de modifications ciblées des génomes (insertion, mutation, délétion de gènes) ont considérablement évolué au cours de la dernière décennie au travers du développement rapide de l'utilisation d'ADN synthétique et de procédures d'assemblage automatisables. La première révolution doit finalement beaucoup à la chimie et résulte de la baisse majeure des coûts de synthèse des oligonucléotides. Néanmoins, ce progrès initialement incrémentiel a permis une vraie révolution avec la possibilité d'étendre la taille des séquences pouvant être synthétisées de plusieurs ordres de grandeur (jusqu'au million de paires de base et même au-delà). Ceci a été rendu possible par l'association de différentes technologies complémentaires où des séquences courtes (quelques dizaines de nucléotides) synthétisées chimiquement sont d'abord assemblées *in vitro* jusqu'à des tailles atteignant la dizaine de kilobases, puis les blocs résultant assemblés par recombinaison chez la levure pour atteindre la mégabase [13]. Néanmoins, même si cette échelle est adaptée aux génomes bactériens, elle reste encore de deux à trois ordres de grandeurs trop faible pour la synthèse des génomes eucaryotes de plantes ou d'animaux.

La seconde évolution majeure provient du développement alternatif de stratégies d'édition massive beaucoup plus simples que la synthèse totale et mieux adaptées aux grands génomes. Il ne s'agit pas dans ce cas de la resynthèse complète d'un génome, mais d'y apporter un très grand nombre de modifications définies et ciblées (voir encadré 3). L'approche développée dans plusieurs articles récents [14] associe des techniques robotisées de mutagenèse dirigée à des méthodes de recombinaison permettant de combiner ensemble des groupes de mutations générés séparément. De telles approches ont été utilisées pour créer par exemple des souches de la bactérie *E. coli* dans le génome duquel un des codons de terminaison de la synthèse protéique a été systématiquement changé en un autre. Il en résulte un « codon libre » dans le code génétique qui peut être reprogrammé

pour coder pour un acide aminé non naturel, étendant ainsi le code génétique et la gamme des acides aminés utilisables dans les biocatalyseurs. Un autre exemple d'édition massive est la création de souches bactériennes « minimales » au sein desquelles un très grand nombre de gènes non essentiels ont été retirés. Le concept est alors de créer des « usines biologiques » minimales pouvant être reprogrammées plus efficacement pour des fonctions d'intérêt.

## L'assemblage combinatoire de gènes reste encore à maîtriser

Les approches permettant d'associer ensemble dans un organisme un grand nombre de gènes synthétiques constituent un élément clé de toute ingénierie métabolique. Ces dernières années ont vu le développement de nouvelles approches rapides et automatisables. Certaines, permettent par ailleurs de ne pas cibler une combinaison unique de gènes d'intérêt, mais de générer des bibliothèques combinatoires, créant ainsi des combinaisons variables de voies métaboliques synthétiques. Dans la pratique, deux classes complémentaires de solutions ont été développées. Les approches *in vitro* sont dérivées soit d'approches classiques de ligation enzymatique de mélange de gènes, soit d'approches plus originales dérivées des mêmes techniques que celles utilisées pour reconstruire des génomes synthétiques. L'une des plus performantes, développée par Gibson *et coll.* [13], permet d'assembler en une seule étape un grand nombre de fragments d'ADN (voir encadré 2). Les approches *in vivo* visent, elles, à construire directement au sein d'un micro-organisme d'intérêt l'assemblage combinatoire de gènes conduisant à la bibliothèque de voie métabolique. Plusieurs solutions techniques ont été publiées pour automatiser le procédé en résolvant en particulier le problème des marqueurs et des réactions parasites [15-16]. Néanmoins, ces solutions présentent généralement la limitation commune de conduire à l'assemblage d'un cluster de gènes synthétiques très délicat à modifier une fois formé. D'autres approches évitant cette limitation, grâce à des intégrations dispersées dans le génome, ont au contraire l'inconvénient d'être difficilement automatisables ou de conduire à des combinatoires biaisées.

## Le mélange de génomes

Mélanger deux génomes, c'est ce que font deux parents pour faire un enfant. Chez la plupart des organismes eucaryotes, ce processus conduit à une redistribution aléatoire des gènes parentaux au cours de la méiose notamment. Chez les micro-organismes tels que les levures, la méiose est un moyen simple pour redistribuer des caractères biotechnologiques d'intérêt, qu'ils soient naturels ou synthétiques. Une condition cependant : le processus n'est efficace que si les gènes à mélanger ne sont pas trop proches (> 100 kilobases) sur les chromosomes. Dans le cas d'une ingénierie impliquant des gènes synthétiques et/ou modifiés, ceux-ci doivent donc être relativement dispersés dans le génome, ce qui est peu compatible avec la majorité des méthodes d'ingénierie combinatoires déjà discutées. L'autre problème est que la méiose va mélanger au hasard, non seulement les gènes d'intérêt, mais aussi de nombreux autres dont l'effet est souvent inconnu. Ce phénomène va « brouiller » la combinatoire et rendre difficile l'utilisation pratique du mélange de génomes comme outil d'optimisation de voies métaboliques synthétiques. La solution, qui fait l'objet de recherches

actives, consiste d'une part à développer des techniques pour créer des banques combinatoires de qualité dispersées dans le génome et non plus localisées, d'autre part à une ingénierie de la machinerie de recombinaison méiotique elle-même, afin de contrôler quels gènes ou zones du génome seront autorisés ou non à s'échanger [17] (voir encadré 3).

## Biologie synthétique et biologie des systèmes : comprendre la Nature pour mieux l'employer

Bien connaître le fonctionnement métabolique de la cellule receveuse, châssis sur lequel se greffe la voie synthétique, est essentiel pour que la voie s'y intègre harmonieusement. La biologie des systèmes étudie des systèmes biologiques complexes comme un tout intégré, en utilisant des outils de modélisation, de simulation, et en les confrontant aux résultats expérimentaux. Elle s'appuie sur la quantité colossale de données issues des approches en « omiques » (génomique, protéomique, métabolomique, fluxomique, etc.). Traiter cette masse de données nécessite des techniques computationnelles basées sur la fusion de concepts de disciplines différentes, incluant la biologie, l'informatique, les mathématiques appliquées, la physique et les sciences de l'ingénieur. La biologie synthétique se nourrit de la biologie des systèmes par la prise en compte de la logique du vivant où l'on doit réaliser l'équation complexe de « respecter » un minimum l'hôte, si on veut qu'il travaille pour nous, tout en l'exploitant au maximum. L'ingénierie associée à la biologie synthétique va perturber de manière profonde le système hôte. Comprendre au travers d'approches de biologie des systèmes comment l'hôte va accommoder ces bouleversements représente aussi un enjeu majeur pour aboutir à des ingénieries plus efficaces ; « *La vie est une interaction entre molécules et non la propriété d'une seule* » (Linus Pauling). La biologie de synthèse construit des systèmes complexes en assemblant des briques biologiques sans trop se soucier de la logique du vivant (voir encadré 4). Mais couplée aux approches quantitatives, prédictives et dynamiques de la biologie des systèmes, elle permet d'appréhender le maillage essentiel des réseaux qui gouverne la stabilité du vivant, tant aux niveaux moléculaire que cellulaire. La biologie de synthèse nous place en tant qu'acteur sur cette scène. Il est donc essentiel pour gagner de découvrir les « règles du jeu » qualitatives et quantitatives qui gouvernent le comportement et la plasticité des systèmes biologiques.

## Références

- [1] Godzik A., Metagenomics and the protein universe, *Curr. Opin. Struct. Biol.*, **2011**, 21(3), p. 398.
- [2] Brustad E.M., Arnold F.H., Optimizing non natural protein function with directed evolution, *Curr. Opin. Chem. Biol.*, **2011**, 15(2), p. 201.
- [3] Ekkers D.M. *et al.*, The great screen anomaly: a new frontier in product discovery through functional metagenomics, *Appl. Microbiol. Biotechnol.*, **2012**, 93, p. 1005.
- [4] Wishart D.S., Computational approaches to metabolomics, *Methods Mol. Biol.*, **2010**, 593, p. 283.
- [5] Carbonell P. *et al.*, Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms, *B.M.C. Syst. Biol.*, **2012**, 6, p. 10.
- [6] Cox R. *et al.*, Programming gene expression with combinatorial promoters, *Mol. Syst. Biol.*, **2007**, 3, p. 145.
- [7] Sandoval N. *et al.*, Strategy for directing combinatorial genome engineering in *Escherichia coli*, *Proc. Natl. Acad. Sci. USA*, **2012**, 109, p. 10540.
- [8] Church G.M., Programming cells by multiplex genome engineering and accelerated evolution, *Nature*, **2009**, 460, p. 894.
- [9] Szczepara F. *et al.*, Total biosynthesis of hydrocortisone from a simple carbon source in yeast, *Nat. Biotechnol.*, **2003**, 21, p. 143.
- [10] Ro D.-K. *et al.*, Production of the antimalarial drug precursor artemisinic acid in engineered yeast, *Nature*, **2006**, 440, p. 940.

### Encadré 4

#### La biologie orthogonale

Un système orthogonal d'information génétique vise à créer un code génétique synthétique ne pouvant être interprété par les organismes naturels avec deux avantages potentiels : limiter les possibilités de dispersion dans l'environnement de l'information génétique artificielle et donner accès à un code génétique étendu pouvant être codé sur plus de quatre bases par exemple. Cette approche est basée sur le développement d'acides xéno-nucléiques (XNA\*) qui contiennent soit des bases non naturelles, soit un squelette s'éloignant du motif poly-désoxyribose phosphate [22]. Les limitations actuelles viennent d'une part de la nécessité de créer des processus de synthèse biologique de ces bases modifiées, d'autre part d'apprendre à l'organisme à répliquer de manière stable et à exprimer ce matériel génétique alternatif. Les acides aminés non naturels constituent une extension de cette approche. La combinatoire de la vingtaine d'acides aminés naturels reste limitée au niveau des fonctionnalités chimiques exploitables par les biocatalyseurs. La synthèse chimique en phase solide de peptides, les systèmes de traduction *in vitro* ou le développement de nouveaux aminoacyl-ARNt et de leur synthétase ont déjà permis de tester la fonctionnalité d'une centaine d'acides aminés non naturels, comportant une réactivité chimique, une fluorescence ou des possibilités de modification particulières. Toutefois, l'utilisation dans des systèmes vivants de ces analogues reste délicate et conduit à des biocatalyseurs souvent instables. Ces limitations pourront être atténuées par approches semi-rationnelles d'évolution dirigée et de modélisation [2].

- [11] Harada H., Misawa N., Novel approaches and achievements in biosynthesis of functional isoprenoids in *Escherichia coli*, *Appl. Microbiol. Biotechnol.*, **2009**, 86(6), p. 1021.
- [12] Umeno D., Arnold F.H., A C35 carotenoid biosynthetic pathway, *Appl. Environ. Microbiol.*, **2003**, 69, p. 3573.
- [13] Gibson D. *et al.*, Creation of a bacterial cell controlled by a chemically synthesized genome, *Science*, **2010**, 329, p. 52.
- [14] Isaacs F. *et al.*, Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement, *Science*, **2011**, 333, p. 348.
- [15] Winkler L. *et al.*, Reiterative recombination for the *in vivo* assembly of libraries of multigene pathways, *Proc. Natl. Acad. Sci. USA*, **2011**, 108, p. 15135.
- [16] Shao Z., Zhao H., Zhao H., DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways, *Nucleic Acids Res.*, **2009**, 37, p. e16.
- [17] Lichten M., de Massy B., The impressionistic landscape of meiotic recombination, *Cell*, **2011**, 147(2), p. 267.
- [18] Briggs A.W. *et al.*, Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers, *Nucleic Acids Res.*, **2012**, p. e117.
- [19] Sarrion-Perdigones A. *et al.*, GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules, *PLoS ONE*, **2011**, 6(7), p. e21622.
- [20] Li T. *et al.*, Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes, *Nucleic Acids Res.*, **2011**, 39, p. 6315.
- [21] Arnould S. *et al.*, The I-Crel meganuclease and its engineered derivatives: applications from cell modification to gene therapy, *Protein Eng. Des. Sel.*, **2011**, 24(1-2), p. 27.
- [22] Chaput J.C. *et al.*, The emerging world of synthetic genetics, *Chem. Biol.*, **2012**, 19(11), p. 1360.



D. Pompon



T. Lautier



G. Truan



P. Urban

Denis Pompon (auteur correspondant), Thomas Lautier, Gilles Truan et Philippe Urban sont chercheurs CNRS au Laboratoire d'ingénierie des systèmes biologiques et des procédés, LISBP-INSA, à Toulouse\*.

\* LISBP-INSA Toulouse, UMR INSA/CNRS 5504 - UMR INSA/INRA 792, 135 avenue de Ranguieu, F-31077 Toulouse Cedex 4.  
Courriel : denis.pompon@insa-toulouse.fr  
www.lisbp.fr