

CHEMOGÉNOMIQUE

**DES PETITES MOLÉCULES
POUR EXPLORER LE VIVANT**

**Une introduction à l'usage des biologistes,
chimistes et informaticiens**

sous la direction de

Eric MARÉCHAL - Sylvaine ROY - Laurence LAFANECHÈRE



17, avenue du Hoggar
Parc d'Activité de Courtabœuf - BP 112
91944 Les Ulis Cedex A - France

Grenoble Sciences

Grenoble Sciences poursuit un triple objectif :

- ▶ réaliser des ouvrages correspondant à un projet clairement défini, sans contrainte de mode ou de programme,
- ▶ garantir les qualités scientifique et pédagogique des ouvrages retenus,
- ▶ proposer des ouvrages à un prix accessible au public le plus large possible.

Chaque projet est sélectionné au niveau de Grenoble Sciences avec le concours de referees anonymes. Puis les auteurs travaillent pendant une année (en moyenne) avec les membres d'un comité de lecture interactif, dont les noms apparaissent au début de l'ouvrage. Celui-ci est ensuite publié chez l'éditeur le plus adapté.

(Contact : Tél. : (33)4 76 51 46 95 - E-mail : Grenoble.Sciences@ujf-grenoble.fr)

Deux collections existent chez EDP Sciences :

- ▶ la ***Collection Grenoble Sciences***, connue pour son originalité de projets et sa qualité
- ▶ ***Grenoble Sciences - Rencontres Scientifiques***, collection présentant des thèmes de recherche d'actualité, traités par des scientifiques de premier plan issus de disciplines différentes.

Directeur scientifique de Grenoble Sciences

Jean BORNAREL, Professeur à l'Université Joseph Fourier, Grenoble 1

Comité de lecture pour Chemogénomique

- ▶ **Jean DUCHAINE**, Conseiller principal de la Plate-forme de criblage, Institut de Recherche en Immunologie et en Cancérologie, Université de Montréal, Canada
- ▶ **Yann GAUDUEL**, Directeur de recherche à l'INSERM, Laboratoire d'Optique Appliquée (CNRS), Ecole Polytechnique, Palaiseau
- ▶ **Nicole MOREAU**, Professeur à l'Ecole Nationale Supérieure de Chimie, Université Pierre et Marie Curie, Paris
- ▶ **Christophe RIBUOT**, Professeur de Pharmacologie à la Faculté de Pharmacie, Université Joseph Fourier, Grenoble

Grenoble Sciences bénéficie du soutien du **Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche** et de la **Région Rhône-Alpes**.
Grenoble Sciences est rattaché à l'**Université Joseph Fourier de Grenoble**.

Réalisation et mise en pages : **Centre technique Grenoble Sciences**

Illustration de couverture : **Alice GIRAUD**

(d'après un cliché de puce - *Laboratoire Biopuces/Direction des Sciences du Vivant/CEA* et une photo de cellule embryonnaire de souris - *Yasmina SAOUDI, INSERM U836 Grenoble*)

ISBN 978-2-7598-0005-6

© EDP Sciences, 2007

EXTRAITS

INTRODUCTION

André TARTAR

Au cours des deux dernières décennies, la recherche en biologie a connu une mutation sans précédent qui s'est souvent traduite par l'adoption de techniques massivement parallèles, qu'il s'agisse du séquençage de génomes entiers, de l'utilisation des puces à ADN ou de la chimie combinatoire. Ces approches, qui ont en commun de multiplier les essais et erreurs pour en extraire quelques événements significatifs n'ont été rendues possibles que grâce aux progrès de l'informatique de la miniaturisation et de la robotique.

L'un des premiers secteurs à mettre cette approche en pratique a été celui de la recherche pharmaceutique avec l'utilisation systématique du criblage à haut débit pour la découverte de nouvelles cibles thérapeutiques et de nouveaux candidats médicaments. La recherche académique est restée longtemps à l'écart de cette démarche, tant pour des raisons de coûts que pour des motifs culturels. Depuis quelques années cependant, la banalisation de ces techniques a eu pour conséquence d'en diminuer considérablement le coût d'accès et de permettre ainsi à des groupes académiques de les mettre en oeuvre dans des projets ayant des objectifs généralement plus cognitifs.

Il n'en demeure pas moins capital, comme avec toute méthode extensive de prendre en compte le facteur coût comme une donnée fondamentale dans l'élaboration d'un protocole expérimental et de le mettre en rapport avec le bénéfice attendu. La valeur d'une chimiothèque est en effet une notion évolutive résultant de la somme de deux grandeurs qui évoluent en sens inverse :

- ▶ D'une part l'ensemble des échantillons physiques dont la valeur ira fatalement en décroissant du fait de leur consommation dans les tests, mais surtout de la dégradation des composés. L'expérience des dernières années montre également qu'elle subira les phénomènes de mode qui contribueront rapidement à son obsolescence : nul n'assemblerait aujourd'hui une chimiothèque comme il l'aurait fait il y a seulement cinq ans. Au vertige des grands nombres qui a régné sur les premières chimiothèques combinatoires s'est progressivement substituée une série de critères plus réalistes, témoignant les difficultés rencontrées. La "drugabilité" est ainsi devenue un mot clé avec la règle des 5 de LIPINSKI, les "frequent hitters" sont devenus la bête noire des cribleurs après leur avoir fait entrevoir tant d'espoirs non fondés.

- ▶ A l'opposé, la masse des informations accumulées au cours des différents tests de criblage ira en croissant et remplacera progressivement la chimiothèque physique. A une échéance plus ou moins lointaine, la chimiothèque physique aura disparu et seules resteront les informations qu'elle aura permis d'accumuler. Ces informations pouvant être utilisées soit directement, constituant la "fiche signalétique" d'un composé donné, soit comme source de références pour les exercices de criblage virtuel ou de prédiction *in silico* des propriétés de composés nouveaux.

Une analyse stratégique très simple montre qu'avec les moyens limités dont disposent les équipes académiques, il est plus facile d'être compétitif sur le second point (quantité et qualité de l'information) que sur le premier (nombre de composés et hauts débits). Elle montre également que la valeur d'une somme d'informations isolées est bien moindre que celle d'une matrice de données organisée de manière cohérente selon deux dimensions principales : la diversité des composés et la cohérence des tests biologiques.

C'est dans cette perspective qu'un criblage à haut-contenu d'information devrait s'imposer, permettant de collecter et stocker le maximum de données pour chaque expérimentation. Ce criblage à haut contenu sera le garant d'une valorisation optimale des collections physiques. Il est intéressant de constater que la problématique de la perte d'information lors d'une mesure a été au centre des préoccupations des spectroscopistes il y a des dizaines d'années. Aux systèmes dispersifs (prismes, réseaux...) qui sélectionnaient séquentiellement chaque longueur d'onde d'observation mais laissaient échapper toutes les autres, ils ont substitué les techniques d'analyse non dispersive confiant à des algorithmes de déconvolution et à des analyseurs multicanaux la charge de traiter l'information globale. La biologie est en pleine mutation à ce niveau : alors qu'il y a une dizaine d'années, on se contentait de suivre l'expression d'un gène sous l'effet d'un stimulus, aujourd'hui, grâce aux puces pan-génomiques, c'est le profil d'expression de l'ensemble du génome qui est devenu accessible. Il est impératif que le criblage suive cette évolution : ne plus perdre aucune information deviendra la règle. A plus longue échéance, il faudra que cette information soit formatée et stockée de manière durable et réutilisable.

Dans cette perspective, cet ouvrage tombe à point nommé car il constituera un outil de référence permettant aux différents spécialistes de parler un même langage, passage incontournable pour assurer la pérennité des informations accumulées.

Chapitre 1

LE PROCESSUS DE CRIBLAGE PHARMACOLOGIQUE : LA PETITE MOLÉCULE, LA CIBLE BIOLOGIQUE, L'AUTOMATE, LE SIGNAL ET L'INFORMATION

Eric MARÉCHAL - Sylvaine ROY - Laurence LAFANECHÈRE

1.1. INTRODUCTION

Le **criblage pharmacologique** met en œuvre des moyens techniques et technologiques pour sélectionner dans une collection de molécules, celles qui sont actives sur une entité biologique. Les pharmacopées antiques ou médiévales, dans lesquelles sont décrits les effets thérapeutiques de substances minérales et d'extraits de plantes, sont issues de criblages pharmacologiques dont les modes opératoires sont soit inconnus, soit très imprécis. Faute de documentation, on ne sait pas si la connaissance médicinale ancienne est issue d'une étude systématique réalisée avec méthode ou de la consignation d'un savoir collectif ayant grandement bénéficié d'expériences individuelles. Au cours des siècles, en plus de l'archivage de savoirs traditionnels et de leurs classifications, les recherches de nouveaux composés actifs se sont orientées vers des stratégies exploiratoires rationalisées, des cribles, en particulier des flores et de leurs extraits. Les approches basées sur un tri systématique ont fait leur preuve, par exemple lors de la recherche d'antibiotiques. Les progrès récents de la chimie, de la biologie, de la robotique et de l'informatique ont permis depuis les années 1990 une augmentation de la cadence des tests pour des criblages dits à **haut-débit** (en anglais, *high-throughput screening*), ainsi que des mesures de signaux multiparamétriques pour des criblages dits à **haut-contenu d'information** (en anglais, *high-content screening*). Au-delà des applications médicales qui ont motivé l'épanouissement des technologies de criblage dans les sociétés pharmaceutiques, la **petite molécule** (en anglais, *small molecule*) est devenue un outil formidable et accessible pour la recherche fondamentale. Les savoir-faire et concepts originaux nés autour du criblage automatisé ont donné naissance à une nouvelle discipline, la **chemogénomique** (BREDEL et JACOBY, 2004) dont un volet pratique est la **génétique chimique**, que nous aborderons plus particulièrement dans cet ouvrage (deuxième partie).

Le criblage pharmacologique implique des métiers très divers, qui ont leur propre culture, leur jargon, qu'il est difficile de comprendre non seulement entre biologistes, chimistes et informaticiens, mais aussi au sein de chacune de ces disciplines. Qu'est-ce qu'une "activité" pour un chimiste ou un biologiste, un "test" pour un biologiste ou un informaticien, ou encore un "contrôle" ? Une **terminologie commune** doit lever ces ambiguïtés. Ce chapitre introductif décrit brièvement les étapes d'un **processus** de criblage automatisé, donne un aperçu sur les différents types de collections de molécules, ou **chimiothèques**, et enfin aborde la question difficile des définitions de la **cible** et de la **bio-activité**.

1.2. LE PROCESSUS DE CRIBLAGE : APERÇU TECHNOLOGIQUE

1.2.1. PLAQUES MULTIPUITS, AUTOMATE ET DÉTECTEURS

Le criblage pharmacologique automatisé permet de tester en parallèle un grand nombre de molécules sur une cible biologique (extraits, cellules, organismes). Pour chaque molécule de la collection, le test permettant de mesurer un effet sur une cible biologique est mis en œuvre et un signal correspondant est mesuré. C'est sur la base de ce signal qu'un choix est effectué pour retenir des molécules intéressantes (fig. 1.1).

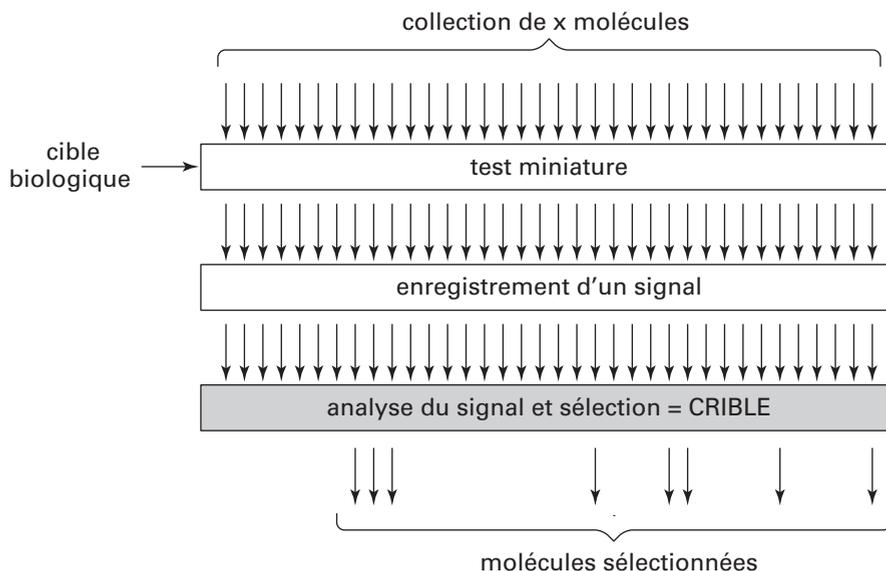


Fig. 1.1 - Schéma d'un processus de criblage pharmacologique

Les mélanges des molécules et de la cible ainsi que les processus nécessaires au test sont réalisés dans des plaques composées de puits (dites **plaques multipuits**, ou **microplaques**, fig. 1.2). Ces plaques sont de dimensions standardisées de 12, 24, 48, 96, 192, 384 ou 1536 puits. Lorsqu'on crible plus de 10 000 composés par

jours, on parle de **criblage "ultraminiaturisé"** (en anglais, μ HTS ou u HTS). Les gains de miniaturisation représentent des économies non négligeables. Les développements technologiques récents s'orientent aussi vers un criblage miniaturisé sur **puces**. Toutefois, le criblage étant un compromis technologique et expérimental visant la simplicité de mise en œuvre, la robustesse des tests et la fiabilité des résultats, le criblage automatisé n'est pas engagé dans une course générale à la miniaturisation. Plutôt qu'une augmentation effrénée de la cadence des tests et d'une sur-miniaturisation, les développements actuels du criblage s'orientent vers des méthodes permettant de renseigner au maximum les effets des molécules testées grâce à la mesure de paramètres multiples, voire la capture d'images obtenues à l'aide de microscopes (criblage à haut contenu ; deuxième partie). L'état actuel de la technologie de criblage automatisé fait donc encore, et probablement pour longtemps, appel aux manipulations sur plaques de 96 ou 384 puits.

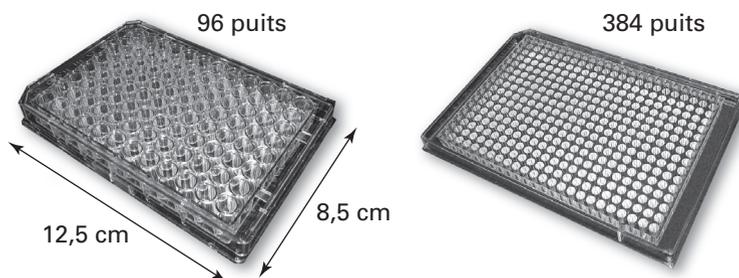


Fig. 1.2 - Plaques multipuits de 96 et 384 puits. (85 x 125 mm)

Les manipulations parallélisées sont réalisées à l'aide d'automates (fig. 1.3). Ces automates sont capables de réaliser des tâches séquentielles indépendantes telles que dilution, pipetage et répartition de composés dans chacun des puits des plaques multipuits, agitation, incubation, lecture de résultats. Ils sont pilotés par des logiciels spécifiquement adaptés au type d'opération à réaliser. Les essais sont effectués dans des microplaques standard identifiées par un code à barres et manipulées par la "main" d'un robot : par exemple, celui-ci prend la plaque vide, ajoute les réactifs nécessaires et les composés à tester, gère le temps de la réaction puis passe la plaque à un lecteur pour connaître le résultat. Pour visualiser les réactions issues de la mise en contact, dans les puits, de la molécule et de la cible, on utilise différentes méthodes basées sur des mesures d'absorbance, de radioactivité, de luminescence, de fluorescence ou encore des méthodes d'imagerie. Le processus de criblage et la collecte des données sont pilotés grâce à une station informatisée. Certaines étapes peuvent être réalisées indépendamment de l'ensemble robotique telles que la lecture de radioactivité, des analyses microscopiques manuelles...

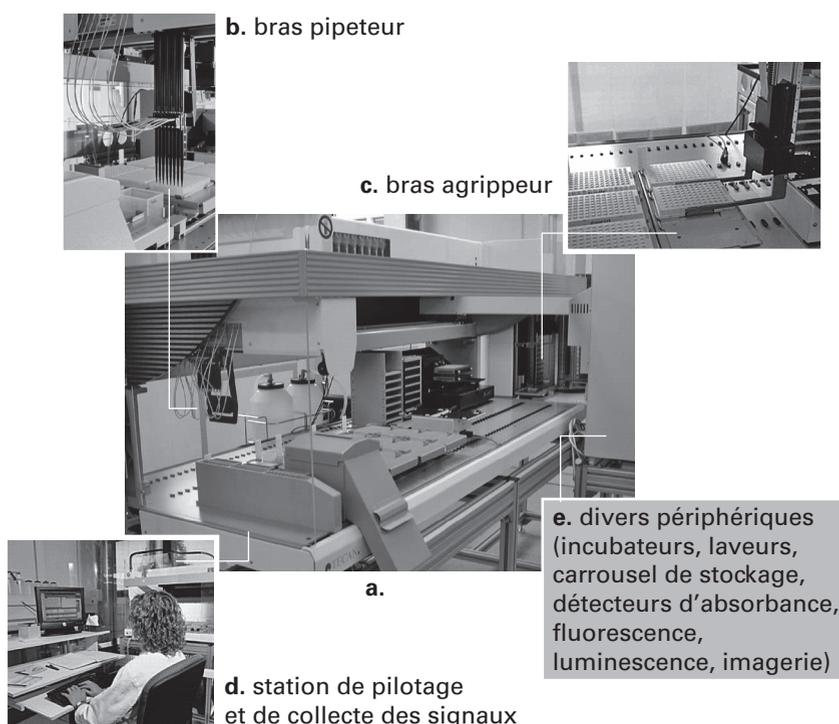


Fig. 1.3 - Un exemple d'automate de criblage

(Centre de Criblage pour des Molécules Bio-Actives, Grenoble)

L'automate assure une séquence de traitements auxquels les microplaques sont soumises (pipetage, mélanges, incubations, lavages...) et de mesures de signaux différents selon les tests effectués (absorbance, fluorescence, imagerie...). Chaque microplaque a donc une "vie" sur l'automate. Un ensemble de logiciels de pilotage est mis en oeuvre pour optimiser le traitement de plusieurs microplaques en parallèle (planification, en anglais, *scheduling*).

1.2.2. CONSOMMABLES, RÉPLIQUES DE CHIMIOTHÈQUES ET STOCKAGES

Pour chaque criblage, un stock complet de réactifs est consommé. En particulier, une série de microplaques correspondant à une réplique de la collection de molécules criblée est utilisée (fig. 1.4).

1.2.3. CONCEPTION DU TEST, CRIBLAGE PRIMAIRE, HIT-PICKING, CRIBLAGE SECONDAIRE

Le criblage est réalisé en plusieurs phases. Avant toute chose, une **cible** est définie pour un projet scientifique motivé par une recherche appliquée ou fondamentale. Nous verrons ci-dessous que la définition de la cible est une question difficile. Un **test** est mis au point afin de permettre d'identifier les perturbations intéressantes (inhibition ou activation) causées par les petites molécules et collectivement qualifiées de **bio-activité**.



Fig. 1.4 - Préparation d'une chimiothèque pour le criblage

Une chimiothèque est répliquée en lots à usage unique. Une réplique (un lot), conservée au froid, est utilisée par criblage.

Souvent, différents types de tests peuvent être envisagés pour rechercher des molécules actives sur une même cible. A cette étape, une réflexion approfondie est indispensable. Cette réflexion doit notamment prendre en compte les caractéristiques des chimiothèques utilisées et tenter de **prévoir** autant que possible dans quelles circonstances la mise en œuvre du test pourra aboutir à des résultats erronés ("faux positifs" et "faux négatifs", par exemple). La pertinence biologique du test est déterminante pour la pertinence des molécules sélectionnées à l'issue du criblage, au plan de leur intérêt ultérieur comme candidat médicament ou comme outil de recherche (chap. 3).

Un **criblage primaire** de l'ensemble de la chimiothèque est réalisé afin de sélectionner par seuillage une première série de composés candidats. Un **criblage secondaire** sur les molécules candidates permet une validation ou une infirmation avant de poursuivre l'étude. Pour effectuer ce criblage secondaire, les molécules sélectionnées sont regroupées sur de nouvelles microplaques. Le prélèvement de ces molécules dites **touches** (en anglais, **hit**) est réalisé à l'aide du robot. Cette étape est appelée **hit-picking**.

1.3. LA PETITE MOLÉCULE : APERÇU SUR LES DIFFÉRENTS TYPES DE CHIMIOTHÈQUES

1.3.1. LA PETITE MOLÉCULE

La **petite molécule** est un terme souvent employé pour qualifier les composés des chimiothèques. Il résume une des propriétés recherchées, à savoir une masse moléculaire (corrélée évidemment avec la taille) inférieure à 500 daltons. La petite molécule active est recherchée dans des collections de composés purs ou en mélange, issus de substances naturelles ou de synthèse chimique.

1.3.2. LE DMSO, SOLVANT POUR LES CHIMIOTHÈQUES

Le diméthylsulfoxyde (DMSO, fig. 1.5) est un solvant couramment utilisé pour les composés d'une chimiothèque issue de la chimie de synthèse. Le DMSO améliore la solubilité des composés hydrophobes ; il est miscible dans l'eau.

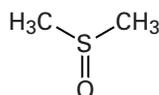


Fig. 1.5 - Le diméthylsulfoxyde (DMSO), solvant pour les chimiothèques

Une des propriétés du DMSO est aussi de déstabiliser les membranes biologiques et de les rendre poreuses, permettant un accès aux zones profondes des cellules et pouvant provoquer selon la dose des phénomènes de toxicité. Bien que le DMSO soit réputé inerte sur la majorité des cibles biologiques, il est important d'en contrôler l'effet avant tout criblage. Dans le cas où le DMSO est nocif pour la cible, il faut rechercher quelle est la concentration de DMSO sans effet sur la cible et adapter en conséquence la dilution des molécules de la chimiothèque. Parfois, la recherche d'un solvant mieux adapté doit être engagée.

1.3.3. LES COLLECTIONS DE SUBSTANCES NATURELLES

Les substances naturelles sont connues pour leur diversité (HENKEL *et al.*, 1999) et leur complexité structurale (TAN *et al.*, 1999 ; HARVEY, 2000 ; CLARDY et WALSH, 2004). Ainsi, 40 % des archétypes structuraux décrits dans les banques de données de produits naturels sont absents de la chimie synthétique. D'un point de vue historique, le succès des substances naturelles comme source de médicaments et de molécules bio-actives est évident (NEWMAN *et al.*, 2000).

Les méthodes actuelles pour isoler un produit naturel bio-actif, connues sous le nom de purifications bio-guidées, sont des processus itératifs consistant à extraire les échantillons à l'aide de solvants et à tester leur activité biologique. Le cycle de purification et de test est répété jusqu'à obtention d'un composé actif pur. Bien que permettant d'identifier des composés originaux, issus de la biodiversité, ce type d'approche présente un certain nombre de limites (LOCKEY, 2003). Tout d'abord,

1.6. UNE NOUVELLE DISCIPLINE AU PIVOT DE LA BIOLOGIE, LA CHIMIE ET L'INFORMATIQUE : LA CHEMOGÉNOMIQUE

Les développements technologiques coûteux (robotique, miniaturisation, standardisation, parallélisation, détection...) qui ont conduit à la création des plates-formes de criblage, étaient initialement motivés par la découverte de nouveaux médicaments. Comme pour toute innovation, il y eut des enthousiastes et des sceptiques. Faire un **bilan de l'apport du criblage automatisé pour la découverte de nouveaux médicaments** est difficile, en particulier du fait de la longueur des cycles de découverte (en moyenne 7 ans) et de développement (en moyenne 8 ans) de nouvelles molécules (MACARRON, 2006). On recensait 62 molécules candidates issues du HTS en 2002, 74 en 2003 et 104 en 2005, chiffres très largement sous-estimés du fait que ces enquêtes ne recouvraient pas tous les laboratoires, et en raison des contraintes de confiance. Le taux de succès est de plus variable en fonction de cibles, avec un succès régulier sur certaines cibles et un échec systématique sur d'autres que l'on pourrait croire **non "médicables"** (en anglais, *not druggable*). MACARRON (2006) rapporte à ce sujet que l'échec peut aussi être dû à la qualité de la chimiothèque. Ainsi, lorsque les sociétés GlaxoWellcome et SmithKline Beecham ont fusionné, certains criblages qui avaient échoué avec la chimiothèque de l'une des sociétés initiales, réussirent avec la chimiothèque de l'autre. La question de la "bonne chimiothèque" se pose alors comme une question critique, un champ d'étude ouvert.

Pour clore ce chapitre, nous posons la question de la place du criblage automatisé dans les disciplines scientifiques. Est-ce que cette technologie n'est qu'un progrès scientifique permettant de miniaturiser des essais et d'augmenter la cadence des pipetages manuels ? Suivant les arguments de Stuart SCHREIBER et Tim MITCHISON de la Harvard Medical School, Massachussetts, une nouvelle discipline à la charnière de la Biologie génomique et post-génomique et de la Chimie est née grâce à l'avance technologique apportée par le développement du criblage automatisé en milieu académique : la **chemogénomique** (en anglais, *chemogenomics*). Cette discipline émergente combine les derniers développements en génomique et en chimie et les applique à la découverte tant de molécules bio-actives que de cibles (BREDEL et JACOBY, 2004). Plus largement, la chemogénomique se donne pour objet l'étude des relations entre l'espace des cibles biologiques (appelé **espace biologique**, en anglais, *biological space*) et l'espace des petites molécules (**espace chimique**, en anglais, *chemical space*). Cet objectif ambitieux nécessite que les données des espaces biologiques et chimiques soient structurés au mieux dans des **bases de connaissances** (en anglais, *knowledge bases*) afin d'être explorés efficacement selon les techniques de **fouilles de données** (en anglais, *data mining*). La fig. 1.9 montre, sur un exemple de stratégie de génétique chimique inverse, la place de la chemogénomique au pivot des disciplines de la biologie, de la chimie et de l'informatique.

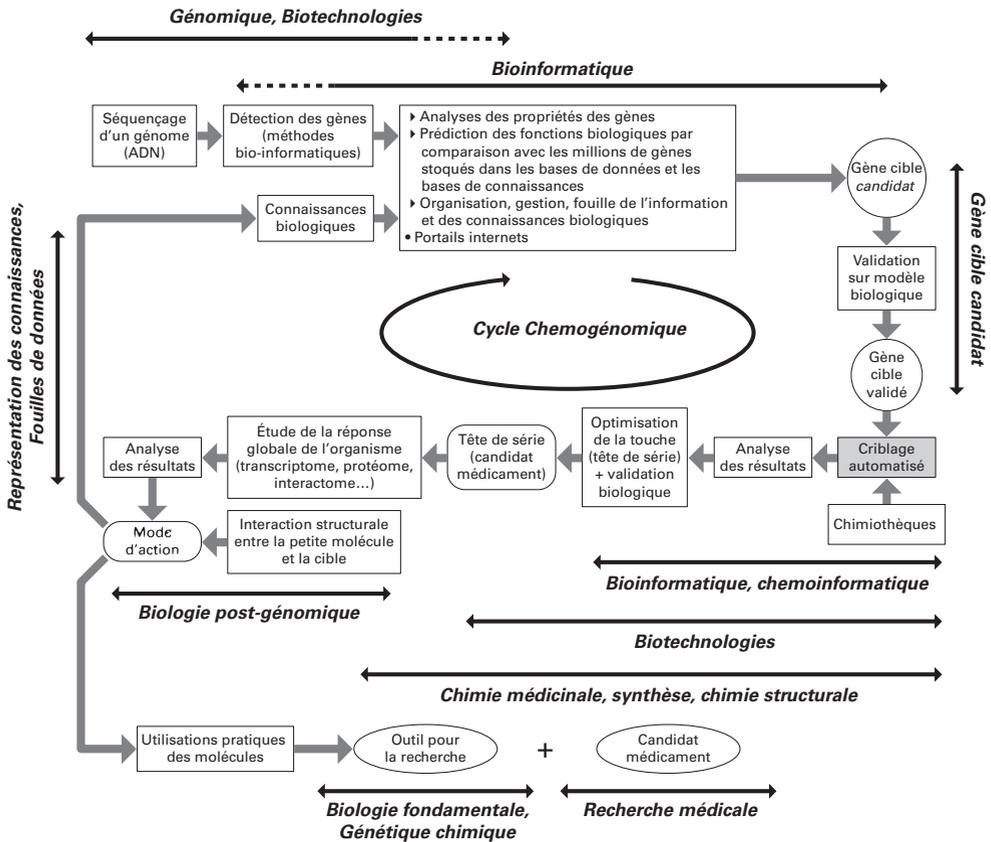


Fig. 1.9 - La chimogénomique, à la charnière de la biologie génomique et post-génomique, de la chimie et de l'informatique

La chimogénomique vise à comprendre la relation entre l'espace biologique des cibles et l'espace chimique des molécules bioactives. Cette discipline est rendue possible par la constitution de collections de molécules, l'accès aux technologies de criblage automatisé et une recherche importante en bioinformatique et chemoinformatique.

Cet ouvrage n'aborde pas la chimogénomique comme une discipline mûre, mais comme une discipline naissante. Nous éclairons surtout ce que biologistes, chimistes et informaticiens peuvent aujourd'hui apporter, et trouver, lorsque leur curiosité les mène à s'interroger sur cette rencontre entre le monde vivant et la diversité chimique.