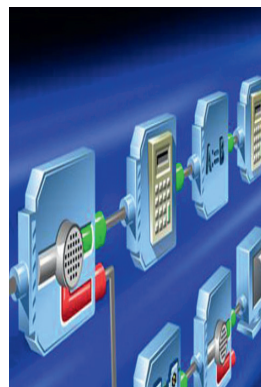


➤ À l'aube des années 2000, la séquence complète du génome humain a été déchiffrée. Cet événement historique a immédiatement lancé d'énormes défis aux chercheurs, tant sur le plan des concepts que sur celui des méthodes de travail. Chaque communauté a considéré la manière dont elle pouvait apporter une aide. Il est rapidement apparu que l'utilisation des petites molécules était un moyen de découvrir et de caractériser la fonction d'une protéine et, par inférence, l'importance du gène qui l'encode, dans la construction des cellules, des organes et des organismes. À l'initiative d'une poignée de chercheurs, les chimistes français se sont organisés pour collecter leurs molécules et les mettre à disposition de biologistes. Ils faisaient ainsi d'une pierre deux coups : ils donnaient une nouvelle vie à leurs molécules en créant la première chimiothèque académique d'une part, et ils permettaient pour la première fois à des biologistes du domaine public de se lancer dans des programmes de découverte de molécules biologiquement actives d'autre part. Il fallait toutefois réunir des composés en grand nombre et assurer un contrôle qualité uniforme, ce qui est rapidement devenu une priorité pour que les chimiothèques deviennent des outils fiables. ➤

Les chimiothèques sont des collections de molécules conçues pour la découverte de composés biologiquement actifs. Elles s'utilisent dans des opérations de sélection appelées criblage, dont le but consiste à identifier des molécules qui perturbent les systèmes biologiques afin d'en comprendre le fonctionnement. Elles permettent aussi parfois des actions correctrices des fonctions biologiques, à l'image des médicaments. On exploite ainsi des chimiothèques dans l'industrie du médicament et dans des laboratoires de recherche publique, et, bien

## Chémobiologie (12) Contrôle qualité des chimiothèques

Jean-Luc Galzi<sup>1</sup>, Fiorella Ruggiu<sup>2</sup>, Patrick Gizzi<sup>1</sup>,  
Bruno Didier<sup>3</sup>



<sup>1</sup> CNRS-Unistra UMR 7242, biotechnologies et signalisation cellulaire, 300, boulevard Sébastien Brant, CS 10413, 67412 Illkirch, France ;

<sup>2</sup> CNRS-Unistra UMR 7140, laboratoire de chémoinformatique, 67000 Strasbourg, France ;

<sup>3</sup> CNRS-Unistra UMR 7200, laboratoire d'innovation thérapeutique, 67412 Illkirch, France.  
[galzi@unistra.fr](mailto:galzi@unistra.fr)

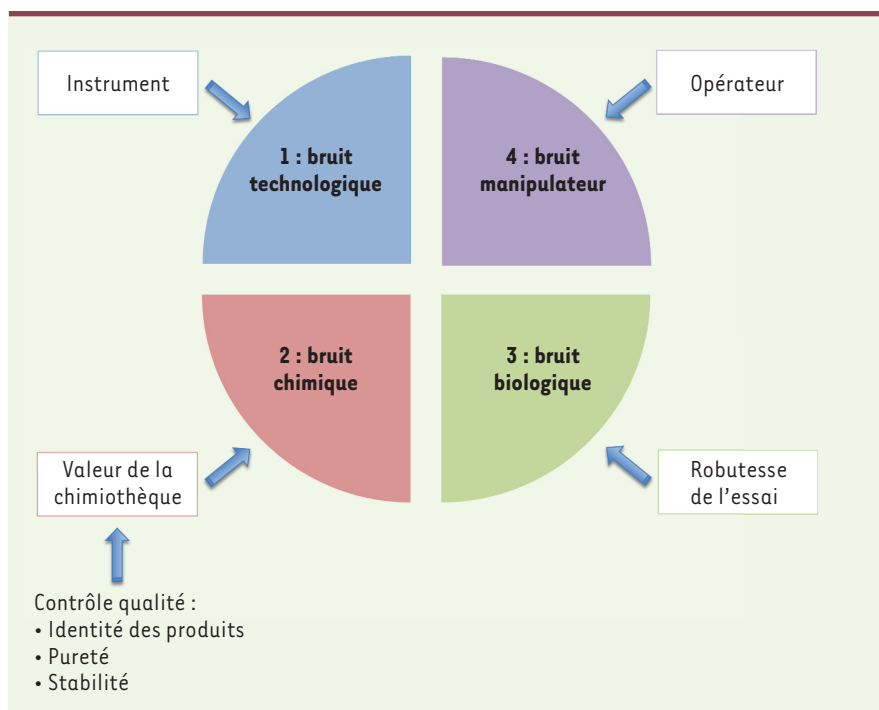
que ces deux secteurs poursuivent des buts différents, leurs démarches expérimentales comportent des points communs.

Chaque opération de criblage de collections de molécules conduit à l'identification de composés appelés touches, ou *hits* en anglais, qui sont des substances ayant exercé un effet biologique dans l'essai mis en œuvre. Les essais de criblage peuvent reposer sur des mesures d'interactions (liaison d'une molécule à une protéine nécessairement identifiée) ou sur des mesures d'une réponse biologique (stimulation ou blocage d'une activité cellulaire, ou animale, sur une cible généralement non identifiée). L'activité de la touche doit être confirmée, c'est-à-dire répétée, et caractérisée par l'établissement, par exemple, d'une relation dose-effet. Nous avons participé à plus de 40 campagnes de criblage qui ont porté sur un total de 250 000 molécules testées et plusieurs centaines de touches identifiées. Il apparaît que le taux de confirmation d'une touche, ainsi que le taux d'identification de la même touche par deux centres de criblage, peuvent varier de manière importante. Ces taux se situent en moyenne aux alentours de 30 % (S. Gioria, P. Villa, communication personnelle) [1]. Les raisons de ces différences sont analysées dans cet article ; elles proviennent de quatre sources dont la qualité des collections de molécules fait partie. Une collection de molécules contient entre mille et plusieurs millions de composés. La caractérisation des molécules qui la composent est un point important, mais il est évident que les coûts de gestion de grosses collections peuvent exploser. Entrons dans le cœur du problème.

### Pourquoi faire un contrôle qualité : les sources de bruit dans un criblage

L'identification d'une touche dans un essai de criblage est la résultante de plusieurs opérations indépendantes les unes des autres, mais dont la combinaison crée des conditions potentiellement très fluctuantes.

Cet article fait partie de la série « Chémobiologie » dont la parution a débuté dans le n° 12, vol. 30, décembre 2014 ([www.medecinesciences.org](http://www.medecinesciences.org)).



**Figure 1.** Les sources de bruit dans un criblage.

- En premier lieu figure le *bruit technologique* (Figure 1). Ce bruit est occasionné par la fiabilité et la sensibilité de l'instrumentation utilisée. Le criblage d'une collection de plusieurs milliers de molécules est une opération qui dépasse les capacités raisonnables de la manipulation manuelle. En effet, un point de mesure requiert la préparation d'un mélange reproductible contenant au minimum trois éléments (la cible, la molécule à tester et la molécule contrôle), utilisant des petits volumes afin d'économiser les réactifs, et réalisé à une cadence la plus élevée possible. Cela peut être atteint en ayant recours à des robots de pipetage capables de travailler rapidement et de manière très reproductible. Ces instruments peuvent alimenter des dispositifs de criblage (des plaques multipuits ou des lames de microscope) dont les formats s'étendent de 96 à plus de 6 000 points de mesure par plaque ou lame. Bien entendu, les technologies de miniaturisation qui permettent le haut débit sont plus coûteuses et nécessitent souvent un personnel formé au plus haut niveau de compétences ingénierales. Il faut en outre toujours maîtriser les domaines d'exploitation des instruments et des accessoires qui y sont ajoutés, connaître leurs failles et s'en écarter afin d'éviter de reproduire à très grande échelle des erreurs.
- En deuxième lieu, il convient de mentionner le *bruit chimique*. Ce bruit a au moins trois origines : l'identité des produits de la chimiothèque (le produit dans le puits est-il bien celui inscrit dans la base de données associée ?), la pureté des produits (combien de molécules distinctes sont présentes dans un seul échantillon ?) et la stabilité ou réactivité (la molécule peut-elle produire des réactions spontanées ou provoquées par l'opérateur lors des manipulations ?). Nous reviendrons plus loin sur ces points.
- En troisième lieu, le *bruit biologique* joue un rôle prépondérant surtout dans des essais cellulaires. Le biologiste souhaitant identifier un

composé chimique actif sur une cible identifiée (par exemple une enzyme ou un récepteur) doit produire cette cible de façon à respecter le cahier des charges de pureté, d'homogénéité et de fonctionnalité de la protéine. Ces critères peuvent changer d'un essai à l'autre, selon le type de détection mis en œuvre et sa sensibilité à la présence d'espèces contaminantes. Ils doivent en revanche être strictement respectés une fois le protocole établi. Le biologiste doit en outre être capable de produire des quantités généralement importantes de la cible, et doit pouvoir la conserver fonctionnelle assez longtemps, du moins durant toute la campagne de criblage. Parfois, la cible n'est pas identifiée. Le biologiste se repose alors sur un phénotype cellulaire dont la modification par un composé bioactif devient observable. Ce cas

est celui qui génère habituellement le plus de bruit car il faut tenir compte de paramètres complexes difficiles à maîtriser, tels que le caractère proliférant ou quiescent des cellules, le taux de confluence au moment des mesures, leur caractère synchrone, l'homogénéité de la lignée cellulaire, sa fragilité aux manipulations, etc.

- Vient enfin le dernier type de bruit qui est le *bruit du manipulateur*. Chaque essai de criblage requiert une mise au point plus ou moins longue dont l'objectif est la robustesse de l'essai de criblage : sa sensibilité, sa reproductibilité et sa stabilité sont des points cruciaux qui nécessitent souvent d'identifier les paramètres cachés de la mesure et de parvenir à les contrôler. À cet égard, il est critique pour les opérateurs d'avoir suivi des formations, non seulement dans le domaine technique du criblage proprement dit, mais aussi dans celui de l'assurance qualité, qui propose des méthodes de travail pour cadrer des procédures et pour en améliorer l'efficacité de manière continue. Chacun comprend l'importance de maîtriser chaque étape de ce processus et de prêter une attention d'autant plus grande à ces sources de bruit que la taille de la chimiothèque et celle du projet de criblage sont importantes.

### Notion de qualité d'une chimiothèque

Comme nous l'avons brièvement évoqué ci-dessus, le succès d'un criblage dépend en partie de la qualité de

la collection de molécules de la chimiothèque. Cette notion de qualité prend plusieurs significations : en premier lieu, le potentiel d'activité biologique des molécules incluses dans la collection et leur potentiel de variations structurales (possibilité de produire rapidement et facilement des analogues) sont considérés. La diversité des châssis moléculaires est également importante afin de couvrir l'espace chimique et de maximiser les chances de trouver des touches. Ensuite, les molécules « touche-tout » [2], qui sont des vraies touches, mais dénuées de sélectivité d'action [3], sont écartées dans la mesure du possible. La seconde manière de considérer la qualité d'une chimiothèque s'attache à des aspects plus quantitatifs tels que l'identité des produits, leur pureté et leur stabilité. Le taux des produits incorrects dans une chimiothèque non contrôlée est évalué à environ 30-50 %, ce qui est considérable. Le contrôle qualité des produits permet de réduire ce taux à 5-15 % avec un coût associé significatif [4]. Le prix moyen d'un composé chimique unique, pur et disponible à raison de plusieurs dizaines de milligrammes, est estimé communément à environ 500-1000 euros. Cela porte donc le prix d'une collection de 10 000 molécules à 5-10 millions d'euros. Une même collection fournie sans contrôle qualité perd 90 % de sa valeur, pour une raison simple : le travail à fournir après criblage pour curer les données, répéter les mesures et identifier les produits est un effort considérable que plus personne ne consent à faire aujourd'hui.

### Quelles techniques de contrôle qualité d'une chimiothèque ?

Deux grandes familles de méthodes sont utilisées pour contrôler la qualité d'une chimiothèque. Il s'agit des méthodes d'identification des molécules et des mesures de pureté.

- L'identification des molécules est réalisée en fonction des quantités disponibles, par spectrométrie de résonance magnétique nucléaire (échelle du milligramme) ou par spectrométrie de masse (échelle du microgramme). La spectrométrie de masse peut être considérée comme un simple outil de contrôle (confirmation de masses calculées avec une précision à 1 décimale - basse résolution) autant que comme un outil d'analyse structurale (masse exacte à 4 décimales - haute résolution). Elle ne permet pas en revanche l'analyse des phénomènes d'isomérisation ni celle, éventuellement, de conformations particulières, ce que permet la résonance magnétique nucléaire [5]. L'analyse des données expérimentales peut, elle aussi, être consommatrice de temps, tout particulièrement pour les mélanges de composés.

- La pureté des molécules peut être évaluée par chromatographie liquide et détection dans l'ultra-violet/visible. Cette méthode est la plus rapide et la moins onéreuse. Elle est adaptée aux grands nombres de molécules. En choisissant une longueur d'onde de détection basse (210-220 nm), les molécules portant une ou plusieurs insaturations (doubles liaisons) seront détectées. Cette méthode se limite au contrôle/confirmation de pureté car, dès lors que plusieurs produits sont détectés, il est nécessaire soit de disposer des substances de référence, soit d'associer des méthodes d'analyse en ligne. Afin de répondre à cette attente, la technologie de chromatographie liquide couplée à la spectrométrie de masse a été développée. Cette tech-

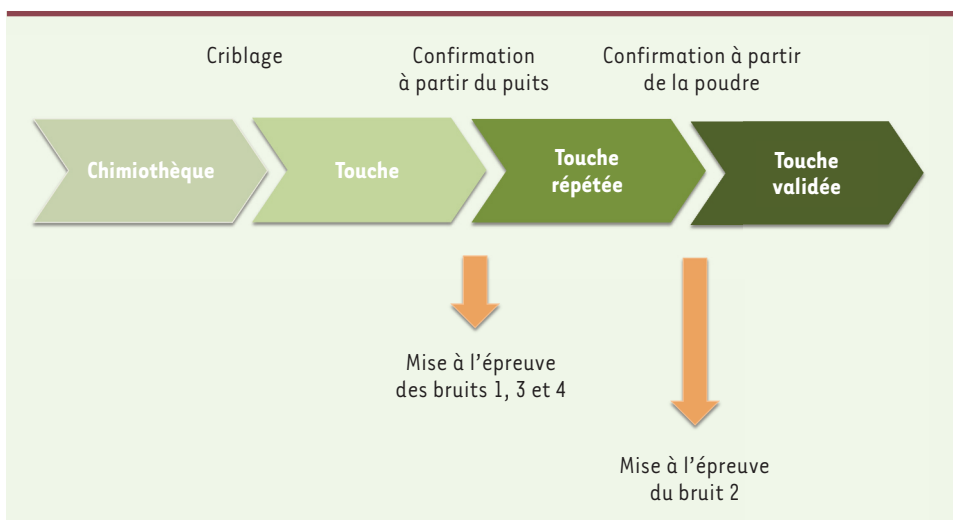
nique permet d'obtenir le pic de masse de la molécule (ou l'ensemble des pics qui constituent sa signature) et, ainsi, d'attribuer les produits aux pics du chromatogramme correspondants. En utilisant des spectromètres en tandem, l'ion relatif à la molécule recherchée peut être sélectionné dans un premier analyseur, puis fragmenté. Les fragments sont ensuite caractérisés dans un second analyseur, ce qui permet une identification plus certaine. Évidemment, cette mesure ne constitue pas une analyse de pureté, mais seulement une confirmation de la présence d'un produit.

### Quand faut-il faire le contrôle qualité d'une chimiothèque ?

La construction d'une chimiothèque comprend plusieurs étapes : (1) entrée de la structure dans une base de données, (2) pesée d'un produit, (3) constitution d'un stock de poudre en format tubes, enchâssés dans des plaques au format standard des chimiothèques (soit 96 tubes par plaque), (4) dissolution dans une plaque multipuits appelée plaque mère, d'une partie de la poudre dans un solvant, généralement du diméthylsulfoxyde (DMSO)<sup>1</sup>, pour atteindre une concentration de 10 mg/ml, (5) saisie des coordonnées cartésiennes de la molécule dans la base de données, (6) préparation de plaques filles (diluées pour les campagnes de criblage) et (7) stockage des plaques.

L'expérience montre que malgré les efforts fournis pour sélectionner des molécules caractérisées lors de la constitution d'une chimiothèque, des écarts entre le contenu réel d'un puits de plaque et le fichier qui décrit le produit existent toujours. Ce niveau d'écart varie de manière aléatoire et peut, pour certains sous-ensembles de molécules, atteindre des valeurs importantes, ce qui est heureusement fort rare. La question d'un contrôle qualité *a priori* se pose donc pour les chimiothécaires. Cette question se heurte cependant à celle, corrélée, du coût de l'opération. Un contrôle de bas niveau coûte de l'ordre de quelques euros par molécule. Il s'agit d'une analyse par spectrométrie de masse dans laquelle la question posée est : le produit désiré est-il présent dans le puits ? Une fraction importante des produits analysés de cette manière sont « invisibles », soit parce qu'ils ont précipité, soit parce qu'ils ne s'ionisent pas dans le spectromètre. La réponse est de type « oui-non » avec aucune information utilisable pour les puits « non ». Cette analyse est généralement peu satisfaisante, mais elle permet d'avoir

<sup>1</sup> Le diméthylsulfoxyde (DMSO) est un solvant polaire organique utilisé pour sa capacité à solubiliser de nombreux composés organiques.



**Figure 2. Démarche générale du contrôle qualité.**

une idée de la qualité des fournisseurs choisis. Un second niveau, cette fois plus poussé, exploite une analyse de la totalité de l'éluant de chromatographie. Le prix de l'analyse augmente significativement (20 euros par analyse) et le temps à consacrer au dépouillement des résultats devient considérable. Ainsi, une analyse complète d'une chimiothèque de 10 000 composés dépasse 200 000 euros, auxquels doivent s'ajouter plusieurs semaines d'analyse. L'achat d'une chimiothèque commerciale ne conduit pas à un produit figé. Le fournisseur ne pourra pas refournir la totalité des produits après un an. Entre-temps, les chimistes auront élaboré des analogues de molécules bioactives découvertes dans la collection. Ainsi, il est justifié de considérer qu'une chimiothèque est un produit vivant évoluant dans le temps, contrairement à une bibliothèque avec laquelle elle est communément comparée. Un produit chimique testé dans un essai biologique ne revient jamais dans la collection : ceci explique l'amenuisement des quantités disponibles et, ultimement, l'épuisement du stock.

La stratégie du contrôle qualité est à présent clairement posée : doit-on faire un contrôle exhaustif *a priori*, sachant que cela aura une conséquence sur le prix d'un criblage et que cet effort n'aura été réellement utile que pour les quelques molécules qui seront des touches ? C'est en tentant de répondre à cette question que les approches du contrôle qualité se sont mises en place chez les utilisateurs et les chimiothécaires. La démarche générale est décrite dans la *Figure 2*. Elle permet de prendre en compte toutes les sources de bruit et réduit l'effort d'analyse pour l'adapter aux réels besoins. Le processus commence avec l'utilisation de la chimiothèque pour réaliser un criblage et identifier des touches. Cette opération est réalisée à partir de plaques multipuits dans lesquelles les produits sont dissous parfois depuis longtemps, et ont pu subir des cycles répétés de congélation-décongélation. L'expérimentateur est pragmatique : il considère *a priori* que la molécule est bien celle qui figure dans le fichier. Il confirme donc la réalité de l'effet en répétant l'essai biologique à l'aide du produit issu du même lot (même plaque, même puits). Il met ainsi à l'épreuve les sources de bruit 1, 3 et 4 mentionnées plus haut, c'est-à-dire qu'il confirme qu'il n'y a pas eu de problème de pipetage et de lecture du signal, que l'essai biologique a bien fonctionné et qu'il n'a pas

d'aller plus loin dans le développement d'un produit, il demande une vérification de l'identité et de la pureté de la molécule. Le résultat de l'ensemble de ce processus est appelé « validation de la touche ».

La question de la stabilité des produits chimiques stockés dans un solvant peut être un point important. L'hydratation du solvant lors des cycles de congélation et décongélation peut faciliter des réactions d'hydrolyse, et l'apport en oxygène peut, quant à lui, favoriser des oxydations [6, 7]. Ces notions sont normalement prises en compte lors de la constitution d'une chimiothèque [8] (→) mais il arrive que de telles molécules instables échappent à l'attention des chimistes et soient présentes dans les grandes collections. À nouveau, une méthode unique pour traiter ce point n'existe pas. Certains chimiothécaires analysent périodiquement une fraction de leur chimiothèque, d'autres limitent leurs efforts à la caractérisation approfondie des touches.

Il arrive malheureusement que la molécule active, la touche, soit un contaminant. Les chercheurs ne publient pas ce type d'information, mais il est communément admis que lorsque l'effet observé avec la solution issue de la plaque de criblage n'est pas reproduit à l'aide de la molécule en poudre, la touche est abandonnée. Il n'est pas rare que l'identification de la molécule active prenne un temps considérable. Le contaminant n'est pas nécessairement un sous-produit d'une réaction qui pourrait être caractérisé, mais plus fréquemment, il s'agit d'une impureté présente dans un des lots de réactifs utilisés. Le traitement d'extraits naturels est soumis lui aussi à des aléas de même nature. C'est pourquoi la qualité du fichier décrivant leurs préparations est critique au succès et à la répétabilité de la qualité des lots. Il arrive par ailleurs que la séparation

(→) Voir la Synthèse de F. Mahuteau-Betzer, *m/s* n° 4, avril 2015, page 417

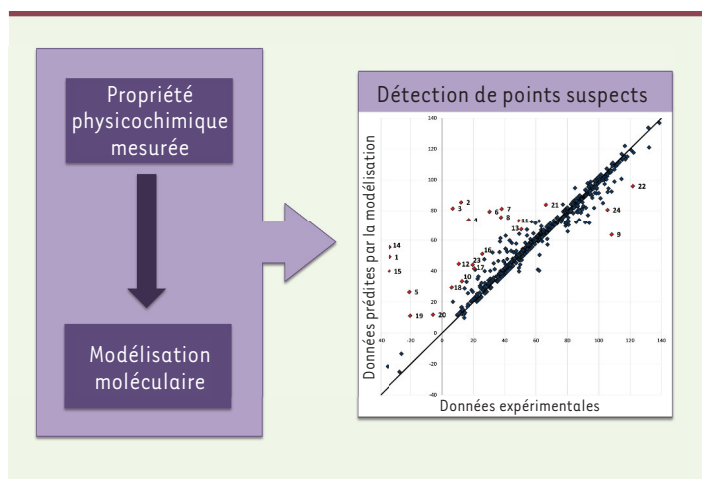


Figure 3. Apport de la chémo-informatique dans le contrôle qualité.

de composants présents dans le mélange entraîne la perte d'activité biologique [9].

Dans certains cas plus heureux, le contaminant a pu être identifié. Ceci s'est produit, entre autres, dans une campagne de criblage pour identifier des inhibiteurs du récepteur des chimiokines CXCR4 (*C-X-C chemokine receptor type 4*) qui sert de point d'entrée au virus VIH (virus de l'immunodéficience humaine) dans les cellules hôtes [10]. Le criblage d'une collection de molécules cycliques, les cyclames<sup>2</sup>, a conduit à une touche dont le produit, après resynthèse ne présentait plus d'activité biologique. L'examen attentif du contenu du puits dans la plaque a révélé une impureté dont la nature était un dimère covalent de cyclame, l'AMD 3100. La structure de l'AMD 3100 a été confirmée ; il a été resynthétisé et son effet biologique répété et validé<sup>3</sup>. Cette molécule a atteint le stade clinique puis la mise sur le marché a été autorisée [11].

### Annotation + modélisation = contrôle qualité

L'annotation d'une collection de molécules consiste à mesurer leurs propriétés et à les associer à leur structure. La modélisation conduit au même exercice mais de façon théorique, avec un système informatique capable d'associer entre eux des éléments de structure et des informations physiques (solubilité, ionisation, lipophilicité, etc.) ou biologiques (activateur versus inhibiteur, liaison à la protéine X et à la protéine Y mais pas à la protéine Z, etc.). Ce système utilisera cet « apprentissage » pour prédire des propriétés physiques (ou biologiques) à partir d'éléments structuraux. Quand le système informatique conduit à des prédictions fiables, la comparaison des données expérimentales aux valeurs prédites constitue un contrôle de la qualité de l'échantillon. Dans le cas contraire, les tentatives de modélisation nous démontrent que certains paramètres régissant une

association structure-fonction demeurent incompris. Prenons des exemples.

La méthode de modélisation la plus utilisée est la relation quantitative structure à activité. Elle consiste à calculer des descripteurs moléculaires à partir de la structure, puis à les analyser par des méthodes d'apprentissage par machine. Essentiellement, les méthodes d'apprentissage identifient les descripteurs faisant varier la propriété ou l'activité, et quantifient leurs effets. Le modèle résultant permet la prédiction de molécules non testées. Il est à noter que seules les prédictions de molécules similaires à celles utilisées pour construire le modèle sont fiables. Il faut donc sélectionner soigneusement son sous-ensemble de départ et définir la partie de l'espace chimique dans lequel le modèle sera fiable, c'est-à-dire le domaine d'applicabilité.

À ce jour, l'ionisation d'une molécule « *drug-like* »<sup>4</sup> est considérée comme pouvant être modélisée de manière fiable [12] et prédite d'une manière utile pour les chimistes qui souhaitent la contrôler. Cette propriété est importante, car elle régit la solubilité des molécules ainsi que leur capacité à franchir les barrières biologiques.

La solubilité est précisément une des propriétés physicochimiques que tout biologiste souhaiterait maîtriser : elle définit la gamme des concentrations dans laquelle on peut tester l'effet d'un composé. Elle permet de comprendre en partie l'affinité que la molécule présente pour sa cible, ainsi que sa biodisponibilité. C'est malheureusement une propriété qui demeure difficile à prédire, vraisemblablement en partie à cause de propriétés amphiphiles conduisant à des interactions qui ne sont pas prises en compte dans la modélisation. Ce comportement induit la formation d'agrégats solubles [13] dont les caractéristiques sont celles d'une macromolécule et non celle de la molécule isolée. Il arrive que ces agrégats soient des touches dans les criblages, ce qui les rend intéressants. Leur propriété caractéristique est une disparition brutale de l'effet biologique lors de la dilution de l'échantillon, ce qui prévient toute mesure fiable de relation dose-effet.

Il existe une troisième propriété des molécules qui est prédictive de leur comportement *in vivo*, en particulier pour le franchissement des barrières biologiques. Il s'agit du coefficient de partage entre deux phases non miscibles (généralement le 1-octanol et l'eau), qui donne lieu à un paramètre, log P, ou sa variante log D si l'on remplace l'eau par un tampon à pH physiologique. Ce paramètre décrit assez fidèlement la capacité

<sup>2</sup> Les cyclames sont des molécules de formule [14]aneN4.

<sup>3</sup> L'AMD 3100, constitue l'antagoniste le plus performant du récepteur CXCR4, récepteur naturel de la chimiokine CXCL12. Il est surtout utilisé dans certaines indications de mobilisation des cellules souches hématopoïétiques dans le sang périphérique (en vue d'un prélèvement pour transplantation autologue).

<sup>4</sup> Une molécule « *drug-like* » présente les caractéristiques requises pour en faire un médicament potentiel.

d'une molécule à franchir la barrière intestinale, ainsi que la barrière hémato-encéphalique. Il est très utile pour décrire la biodisponibilité d'une molécule *in vivo*, ainsi que pour déterminer la meilleure voie d'administration [14]. Cette variable peut être modélisée de façon fiable avec les méthodes chémoinformatiques.

L'utilisation de la modélisation ne se limite pas à la prédiction de molécules non testées, elle peut aussi être utile dans le cadre du contrôle qualité. La confrontation des données expérimentales avec les prédictions peut révéler des problèmes expérimentaux quand les prédictions ne correspondent pas aux valeurs obtenues (Figure 3). Une étude a été menée sur l'indice chromatographique d'hydrophobicité, correspondant au coefficient de partage entre une colonne chromatographique en phase inverse et un gradient d'acétonitrile et d'eau [4]. La modélisation de quelques centaines de molécules a révélé des erreurs significatives et récurrentes dans la prédiction de certaines, même en changeant la méthode d'apprentissage. Les molécules concernées ont été testées à nouveau et l'examen a montré que (1) certaines molécules avaient évolué dans les puits de plaque de la chimiothèque (essentiellement par hydrolyse de fonctions sensibles, telles que des esters), (2) le produit présent dans le puits était inconnu et ne correspondait pas à la structure indiquée dans la base de donnée. Ce fait n'est pas rare. Ainsi, les chercheurs qui prennent le soin de faire contrôler les produits commerciaux qu'ils se procurent, constatent que près de 10 % des lots qui leur sont fournis ne correspondent pas au produit commandé (D. Rognan, communication personnelle). Cela est particulièrement vrai pour les chimiothèques achetées par lots de plusieurs milliers, voire dizaines de milliers, de molécules.

Ainsi, la mesure du coefficient de partage, qui peut être réalisée de manière rapide et peu onéreuse, apparaît être non seulement un paramètre prédictif des propriétés des molécules, représentant un élément d'information très utile pour le biologiste et le chimiste, mais aussi un moyen efficace de réaliser un contrôle qualité d'une chimiothèque.

## Conclusion

L'avènement de l'automatisation, dans le domaine de la biologie avec les méthodes génomiques récentes, ou dans le domaine de la chimie avec les grandes collections de molécules et leur production en masse par la chimie parallèle et combinatoire, nous conduit à changer notre façon d'appréhender le travail de recherche et développement. Nous avons tendance à penser que le passage aux grands nombres est la simple répétition d'opérations unitaires et, ainsi, que le mode de pensée pour le traitement d'une molécule se transpose simplement à l'ensemble de la collection. Cette idée se heurte à une réalité qui se veut consommatrice de temps et d'argent. Une méthode d'analyse qui nécessite un examen au cas par cas des résultats s'avère rapidement trop chronophage. Une méthode dont le prix varie du simple au double a des conséquences énormes à l'échelle de la dizaine ou la centaine de milliers de molécules. Le criblage même d'une collection de molécules coûte encore aujourd'hui de l'ordre de l'euro par point de mesure. La recherche de stratégies permettant d'obtenir des résultats fiables alliés à un coût humain et financier raisonnable, est un enjeu majeur.

Il est vrai que la découverte de nouvelles molécules bioactives est d'autant plus fructueuse et innovante que l'on s'épargne d'effectuer des présélections qui, par définition, réduisent la taille du champ exploré et se font souvent au détriment de l'audace. Dans ce contexte, l'association d'approches expérimentales et de méthodes prédictives est assurément un atout. Pour que cela fonctionne, il est nécessaire d'obtenir des jeux de données expérimentales de grande qualité acquis dans des conditions parfaitement définies et robustes, de les confier aux chémoinformaticiens et d'explorer conjointement l'impact des prédictions sur la méthode de mesure et la valeur d'un point expérimental. Il est tout autant nécessaire de se poser des questions sur la nature et l'importance des descripteurs moléculaires utilisés pour réaliser la modélisation. ♦

## SUMMARY

### Quality control of chemical libraries

The complete sequence of the human genome has been deciphered at the dawn of the new century. This historic event immediately challenged researchers with new needs both in terms of concepts and of working methods. Each scientific community considered how it could tackle these new challenges and it quickly became clear that using small chemical molecules would help discovering and characterizing the function of new proteins. The importance of the genes that the encode new proteins could thus be established in cells, organs and whole organisms. At the initiative of a handful of researchers, French chemists have organized the collection of their molecules and provided them to biologists. By doing so they killed two birds with one stone: on the one hand they created a unique opportunity to add value to their molecules by creating the first academic chemical library, and on the other hand they stimulated the launch of biologically active molecules discovery programs by biologists from the academic sector. It was necessary, however, to raise many compounds and ensure consistent quality control, which quickly became a priority for the chemical libraries to become reliable tools. ♦

## LIENS D'INTÉRÊT

Les auteurs déclarent n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

## RÉFÉRENCES

1. Lindh M, Svensson F, Schaal W, et al. Toward a benchmarking data set able to evaluate ligand- and structure-based virtual screening using public HTS data. *J Chem Inf Model* 2015 ; 55 : 343-53.
2. Baell J, Walters M. Chemical con artists foil drug discovery. *Nature* 2014 ; 513 : 481-3.

## RÉFÉRENCES

3. Rognan D, Bonnet P. Les chimiothèques et le criblage virtuel. *Med Sci (Paris)* 2014 ; 30 : 1152-60.
4. Ruggiu F, Gizzi P, Galzi JL, Hibert M, et al. Quantitative structure-property relationship modeling: a valuable support in high-throughput screening quality control. *Anal Chem* 2014 ; 86 : 2510-20.
5. Klaus A. *On-line LC-NMR and related techniques*. Chichester : Wiley, 2002 : 290 p.
6. Semin D, Malone T, Paley M, Woods P. Novel approach to determine water content in DMSO for a compound collection. *J Biomol Screen* 2005 ; 10 : 568-72.
7. Engeloeh C, Schopfer U, Muckenschnabel I, et al. Stability of screening compounds in wet DMSO. *J Biomol Screen* 2008 ; 13 : 999-1006.
8. Mahuteau-Betzer F. Chimiothèque Nationale. *Med Sci (Paris)* 2015 ; 31 : 417-22.
9. Spelman K, Duke JA, Bogenschutz-Godwin MJ. The synergy principle at work with plants, pathogens, insects, herbivores and humans. In : *Natural products from plants*. Boca Raton : CRC Taylor and Francis, 2006 : 475-501.
10. Davies S, Serradell N, Bolos J, Bayes M. Plerixafor hydrochloride. *Drugs Future* 2007 ; 32 : 123-36.
11. CHMP assesment for Mozobil. *European Medicines Agency*. 2009 : procedure No. EMEA/H/C/001030.
12. Cruciani G, Milletti F, Storchi L, Sforza G, et al. *In silico* pKa prediction and ADMET profiling. *Chem Biodivers* 2009 ; 6 : 1812-21.
13. McGovern S, Caselli E, Grigorieff N, Schoichet B. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J Med Chem* 2002 ; 45 : 1712-22.
14. Van de Waterbeemd H, Lennernäs H, Artursson P. High throughput measurement of logD and pKa. In : *Methods and principles in medicinal chemistry*. Weinheim : Wiley, 2003 : 21-45.

**TIRÉS À PART**

J.L. Galzi



# Questions de santé publique

## Un nouveau bulletin

pour une meilleure visibilité des résultats de la recherche en santé publique

**L**es résultats de la recherche en santé publique souffrent en France d'un réel manque de visibilité. Ceci concerne aussi bien le monde académique (hors santé publique) que le grand public et les décideurs. Pour pallier ce déficit, l'IReSP a créé un bulletin à large diffusion intitulé « *Questions de santé publique* », largement inspiré du bulletin mensuel d'information de l'INED « *Populations et sociétés* ». L'objectif éditorial est de porter à la connaissance d'un large public (enseignants, étudiants, journalistes, décideurs, milieux de la recherche, asso-

ciations, public concerné) les informations les plus récentes concernant des questions importantes de santé publique, rédigées de façon facilement lisible et compréhensible pour des non spécialistes, en garantissant que les informations publiées sont validées scientifiquement. La publication concerne des faits et non des positions. Au-delà de la présentation de résultats, les qualités pédagogiques de *Questions de santé publique* permettent au lecteur de mieux comprendre comment sont formulées et abordées les questions de santé publique et quelles sont les limites de ces études.



Nom .....

Prénom .....

Institution ..... Fonction .....

Spécialité ..... Service .....

Adresse .....

Ville .....

Code postal .....

Pays .....

Adresse électronique .....

à nous retourner par la poste ou par fax au 01 49 85 03 45

Questions de santé publique  
Les Éditions EDK EDP Sciences  
109, avenue Aristide Briand  
92541 Montrouge Cedex  
France

Réservé aux abonnés de M/S  
Recevez gratuitement et régulièrement  
**Questions de santé publique**  
en renvoyant ce document soigneusement rempli.

**Questions de santé publique** est une publication de l'Institut de Recherche en Santé Publique. ■ **Directeur de la publication** : Geneviève Chêne.

■ **Rédacteur en chef** : Claire-Isabelle Coquin. ■ Une réalisation des Éditions EDK.