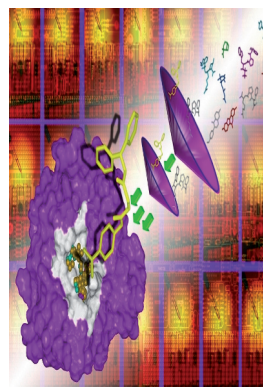


> L'avènement de la biologie moléculaire et l'achèvement du séquençage du génome humain ont conduit l'industrie pharmaceutique à progressivement implémenter des approches dites cible-centriques pour identifier les candidats médicaments. Cependant, la faible productivité de la recherche et du développement en ce début de millénaire, combinée aux évolutions technologiques dans des domaines tels que l'ingénierie cellulaire, le criblage à haut contenu, la robotique, l'analyse d'images et l'intelligence artificielle, ont nourri un fort regain d'intérêt pour les approches phénotypiques. De plus en plus fréquemment, les approches cible-centriques et phénotypiques sont considérées de façon complémentaire, positionnant ainsi les techniques de déconvolution¹ de cible sur le chemin critique de la découverte et du développement de médicaments. Cette revue analyse l'évolution des approches cible-centriques *versus* phénotypiques, en se focalisant plus particulièrement sur le criblage à haut contenu et les différentes techniques de déconvolution de cible aujourd'hui disponibles. <

Du criblage à haut contenu à la déconvolution de cibles

Nouvelle donne pour les approches phénotypiques

Preety Shabajee, Albane Gaudeau, Céline Legros, Thierry Dorval, Jean-Philippe Stéphan



Pôle d'expertise
Criblage pharmacologique,
chimiothèque et biobanques,
Institut de Recherches Servier,
125, Chemin de Ronde,
78290 Croissy-sur-Seine, France.
jean-philippe.stephan@servier.com

pas la phase de commercialisation et ceci essentiellement en raison d'un manque d'efficacité ou, dans une moindre mesure, d'un niveau de sécurité suffisant. Afin de remédier à cette attrition lors des phases cliniques, le secteur a mis en œuvre des stratégies visant à augmenter les chances de succès des candidats précliniques. Ainsi, des outils de biologie moléculaire, de génomique et de bio-informatique sont déployés pour optimiser la découverte de cibles pertinentes, et combinés avec des techniques permettant d'améliorer la découverte de candidats telles que la chimio-informatique, la chimie combinatoire ou encore le criblage à haut débit. Le processus de découverte de médicaments s'est ainsi progressivement éloigné de la sérendipité² du chimiste médicinal, se focalisant principalement sur l'effet biologique d'un composé ou d'une classe de composés pour évoluer vers des approches plus rationalisées et industrialisées dites « cible-centriques » qui permettent d'évaluer l'activité d'un grand nombre de composés directement sur la cible moléculaire d'intérêt. Malgré les énormes avancées scientifiques et techniques réalisées depuis l'avènement de la pharmacologie moderne, le coût considérable de la Recherche et du Développement conduit actuellement la communauté scientifique à s'interroger sur les différentes approches permettant de garantir la mise au point de traitements innovants, abordables et accessibles au plus grand nombre de façon soutenable dans le temps. Plusieurs analyses ont été menées durant les dix dernières années pour tenter de comprendre la genèse des médicaments innovants depuis

Complémentarité des approches cible-centriques et phénotypiques dans la découverte de médicaments innovants

Même si le coût moyen pour développer un nouveau médicament reste un sujet de controverse avec des évaluations allant de 1,3 à 2,8 milliards de dollars US (soit environ de 1,1 à 2,4 milliards d'euros), il est indéniable que ce coût est non-négligeable quelle que soit l'analyse [1]. En effet, l'industrie pharmaceutique est une activité à risques avec 90 % des molécules à usage thérapeutique testées chez l'homme qui n'atteignent

Vignette (Photo © Inserm-Villoutreix, Bruno).

¹ Identification du ou des composé(s) responsable(s) de l'activité du mélange.

² Découverte inattendue due au hasard.

1999, en comparant l'impact des approches dites phénotypiques et celui des approches dites cible-centriques. Une première publication de David C. Swinney et Jason Anthony faisait état, en 2011, parmi les 50 petites molécules approuvées par la FDA (*Food and drug administration*) entre 1999 et 2008 dans une nouvelle classe thérapeutique, de 28 molécules découvertes *via* des criblages phénotypiques [2]. Parmi celles-ci, 25 furent découvertes de façon intentionnelle, en s'intéressant spécifiquement à un phénotype défini, et 3 autres par sérendipité en utilisant des tests phénotypiques. Ce chiffre devançant largement le nombre de molécules découvertes par des cribles cible-centriques, les auteurs soulignaient l'intérêt des approches phénotypiques par rapport aux approches cible-centriques mises en avant par l'industrie pharmaceutique durant cette période. Cependant, une seconde étude publiée en 2014 par Jörg Eder, Richard Sedrani et Christian Wiesmann apportait une lecture quelque peu différente, en analysant les approbations de la FDA sur la période de 1999 à 2013 [3]. En effet, les cinq ans supplémentaires inclus dans cette seconde étude correspondent à un essor majeur des approches cible-centriques suite à l'achèvement du projet de séquençage du génome humain et à l'implémentation de nombreuses techniques de biologie moléculaire en routine dans les processus de Recherche et Développement. Sur les 78 petites molécules approuvées par la FDA dans une nouvelle classe thérapeutique durant cette période, 33 furent découvertes par des approches phénotypiques et 45 par des approches cible-centriques ; pour 9 des 33 molécules découvertes sur la base de tests phénotypiques, le mécanisme d'action est resté mal compris, voire, dans certains cas, complètement inconnu. Ce point souligne le fait que même si la connaissance du mécanisme d'action, incluant l'identité de la cible moléculaire, est un atout, cela ne constitue pas un impératif pour développer avec succès un nouveau médicament. Les auteurs concluaient en mettant en avant l'intérêt des cribles phénotypiques, notamment pour découvrir de nouveaux principes thérapeutiques ou de nouvelles voies moléculaires pour des maladies encore intraitables. Enfin, ils soulignaient le fait que les approches phénotypiques et cible-centriques ne doivent pas être perçues comme étant exclusives, mais plutôt comme complémentaires l'une de l'autre. Ainsi, un objectif majeur pour la recherche serait de rendre les cribles phénotypiques plus performants et efficaces, et de les combiner de façon raisonnée à des stratégies cible-centriques pour aboutir à une démarche globale plus productive. Cette vision rejoint celle décrite par John G. Moffat, Joachim Rudolph et David Bailey qui introduisent, en 2014, la notion d'approche phénotypique définie d'un point de vue mécanistique, et pour lesquels un crible phénotypique est utilisé soit pour optimiser des molécules dont la cible moléculaire est déjà connue, soit pour découvrir des molécules candidates qui seront ensuite caractérisées dans un test moléculaire précis [4]. L'implémentation de telles approches nécessite la maîtrise d'un ensemble de techniques qui incluent l'utilisation de modèles cellulaires susceptibles de reproduire de façon plus précise la physiopathologie d'une maladie, l'utilisation du criblage à haut contenu permettant d'appréhender la complexité des modèles cellulaires dans le contexte d'une démarche industrielle [5, 6] (→) et, enfin, les différentes techniques de

(→) Voir le Dossier technique de R. Prudent et al., *m/s* n° 10, octobre 2013, page 897

déconvolution susceptibles d'identifier l'entité et les mécanismes moléculaires dont la modulation provoque l'effet phénotypique souhaité. L'intégration de ces différentes technologies représente une nouvelle donne pour les approches phénotypiques avec pour objectif, l'établissement d'un lien de causalité entre modulation de l'activité ou de l'engagement d'une cible par une entité moléculaire donnée, de ses impacts au niveau cellulaire et de la réponse phénotypique du modèle, permettant ainsi de mieux anticiper l'activité clinique. La Figure 1 résume l'évolution des approches phénotypiques et celles fondées sur la cible dans la recherche de nouveaux médicaments.

Intégration de la complexité des modèles biologiques *via* le criblage à haut contenu : une réelle opportunité pour la recherche pharmaceutique

L'approche dite de criblage à haut débit (ou *high-throughput screening*, HTS) consiste à tester de façon systématique un grand nombre de molécules, le plus souvent chimiques, dans un test *in vitro* visant à évaluer l'engagement d'une ou plusieurs cibles dont l'activité est supposée être en lien avec une maladie définie. Cette démarche de criblage peut être cible-centrique ou phénotypique, mais dans tous les cas, l'objectif est d'identifier une ou plusieurs entités moléculaires pouvant conduire au développement d'un ou plusieurs candidats médicaments. Depuis la fin des années 1990, le criblage à haut débit s'est enrichi d'approches dites à haut contenu (ou *high content screening*, HCS). Le HCS vise à extraire à partir d'images, un large ensemble de données quantitatives, spatiales et potentiellement temporelles. Par définition, l'approche HCS est multiparamétrique. Elle vise à évaluer de manière précise un phénomène biologique spécifique [7] (→).

(→) Voir la Synthèse de P. Brodin et al., *m/s* n° 2, février 2015, page 187

La précision avec laquelle un phénotype peut être classifié est intrinsèquement liée au nombre de paramètres extraits et au niveau d'information qu'ils véhiculent. Cette spécificité du HCS permet, entre autres, de classer les phénotypes résultant de différents traitements, et d'identifier les faux-positifs lors d'une campagne de criblage [8]. Les données émanant d'un traitement sont généralement agrégées sur l'ensemble des cellules composant le modèle, dans un souci de simplicité, faisant alors l'hypothèse d'une réponse homogène sur l'ensemble de la population cellulaire. Cependant, la pertinence des modèles biologiques s'accompagnant généralement d'une hétérogénéité des valeurs prises par un même descripteur au sein de la population, il est

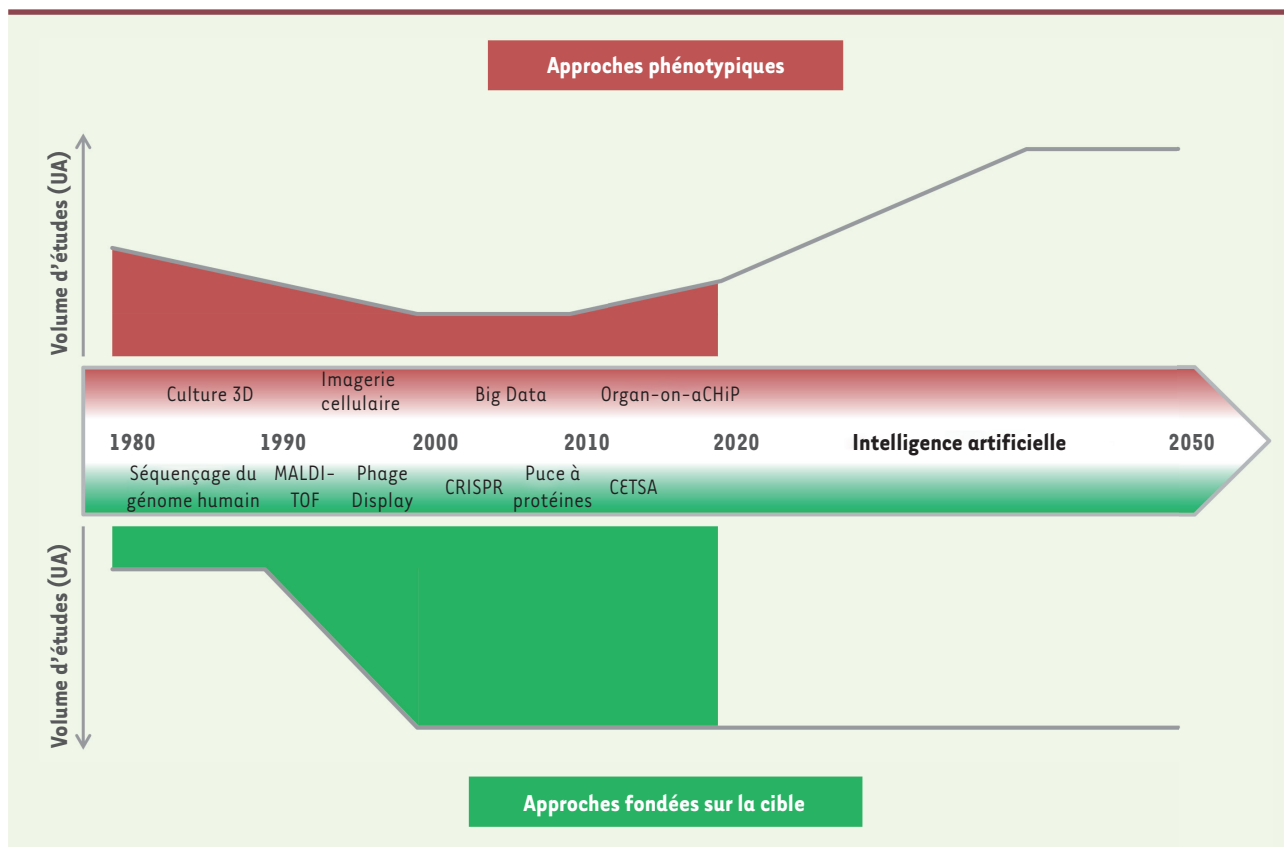


Figure 1. Passé et futur des approches phénotypiques et fondées sur la cible dans la recherche de nouveaux médicaments.

de plus en plus courant de tenir compte de l'ensemble de ces valeurs au niveau de la cellule unique : on parle alors d'analyse en cellule unique (ou *single cell*). Cette hétérogénéité, qui peut provenir de processus comme le cycle des cellules, la différenciation d'une sous-population, ou plus simplement, la proximité spatiale des objets entre eux, peut être à l'origine de variations de la réponse cellulaire. Par exemple, malgré une réponse globale similaire entre plusieurs lignées cancéreuses en termes d'activation de STAT3 par l'interleukin-6 (IL-6), les cellules de chacune de ces lignées ont une réponse différente, avec approximativement 10 % de la population cellulaire qui ne répondent pas au traitement (macro-hétérogénéité), le reste des cellules présentant différents degrés de réponse (micro-hétérogénéité) [9].

Pour illustrer l'efficacité et la qualité d'un criblage à haut contenu (HCS), une étude réalisée par O'Brien *et al.* avait comparé les résultats obtenus pour 611 composés en utilisant 7 tests classiques de cytotoxicité et un test HCS [10] : ce dernier s'est avéré nettement plus sensible (93 %) et spécifique (98 %) que les tests classiques. Ce type de criblage bénéficie des avancées techniques réalisées dans divers domaines, tels que le criblage à haut-débit, l'informatique, la robotique, l'imagerie, l'apprentissage automatique (ou *machine learning*) et l'organisation des données (ou *data management*). La nature et la précision de la signature phénotypique, extraite par HCS pour un modèle traité par un composé donné, permettent de réaliser des regroupements (ou *clustering*) ou des classements de composés

(classification). Ceux-ci peuvent être utilisés pour caractériser les molécules au niveau biologique, comme c'est le cas avec leurs signatures physico-chimiques (Figure 2). La robustesse des marquages, ainsi que la précision des modalités d'acquisition du HCS peuvent alors donner l'occasion d'évaluer avec finesse la similitude entre composés d'un point de vue phénotypique. Les algorithmes d'apprentissage ou d'intelligence artificielle peuvent également bénéficier de ces données multiparamétriques pour inférer l'activité ou la toxicité des composés et ainsi orienter la sélection des molécules actives en sortie de criblage. En combinant ces liens phénotypiques entre molécules et les données historiques, il devient alors possible de conclure *in silico* quant à l'activité de composés et ainsi orienter le choix des molécules à tester *in vitro* [11, 12].

Les récentes avancées en analyse d'images, rendues possibles grâce à l'utilisation de l'apprentissage profond (*deep learning*) et plus particulièrement d'un type de réseau de neurones artificiels appelé réseau neuronal convolutif (ou *convolutional neural network*, CNN), permettent d'augmenter la quantité et la qualité des informations extraites, tout en se dédouanant des *a priori* sur le type de descripteurs à extraire. Ces

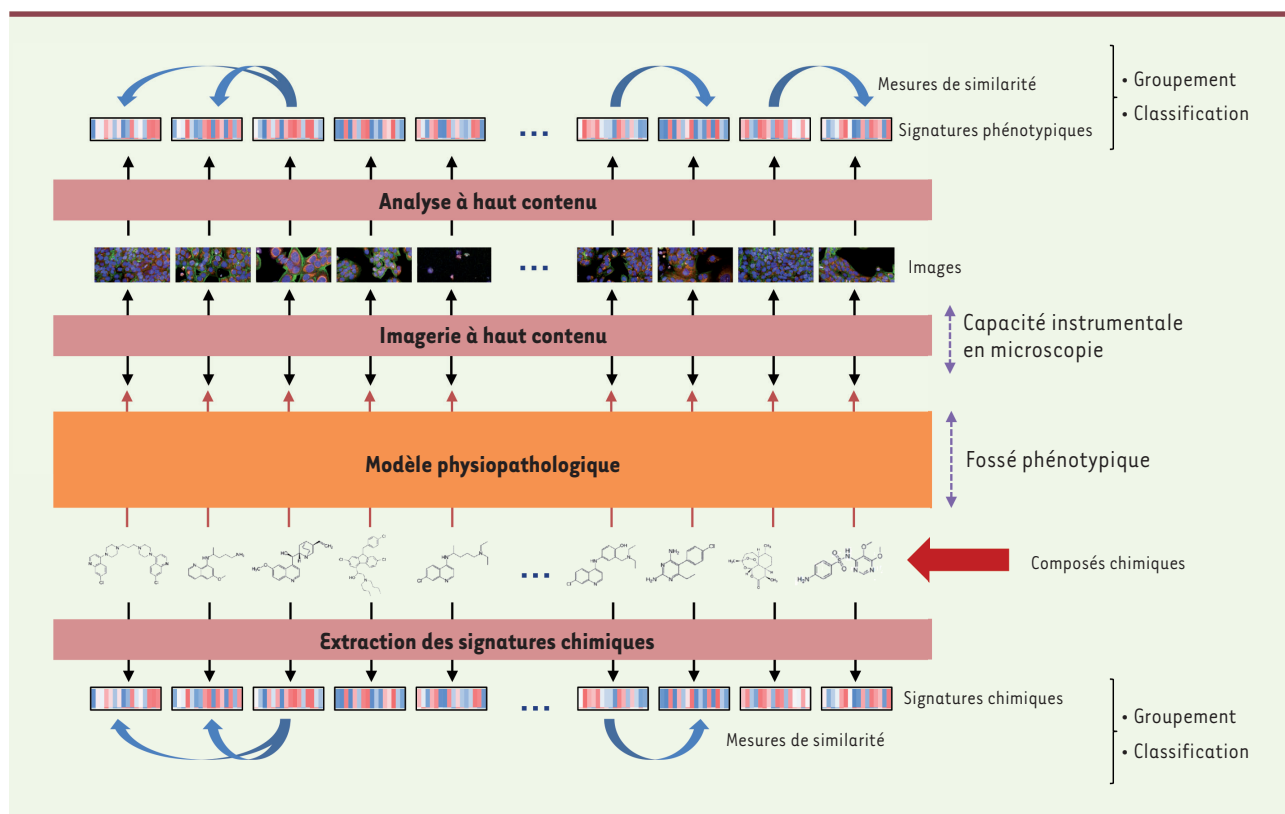


Figure 2. Les composés chimiques (flèche rouge) issus d'une banque peuvent être caractérisés par leurs structures chimiques (partie basse de la figure) ou de façon indirecte par leurs signatures phénotypiques. Ces signatures sont généralement issues de l'analyse automatisée d'images en microscopie d'un modèle physiologique *in vitro*. Ces signatures, accompagnées d'une métrique, permettent le regroupement et potentiellement la classification de ces composés.

deux dernières approches peuvent même être combinées et laissent présager des bouleversements majeurs dans la manière de réaliser des cribles en imagerie.

Précisons qu'au-delà des éléments d'analyse que nous avons présentés, l'impact des résultats générés lors d'un crible HCS sur le développement d'un futur médicament est intimement lié à la pertinence physiopathologique du modèle utilisé. Celui-ci doit en effet répondre à un compromis entre sa capacité à être déployé à grande échelle et sa complexité et /ou sa pertinence physiologique, ce qui nécessite une bonne compréhension des processus biologiques sous-jacents.

La combinaison des modèles cellulaires, physiologiquement plus prédictifs, à des approches de HCS représente une réelle opportunité pour la recherche pharmaceutique, même si la cible impliquée dans la modulation du phénotype d'intérêt n'est pas toujours connue. Selon les recommandations de la FDA, l'identification de la cible moléculaire d'un médicament n'est en effet pas nécessaire pour l'initiation des essais cliniques ou l'approbation finale [13]. Plusieurs exemples montrent en effet que la définition de la cible d'une molécule n'est pas indispensable pour son utilisation, notamment, la cyclosporine, la rapamycine, la rosiglitazone, ou plusieurs classes d'inhibiteurs de canaux calciques, pour lesquels les cibles furent identifiées après leur mise sur le marché [14, 15]. Russell A. Miller *et al.* mettent par

exemple en avant le fait que les biguanides, telle que la metformine, connus depuis 1958, sont utilisés de façon sûre et efficace depuis plusieurs dizaines d'années en première ligne pour le traitement du diabète, bien que l'on ne dispose que de très peu d'informations concernant leur cible moléculaire [16].

Un arsenal méthodologique pour la déconvolution de cibles comme trait d'union entre les approches phénotypiques et cible centriques

La communauté scientifique s'accorde pour faire le constat que les informations relatives à la cible moléculaire, la voie de signalisation impliquée et les mécanismes cellulaires modulés en réponse à un composé, sont autant d'atouts pour mener à bien un programme de découverte de médicaments. Nombreux sont les investigateurs qui considèrent que les actifs issus d'un crible phénotypique sont mieux optimisés par l'utilisation de tests biochimiques. Mais cela suppose de connaître la cible moléculaire du principe actif. En

effet, la connaissance de la cible permet d'implémenter des approches de relation structure/activité, guidées par la connaissance de la structure de la cible, qui sont très appréciées des chimistes médicaux. Ce positionnement résulte très souvent d'un doute sur la qualité statistique des tests phénotypiques ou fonctionnels et, par conséquent, anticipe les difficultés à finement comprendre les impacts structuraux sur l'activité. Néanmoins, même si les tests phénotypiques sont généralement plus difficiles à optimiser et à mettre en place avec un débit suffisant, de nombreux exemples démontrent qu'il est tout à fait possible d'établir des essais aussi robustes opérationnellement et solides statistiquement que des tests biochimiques [17]. Les tests phénotypiques peuvent ainsi être sensibles aux changements structuraux des composés, comme les tests biochimiques [17], même si les tests phénotypiques peuvent être soumis à des barrières d'activité, des facteurs comme la perméabilité, le métabolisme ou l'efflux cellulaire, pouvant compromettre l'interprétation de la relation structure/activité qui est observée. La connaissance de la cible moléculaire représente finalement un point de complémentarité entre approches phénotypiques et approches cible-centriques. Les techniques permettant d'identifier la cible impliquée dans la modulation d'un phénotype d'intérêt (Tableau 1) sont donc clés pour permettre aux approches phénotypiques de délivrer leur plein potentiel. Plusieurs méthodes de déconvolution de cible existent depuis de nombreuses années, mais le regain d'intérêt pour les approches phénotypiques, ainsi que les récentes évolutions techniques élargissent le champ des possibles.

Les méthodes de déconvolution de cible

Une des méthodes les plus classiques est la chromatographie d'affinité. Cette approche consiste tout d'abord à modifier chimiquement la molécule d'intérêt afin de pouvoir l'immobiliser sur un support solide (des billes en général). L'extrait protéique contenant la cible est ensuite incubé avec la molécule immobilisée. Après leur élution, les protéines qui s'étaient fixées à la molécule immobilisées sont identifiées par spectrométrie de masse ou immuno-détection (à l'aide d'anticorps spécifiques). Un des avantages de cette technique est qu'elle permet de tester des extraits protéiques natifs et ainsi de conserver la conformation tridimensionnelle originelle des protéines de l'extrait. Cette approche ne peut néanmoins aboutir que si le composé d'intérêt a une affinité suffisante pour la cible dont la modulation est à l'origine de l'impact phénotypique. De plus, pour permettre son identification, la cible doit être suffisamment exprimée dans le modèle cellulaire utilisé à partir duquel sera réalisé l'extrait de protéines. Malgré ces limitations, cette méthode a, par exemple, permis d'identifier la kinase pyridoxal (PDXK) comme cible de la roscovitine (CYC202), un inhibiteur compétitif de l'ATP, dans différents tissus et cellules, et cela en plus de ses autres cibles déjà connues (ERK[*extracellular signal regulated kinase*]1/2, et CDK[*cyclin-dependent kinase*]2,5) [18]. Cette molécule est actuellement en essai clinique de phase II pour le traitement de la mucoviscidose.

Le phage display

Des méthodes ont été développées en utilisant différents systèmes d'expression. C'est le cas pour la technique de *phage display*. Celle-ci consiste à présenter des molécules, telles que des peptides ou d'autres protéines, à la surface de phages filamenteux, en fusion avec le domaine amino-terminal de leurs protéines d'enveloppe, pIII ou pVIII. Les phages recombinants, qui expriment la protéine à leur surface, sont ensuite sélectionnés pour leur capacité de se lier à la molécule d'intérêt. Après lavages, les phages s'étant liés sont élués, isolés et amplifiés par infection de bactéries, puis leur matériel génétique séquencé. Cette stratégie fondée sur la sélection des phages est nettement plus puissante qu'une stratégie de criblage classique qui nécessite de nombreuses manipulations. Il est en effet possible de cribler 10^6 à 10^{10} molécules recombinantes différentes dans un volume réduit de quelques microlitres. L'association de la protéine, exposée en surface, avec son ADN, codé par le phage, permet ensuite de rapidement identifier la protéine cible. De nombreux exemples d'application du *phage display* à la déconvolution de cibles existent dans la littérature. On peut citer, par exemple, l'identification de la peptidyl-prolyl isomérase FKBP12 comme cible du FK506 (tacrolimus), un immunosuppresseur, en utilisant la forme biotinylée de la molécule [19], ou celle de la calmoduline, comme cible directe de l'HBC³, un anti-cancéreux dérivé du curcumin [20].

Le système de triple hybride

Un autre système d'expression aussi utilisé pour la recherche de cible est la technique du triple-hybride dans la levure [21]. Cette technique nécessite trois hybrides : un premier qui est une fusion entre un domaine de liaison à l'ADN et un ligand connu (par exemple la streptavidine), le deuxième est constitué de la cible du ligand connu (ici, la biotine) liée à la molécule d'intérêt, et le troisième est une fusion entre une séquence promotrice et un ADN complémentaire (ADNc). Lorsque la molécule se lie à la protéine exprimée par le troisième hybride (l'ADNc), un complexe protéique se forme et se lie à l'ADN de levure pour activer l'expression d'un gène rapporteur, révélant la liaison. Plusieurs succès ont été décrits dans la littérature pour différents types de cibles moléculaires [22, 23]. Malgré la puissance de ce système, les principaux inconvénients résident dans l'existence de nombreux faux-positifs et faux-négatifs. La présence de faux-

³ HBC : 4-{3,5-Bis-[2-(4-hydroxy-3-methoxy-phenyl)-ethyl]-4,5-dihydro-pyrazol-1-yl}-benzoic acid).

Méthodes	Principes	Avantages	Inconvénients	Références
Par chromatographie d'affinité	Petite molécule modifiée immobilisée sur colonne. Incubation avec extrait protéique d'intérêt. Séparation par élution et identification par spectrométrie de masse ou immunodétection	Simple à mettre en place. Cible retenue dans des conditions physiologiques	Modification de la petite molécule : gêne stérique? Activité biologique conservée? Molécule de très haute affinité nécessaire. Forte concentration de protéine cible obligatoire	[18]
Par expression de clones (phage display, système triple-hybride, banques d'ADNc)	Expression de protéines cibles potentielles en système hétérologue et crible de la molécule d'intérêt marquée	Permet de détecter les protéines faiblement abondantes	Expression de protéines modifiées dans des systèmes simples (levure, phage) : QUID de la structure/modifications post traductionnelles (MPT), etc.	Phage display : [19, 20] Y3H : [21-23]
Puce à protéines	La molécule d'intérêt fluorescente ou radiomarquée est incubée avec les puces de protéines. Les protéines ainsi marquées sont identifiées selon leur position	Permet une analyse à haut débit des interactions protéine / molécule. Pas de problème de faible abondance	Position de la protéine : liera-t-elle encore la molécule? Modification chimique de la petite molécule (activité...). Protéines isolées : QUID de la structure / MPT ?	[25]
Puce à cellules	Expression d'ADNc définis dans des cellules sur des puces. Incubation avec la molécule d'intérêt fluorescente ou radiomarquée	Criblage de milliers d'ADNc possible. Système stable sur le long-terme	Variation d'efficacité de transfection. Ratio signal/bruit suffisant ?	[26]
Suppression biochimique	Suppression fonctionnelle de l'inhibition chimique <i>in vitro</i> . Mélange molécule/extrait protéique : activité. Des fractions de cet extrait sont administrées pour supprimer cette activité et les fractions induisant la suppression d'activité sont ensuite déconvoluées	Les molécules de faible affinité peuvent être testées	Développement d'un test fonctionnel nécessaire. Peut mener à l'identification d'une protéine de la voie de signalisation	[27]
Technique d'organisation nématique de protéines (NPOT®)	Méthode d'hétéroassemblage par incubation de lysats tissulaires et de la molécule d'intérêt, isolement de l'interactome, digestion enzymatique et identification des partenaires par LC-MS/MS des peptides digérés	Structure quaternaire conservée. Accès au mécanisme d'action de la molécule. Possible prédiction des effets secondaires	Risque de non identification de la cible si digestion enzymatique imparfaite. Méthode brevetée, seulement effectuée par une PME	[28]
CETSA (Cellular Thermal Stability Assay), DARTS (Drug Affinity Response Target Stability)	Mesure de la stabilisation d'une protéine en réponse à un ligand	Pas de modification chimique	La stabilisation thermodynamique doit être suffisante pour être détectable	[29]

Tableau 1. Méthodes de déconvolution existantes de la plus ancienne à la plus récente.

positifs peut être due au fait que de nombreuses protéines appât possèdent des similitudes avec les protéines de levure. Elles peuvent ainsi activer le gène rapporteur sans réelle interaction (*self-activation*). Ils peuvent aussi être causés par forçage du système, dû à la surexpression des protéines dans la levure. Les faux-négatifs sont généralement dus, d'une part à l'impossibilité de détecter des interactions entre protéines dont la localisation n'est pas nucléaire et, d'autre part, à une structure tertiaire incorrecte des protéines.

Les systèmes d'expression

Toujours dans les systèmes d'expression, l'affichage d'ARN messagers (ARNm) représente une approche qui peut aussi être explorée. Des ARNm sont fusionnés à un complexe puromycine/ADN lieur avant d'être soumis à une traduction *in vitro* partielle pour générer un complexe protéine-ARNm-ADN/puromycine. L'ensemble des complexes sont ensuite purifiés et retro-transcrits avant d'être exposés à la molécule active immobilisée qui les capture. Ils sont ensuite élués avant amplification par PCR (*polymerase chain reaction*) des ADNc précédemment générés. Ces ADNc amplifiés sont ensuite transcrits et les ARNm ainsi produits subissent alors un nouveau cycle de fusion/traduction/immobilisation/transcription inverse/amplification, jusqu'à l'obtention d'un ADNc unique, qui sera alors séquencé pour identifier la protéine se liant à la molécule active [24]. Bien que cette approche présente l'avantage d'amplifier l'expression des cibles potentielles et qu'elle permette l'évaluation de banques de grandes tailles, elle est relativement lourde à implémenter, sans garantie de succès. L'expression se fait en effet dans des modèles simples et la structure tridimensionnelle native de la cible n'est pas toujours assurée. Les modifications post-traductionnelles entrant potentiellement en jeu pour la liaison et l'activité du composé peuvent également ne pas être présentes sur les molécules.

Les techniques sur puces

L'évolution des technologies de miniaturisation a permis le développement de différentes approches fondées sur l'utilisation de puces. Parmi ces approches, les puces à protéines présentent un certain nombre d'avantages. Les protéines sont immobilisées sur des micro-puces, puis incubées avec une version fluorescente ou radiomarquée de la molécule. Après lavage, les protéines marquées sont identifiées par leur position sur la puce. Cette approche permet d'exposer un grand nombre de protéines au composé d'intérêt, et cela de façon identique pour toutes les protéines. Quelques succès ont pu être obtenus avec cette technologie [25]. Cependant, certaines limitations sont à prendre en considération, comme par exemple, le fait que des gênes stériques, liées à l'immobilisation des protéines, puissent limiter leur accessibilité, ou encore l'absence de certaines modifications post-traductionnelles ou de certains partenaires, qui peut diminuer l'affinité du composé pour la protéine isolée. Les puces cellulaires ont l'avantage de présenter les protéines dans un environnement cellulaire. Après immobilisation de certaines d'ADNc à des positions bien définies sur la puce, les cellules sont ensuite ajoutées sur celle-ci, en présence d'un agent de transfection afin que les protéines soient

exprimées sur un support unique. Les cellules sont ensuite incubées avec la/les molécule(s) fluorescentes ou radiomarquées. La localisation du marquage permet d'identifier la ou les protéines ciblées par le composé d'intérêt. Cette approche a récemment permis l'identification du récepteur GFRAL (*GDNF family receptor alpha-like*) comme la cible du facteur de croissance GDF15 (*growth/differentiation factor-15*) [26]. Comme pour toute technique d'expression, l'efficacité de la transfection peut limiter la sensibilité de détection et être à l'origine de faux-négatifs, selon les différentes lignées cellulaires et protéines impliquées. Cette approche est par ailleurs essentiellement adaptée aux récepteurs et aux protéines de surface, les techniques de perméabilisation cellulaire pouvant altérer les sites de liaison des composés.

Les techniques sans marquage

Toutes les techniques présentées jusqu'ici nécessitent une modification chimique de la molécule bioactive. Cet impératif peut être très limitant car il peut être associé à une perte d'affinité ou d'activité de la molécule. Un certain nombre de techniques sans marquage ont donc été développées pour contourner cette limitation. Parmi celles-ci, la suppression biochimique peut être mise en œuvre lorsqu'un test d'activité est disponible. Dans cette approche, un extrait protéique est incubé avec une molécule connue pour inhiber une activité mesurable dans un test : c'est « l'extrait inhibé ». Puis un « extrait non inhibé » est fractionné biochimiquement et les fractions concentrées sont administrées sur l'extrait inhibé et l'activité est mesurée. La fraction supprimant l'inhibition est fractionnée plusieurs fois jusqu'à isoler la protéine d'intérêt. Un exemple d'utilisation de cette approche est l'identification du complexe CDC42 (*cell division cycle 42*) - RhoGDI (*Rho GDP-dissociation inhibitor*) comme cible d'un composé qui avait pour effet d'inhiber la polymérisation de l'actine induite par le phosphatidyl inositol diphosphate (PIP_2) dans des extraits protéiques d'œufs de xénope [27]. Bien que prometteuse, cette technique nécessite le développement d'un test d'activité sensible et robuste, ainsi qu'un travail de validation rigoureux. De plus, l'interprétation des résultats peut être complexe, notamment lorsque la voie de signalisation impliquée n'est pas connue.

Au-delà de l'utilisation de composés sans marquage, les techniques les plus modernes de déconvolution se font sur des extraits protéiques natifs, et reposent sur la liaison du composé sur sa cible, et non sur l'activité du composé. Parmi ces approches, la technique d'organisation nématique de protéines, c'est-à-dire

intermédiaire entre les phases solide cristalline et liquide, consiste à incuber la molécule directement avec les lysats tissulaires ou cellulaires supposés contenir la protéine cible. La séparation des macromolécules est effectuée sur un gradient de pH, et lorsque la molécule se lie à un complexe protéique, l'hétéro-assemblage qui se forme peut être isolé manuellement. Les protéines contenues dans ce complexe sont identifiées par spectrométrie de masse. Une liste des protéines composant l'interactome lié à la molécule d'intérêt est ainsi obtenue, permettant de caractériser la voie de signalisation impliquée et son mode d'action. Bien que relativement récente, cette technique a permis d'identifier les cibles de plusieurs composés d'intérêt comme par exemple ABCA1 (*ATP-binding cassette transporter A1*), un transporteur clé pour l'efflux du cholestérol, et l'évodiamine, un produit naturel utilisé pour la perte de poids [28]. Bien que cette technique soit sensible et permette d'accéder à la cible directement dans son environnement natif, il est important de préciser que certaines protéines, telles que les récepteurs transmembranaires par exemple, sont difficilement hydrolysables par les protéases utilisées lors de la lyse des cellules. Il y a donc un risque que certaines protéines ne soient pas accessibles et donc pas identifiables.

Les techniques DARTS et CETSA

Les techniques DARTS (*drug affinity responsive target stability*) ou CETSA (*cellular thermal shift assay*) permettent l'identification de cibles en évaluant les différences entre les fractions liées et non liées. La méthode DART est fondée sur le concept qu'une protéine liée à son ligand a une stabilité modifiée par rapport à sa forme libre. Ce changement de stabilité est évalué par l'analyse du profil de protéolyse, en présence de différentes protéases, de la protéine associée ou non à son ligand. Celui-là pouvant la protéger, son profil de migration en électrophorèse sur gel est alors différent de celui la protéine libre. Cette technique a permis l'identification de l'eIF4A (*eukaryotic initiation factor 4A*) comme cible du resvératrol, un produit naturel présent dans le raisin et le vin ayant des propriétés bénéfiques sur l'espérance de vie [29]. Une des principales limites de cette méthode est, ici encore, le fait que certaines protéines sont difficilement hydrolysables par les protéases. La méthode CETSA est, quant à elle, fondée sur la stabilisation thermodynamique de la protéine cible par son ligand. Les lysats cellulaires ou tissulaires, traités ou non, sont chauffés à différentes températures et les protéines solubilisées sont séparées de celles qui sont précipitées et déstabilisées, puis détectées par électrophorèse sur gel. Les protéines cibles potentielles sont identifiées par comparaison des profils de migration en fonction des différents paramètres. Pour que les approches DART ou CETSA soient efficaces, il faut que le composé d'intérêt stabilise sa cible et que cette stabilisation soit suffisamment importante pour être mesurée.

Les approches *in silico*

En plus des approches impliquant directement le composé d'intérêt, les technologies « omiques » couplées à des approches *in silico* permettent d'identifier les voies de signalisation impliquées et de définir des protéines pouvant être ciblées. Dans ce domaine, les approches de

génomique fonctionnelle, qui consistent à cribler des collections d'ADNc, de petits ARN interférents (siRNA) ou de petits ARN en épingle à cheveux (shRNA) et plus récemment d'ARN guides (gRNA) *via* la technique de CRISPR/Cas9, sont des options qui peuvent produire d'excellents résultats. Ces différentes modalités sont généralement déployées en utilisant le même test phénotypique que celui utilisé pour identifier la molécule d'intérêt. Ici, la surexpression ou la répression de cibles potentielles permettent de mimer le phénotype d'intérêt et ainsi potentiellement d'identifier la cible du composé d'intérêt. De nombreux exemples utilisant des approches de génomique fonctionnelle existent et sont décrits dans différentes revues [30, 31]. Une application particulièrement puissante de ces approches est la *connectivity map*, créée par le *Broad Institute*⁴, qui associe un large ensemble de signatures phénotypiques obtenues sur la base de perturbations génétiques à des données pharmacologiques. Ces différentes approches, notamment siRNA, shRNA et gRNA, génèrent cependant un certain nombre de faux-positifs. Un travail rigoureux de validation est donc nécessaire pour confirmer la ou les cibles ainsi identifiées. Ces approches sont également coûteuses, surtout lorsqu'elles impliquent des collections ciblant tout le génome. Les technologies de génomique fonctionnelle appliquées à la déconvolution de cibles sont donc très souvent combinées avec des approches *in silico* afin de diminuer la taille des cribles et ainsi réduire les coûts.

Les approches dites de profilage *in silico* tirent parti de l'ensemble des données disponibles. Elles incluent les données « omiques » (génomique, transcriptomique, métabolomique, etc.) qui peuvent être associées au phénotype observé, et les données physico-chimiques et chémo-informatiques associées à la petite molécule d'intérêt. Ainsi, il existe des méthodes qui permettent d'utiliser la structure tridimensionnelle du composé d'intérêt pour évaluer virtuellement toutes les protéines pour lesquelles les structures 3D (en 3 dimensions) sont connues et ainsi prédire les cibles potentielles [32, 33]. Des bases de données biologiques et/ou phénotypiques peuvent aussi compléter l'approche, incluant la comparaison de l'activité du composé d'intérêt à l'activité observée avec un ensemble de composés de référence dont les cibles sont connues, afin d'inférer l'identité potentielle de la cible du composé [34]. Dans ce domaine aussi, l'apprentissage profond et les réseaux convolutionnels permettront des avancées.

⁴ Le *Broad Institute* (Boston, États-Unis) est un institut alliant le *Massachusetts Institute of Technology* (MIT) et l'université Harvard, ainsi que les hôpitaux affiliés à cette université.



Conclusions

L'évolution technologique du criblage à haut contenu combinée aux importants progrès réalisés dans des domaines aussi variés que les modèles cellulaires *in vitro* et les stratégies de déconvolution de cibles, sont à l'origine d'un fort regain d'intérêt pour les approches phénotypiques dans le contexte de la recherche de nouveaux candidats médicaments. La combinaison de ces diverses techniques permet désormais d'appréhender la découverte de médicaments de façon plus intégrée d'un point de vue mécanistique en déployant les approches phénotypiques pour optimiser des molécules dont la cible moléculaire est déjà connue, ou pour découvrir des molécules candidates qui seront ensuite caractérisées dans un test moléculaire précis. Il ne fait, par ailleurs, aucun doute que les évolutions actuelles dans le domaine de l'intelligence artificielle vont amplifier cette tendance et que nous ne sommes qu'à l'aube d'une nouvelle donne pour les approches phénotypiques. ♦

SUMMARY

From high content screening to target deconvolution: New insights for phenotypic approaches

The advent of the molecular biology and the completion of the human genome sequencing prompted the pharmaceutical industry to progressively implement target-centric drug discovery strategies. However, concerns regarding the research and development productivity during the last ten years, combined with technological developments in high-content screening, automation, image analysis and artificial intelligence triggered a renewed interest for the phenotypic drug discovery approaches. Target-centric and phenotypic approaches are more and more considered complementary, hence, positioning the target deconvolution on the critical path. This review analyzes the evolution of the target-centric and phenotypic approaches, focusing more specifically on the high-content screening and the target deconvolution technologies currently available. ♦

LIENS D'INTÉRÊT

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

RÉFÉRENCES

- Wouters OJ, McKee M, Luyten J. Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA* 2020 ; 323 : 844-53.
- Swinney DC, Anthony J. How were new medicines discovered? *Nat Rev Drug Discov* 2011 ; 10 : 507-19.
- Eder J, Sedrani R, Wiesmann C. The discovery of first-in-class drugs: origins and evolution. *Nat Rev Drug Discov* 2014 ; 13 : 577-87.
- Moffat JG, Rudolph J, Bailey D. Phenotypic screening in cancer drug discovery - past, present and future. *Nat Rev Drug Discov* 2014 ; 13 : 588-602.
- Dorval T, Chanrion B, Cattin ME, Stephan JP. Filling the drug discovery gap: is high-content screening the missing link? *Curr Opin Pharmacol* 2018 ; 42 : 40-5.
- Prudent R, Soleilhac E, Barette C, et al. Les criblages phénotypiques ou comment faire d'une pierre deux coups - Découvrir la cible et la molécule pharmacologique capable de la réguler. *Med Sci (Paris)* 2013 ; 29 : 897-905.
- Brodin P, DelNery E, Soleilhac E. Criblage phénotypique à haut contenu pour la chémobiologie et ses enjeux. *Med Sci (Paris)* 2015 ; 31 : 187-96.
- Xu JJ, Henstock PV, Dunn MC, et al. Cellular imaging predictions of clinical drug-induced liver injury. *Toxicol Sci* 2008 ; 105 : 97-105.
- Gough AH, Chen N, Shun TY, et al. Identifying and quantifying heterogeneity in high content analysis : application of heterogeneity indices to drug discovery (Paulmurugan R ed). *PLoS One* 2014 ; 9 : e102678.
- O'Brien PJ, Irwin W, Diaz D, et al. High concordance of drug-induced human hepatotoxicity with *in vitro* cytotoxicity measured in a novel cell-based model using high content screening. *Arch Toxicol* 2006 ; 80 : 580-604.
- Simm J, Klambauer G, Arany A, et al. Repurposing high-throughput image assays enables biological activity prediction for drug discovery. *Cell Chem Biol* 2018 ; 25 : 611-8.e3.
- Hofmarcher M, Rumetshofer E, Clevert DA, et al. Accurate prediction of biological assays with high-throughput microscopy images and convolutional networks. *J Chem Inf Model* 2019 ; 59 : 1163-71.
- Guidance for Industry. *Content and format of investigational new drug applications (INDs) for phase 1 studies of drugs, including well-characterized, therapeutic, biotechnology-derived products*. Center for Drug Evaluation and Research-Center for Biologics Evaluation and Research, 1995.
- Young PR. Perspective on the discovery and scientific impact of p38 MAP-kinase. *J Biomol Screen* 2013 ; 18 : 1156-63.
- Zheng W, Thorne N, McKew JC. Phenotypic screens as a renewed approach for drug discovery. *Drug Discov Today* 2013 ; 18 : 1067-73.
- Miller RA, Chu Q, Xie J, et al. Biguanides suppress hepatic glucagon signalling by decreasing production of cyclic AMP. *Nature* 2013 ; 494 : 256-60.
- Lee JA, Berg EL. Neoclassic drug discovery: the case for lead generation using phenotypic and functional approaches. *J Biomol Screen* 2013 ; 18 : 1143-55.
- Bach S, Knockaert M, Reinhardt J, et al. Roscovitine targets, protein kinases and pyridoxal kinase. *J Biol Chem* 2005 ; 280 : 31208-19.
- Sche PP, McKenzie KM, White JD, Austin DJ. Display cloning: functional identification of natural product receptors using cDNA-phage display. *Chem Biol* 1999 ; 6 : 707-16.
- Shim JS, Lee J, Park HJ, et al. A new curcumin derivative, hbc, interferes with the cell cycle progression of colon cancer cells via antagonization of the Ca²⁺/calmodulin function. *Chem Biol* 2004 ; 11 : 1455-63.
- Licitra EJ, Liu JO. A three-hybrid system for detecting small ligand-protein receptor interactions. *Proc Natl Acad Sci USA* 1996 ; 93 : 12817-1.
- Becker F, Murthi K, Smith C, et al. A Three-hybrid approach to scanning the proteome for targets of small molecule kinase inhibitors. *Chem Biol* 2004 ; 11 : 211-23.
- Chidley C, Haruki H, Pedersen MG, et al. A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nat Chem Biol* 2011 ; 7 : 375-83.
- McPherson M, Yang Y, Hammond PW, Kreider BL. Drug receptor identification from multiple tissues using cellular-derived mRNA display libraries. *Chem Biol* 2002 ; 9 : 691-8.
- Huang J, Zhu H, Haggarty SJ, et al. Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. *Proc Natl Acad Sci USA* 2004 ; 101 : 16594-9.
- Mulligan SE, Lin-Schmidt X, Chin CN, et al. GFRAL is the receptor for GDF15 and the ligand promotes weight loss in mice and nonhuman primates. *Nat Med* 2017 ; 23 : 1150-7.
- Peterson JR, Lebensohn AM, Pelish HE, Kirschner MW. Biochemical suppression of small-molecule inhibitors: a strategy to identify inhibitor targets and signaling pathway components. *Chem Biol* 2006 ; 13 : 443-52.
- Wang L, Eftekhari P, Schachner D, et al. Novel interactomics approach identifies ABCA1 as direct target of evodiamine, which increases macrophage cholesterol efflux. *Sci Rep* 2018 ; 8 : 1-10.
- Lomenick B, Hao R, Jonai N, et al. Target identification using drug affinity responsive target stability (DARTS). *Proc Natl Acad Sci USA* 2009 ; 106 : 21984-9.
- Nijman SMB. Functional genomics to uncover drug mechanism of action. *Nat Chem Biol* 2015 ; 11 : 942-8.
- Neggers JE, Kwanten B, Dierckx T, et al. Target identification of small molecules using large-scale CRISPR-Cas mutagenesis scanning of essential genes. *Nat Commun* 2018 ; 9 : 1-14.
- Kufareva I, Chen YC, Ilatovskiy AV, et al. Compound activity prediction using models of binding pockets or ligand properties in 3D. *Curr Top Med Chem* 2012 ; 12 : 1869-82.
- Tsubaki M, Tomii K, Sese J. Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics* 2019 ; 35 : 309-18.
- Perlman Z, Slack MD, Feng Y. Multidimensional drug profiling by automated microscopy. *Science* 2004 ; 30 : 1194-8.

TIRÉS À PART

J.P. Stéphan