médecine/sciences 2014 ; 30 : 1152-60

Chémobiologie (1)

Les chimiothèques et le criblage virtuel

Didier Rognan[1], Pascal Bonnet[2]

› Le préalable indispensable à tout criblage virtuel est la définition de la ou des chimiothèques à cribler. Comme nous le verrons ici, de nombreuses sources sont disponibles et leur sélection, ou non, dépendra du projet dans lequel le criblage virtuel s'insère. Cette étape de sélection est au moins aussi importante que la méthode de criblage elle-même. Cette revue détaille les principales chimiothèques utilisables à des fins de criblage virtuel, et oriente le lecteur vers le meilleur choix possible en fonction de ses besoins. ‹

[1] Laboratoire d'innovation thérapeutique, UMR 7200 CNRS-université de Strasbourg, MEDALIS *drug discovery center*, 74, route du Rhin, F-67400 Illkirch, France ;
[2] Institut de chimie organique et analytique (ICOA), UMR CNRS-université d'Orléans 7311, université d'Orléans, rue de Chartres, F-45067 Orléans Cedex 02, France.
rognan@unistra.fr

Le criblage virtuel regroupe diverses méthodes informatiques visant à sélectionner, parmi les molécules d'une ou plusieurs chimiothèques, celles qui répondent à un cahier des charges bien précis quant à des propriétés physicochimiques (ex. : solubilité aqueuse), pharmacocinétiques (ex. : perméation membranaire) ou le plus souvent pharmacologiques (liaison à une protéine cible). Cette revue ne s'intéresse pas aux diverses méthodes de criblage virtuel pour lesquelles la littérature est abondante [1-3]. Nous nous focaliserons simplement sur une description des chimiothèques qu'il est possible de cribler virtuellement et en analyserons le contenu.

Le criblage virtuel repose sur le principe inverse de celui de la sérendipité : *on ne trouve que ce que l'on cherche.* Le choix de la ou des chimiothèques criblées influe donc sur la qualité des touches identifiées. Cette affirmation peut paraître bien naïve. Néanmoins, à la question « Pourquoi avez-vous choisi cette chimiothèque ? », il est souvent répondu « Parce qu'elle était disponible au laboratoire ». Cette revue a pour mission de vous montrer combien les chimiothèques peuvent être diverses et de vous assister dans la meilleure sélection possible.

Chimiothèques commerciales

Depuis une dizaine d'années et l'essor du criblage miniaturisé à haut débit, de nombreuses sociétés et institutions académiques proposent des collections de molécules destinées à fournir des touches lors de criblages biologiques *(Tableau I)*. Ces molécules sont disponibles sous forme de poudre en quantité variable (1, 5, 10 mg, voire plus) en l'espace de 3 à 4 semaines sur commande auprès du fournisseur *(Figure 1)*. Initialement issues des patrimoines de laboratoires académiques, ces chimiothèques ont d'abord été alimentées par chimie combinatoire, garantissant ainsi un grand nombre de molécules aux dépens de leur diversité chimique [4]. Les exigences des compagnies pharmaceutiques ont permis à ces chimiothèques d'évoluer vers une meilleure qualité (diversité, nouveauté, pureté, caractérisation analytique). Plus de 20 millions de molécules uniques sont ainsi aujourd'hui accessibles au chémoinformaticien.

Il est juste de reconnaître que la même molécule peut exister chez différents fournisseurs. Il est donc important, au cas où plusieurs chimiothèques sont sélectionnées, d'en éliminer la redondance. Dans la mesure où chaque fournisseur propose une majorité de molécules uniques, il est en général intéressant de les fusionner afin de constituer l'espace chimique le plus diversifié possible. Étant donné le grand nombre de fournisseurs différents, certaines sociétés privées et laboratoires académiques archivent l'ensemble de cette information en un portail unique afin de faciliter les requêtes de l'utilisateur *(Tableau II)*.

Ces portails sont interrogeables soit librement (Ambinter, Bioinfo-DB, ChemSpider, e-molecules, ZINC), soit après abonnement (ChemNavigator), afin de télécharger les structures et propriétés des molécules et se constituer ainsi une version électronique de ces chimiothèques. Il est aussi possible de ne sélectionner que des molécules partageant un même châssis structural *(Figure 2)* afin d'établir une chimiothèque focalisée.

Cet article fait partie de la série « Chémobiologie ».

DOI : 10.1051/medsci/20143012019

Fournisseur	Nombre de molécules	Site internet
AnalytiCon	31 897	http://www.ac-discovery.com/
Asinex	484 299	http://www.asinex.com/
Bionet	41 601	http://www.keyorganics.net
Specs	503 912	http://www.specs.net
Chembridge	919 685	http://www.chembridge.com
ChemDiv	1 320 696	http://www.chemdiv.com
Enamine	1 344 255	http://www.enamine.net
InterBioScreen	508 012	http://www.ibscreen.com/
LaboTest	114 283	http://www.labotest.com/
LifeChemicals	357 951	http://www.lifechemicals.com/
Maybridge	30 125	http://www.maybridge.com/
Otava	437 583	http://www.otavachemicals.com/
Peakdale	14 644	http://www.peakdale.co.uk/
Pharmeks	274 598	http://www.pharmeks.com/
PrincetonBio	543 557	http://www.princetonbio.com/
TimTec	923 182	http://www.timtec.net/
TRC	27 877	http://www.trc-canada.com/
TOSLab	17 585	http://www.toslab.com/
Vitas-M	1 227 380	http://www.vitasmlab.com/
Uorsy	1 408 603	http://www.ukrorgsynth.com/

Tableau I. ***Principales chimiothèques commerciales.***

Il convient toutefois de faire attention aux différences notables entre ces divers portails d'accès, qui portent notamment sur :

– la gestion de la redondance de l'information (entités moléculaires dupliquées) ;

– la gestion des contre-ions (inclus ou non) (*voir Glossaire*) ;

– le filtrage préalable en sous-ensembles d'intérêt (ex : sous-ensembles « *drug-like* » [5], « *lead-like* » [6], « *fragment-like* » [7] dans la base ZINC [8]) ;

– la disponibilité des structures dans des formats moléculaires (2D ou 3D) propices au criblage virtuel (format de type smiles, sdf, ou mol2) [9] ;

– la gestion des formes tautomères (*voir Glossaire*) lors d'une recherche de molécules par sous-structure ou châssis.

Ces chimiothèques, dont le téléchargement des structures est gratuit, sont la source principale de diversité moléculaire pour la très grande majorité des criblages virtuels. Elles évoluent régulièrement avec un taux de jouvence très significatif (de l'ordre de 25-30 % par an). Il convient donc d'être attentif au laps de temps séparant le téléchargement de la chimiothèque, son criblage, puis l'achat des molécules sélectionnées. Ce délai doit être aussi court que possible (idéalement moins de trois mois) afin d'éviter que la touche ait disparu du catalogue fournisseur. Le prix d'achat d'une molécule commerciale (quelques milligrammes de poudre pour une validation expérimentale *in vitro*) peut varier en fonction du fournisseur (compter de 50 à 150 € pour 5 mg), mais également en fonction du nombre de molécules achetées. Si économiquement, il est plus rentable de se fournir auprès d'un fournisseur unique, il est scientifiquement plus logique de multiplier les sources. Le cribleur devra donc en permanence résoudre ce dilemme afin d'optimiser coût et diversité des molécules sélectionnées par criblage virtuel à des fins de confirmation expérimentale.

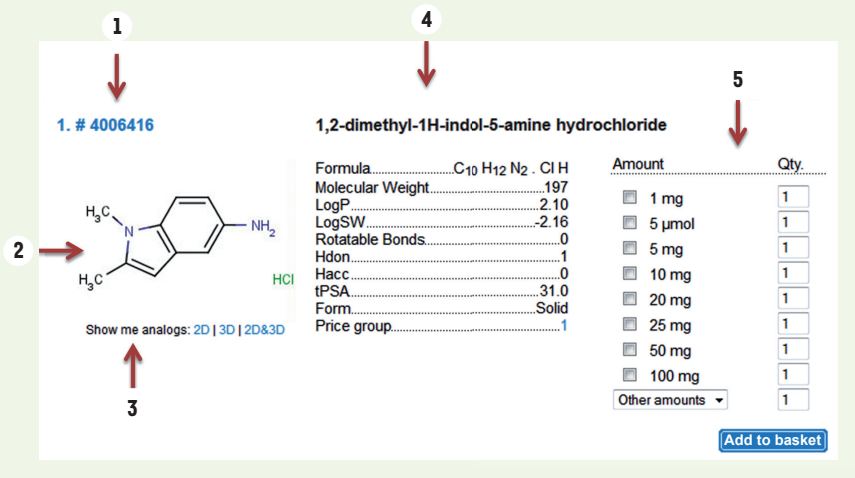

Figure 1. ***Description d'une molécule disponible commercialement dans la chimiothèque Hit2Lead de la société ChemDiv.*** Parmi les informations disponibles figurent la référence catalogue ***(1)***, la structure ***(2)***, les analogues proches ***(3)***, les propriétés physicochimiques ***(4)*** et les quantités disponibles à la commande ***(5)***.

Nom	Nombre de molécules	Nombre de sources	Site Internet
Ambinter	22 000 000	220	http://www.ambinter.com
Bioinfo-DB	3 210 000	22	http://bioinfo-pharma.u-strasbg.fr/bioinfo
ChemNavigator	60 000 000	> 200	http://www.chemnavigator.com
ChemSpider	28 000 000	400	http://www.chemspider.com
e-Molecules	8 941 907	25	http://www.emolecules.com
ZINC	21 000 000	247	http://zinc.docking.org

Tableau II. *Portails d'accès à diverses chimiothèques commerciales.*

Chimiothèques académiques

De nombreuses chimiothèques regroupant le patrimoine de laboratoires académiques ont vu le jour dans de nombreux pays. Le Centre national de la recherche scientifique (CNRS) ainsi que de nombreux partenaires académiques ont joué un rôle pionnier en Europe dans ce domaine par la création de la Chimiothèque nationale [10, 11]. Depuis septembre 2003, cette chimiothèque a pour mission principale de regrouper les collections de produits de synthèse, de composés naturels et d'extraits naturels existants dans les laboratoires publics français et d'en promouvoir la valorisation scientifique et industrielle. Les informations concernant les molécules et extraits disponibles sont regroupées dans deux bases de données nationales (produits de synthèse et composés naturels d'une part, extraits naturels d'autre part), accessibles *via* un portail internet [11]. La Chimiothèque nationale référence aujourd'hui 56 000 molécules de synthèse et 14 300 extraits naturels. Ces produits sont conditionnés soit en vrac, soit en microplaques de 96 puits, afin de permettre leur évaluation biologique ciblée ou systématique, respectivement. Il est à noter qu'un sous-ensemble de cette collection (Chimiothèque nationale essentielle) de 640 molécules sélectionnées selon des critères de diversité et de représentativité, est également disponible.

Les chimiothèques académiques, tout du moins au niveau national, sont généralement très diverses [4], car issues de laboratoires différents apportant chacun un espace chimique propre à son histoire. En cas d'identification de touches, il est assez facile de remonter à l'équipe de recherche dont sont issues les molécules d'intérêt, de sorte qu'une optimisation par chimie médicinale est facile à implémenter grâce à des collaborations scientifiques interdisciplinaires. Dans le cadre du programme Européen ESFRI (*European strategy forum on research infrastructures*), il est à noter que le projet EU-OPENSCREEN [12] prépare actuellement la structuration de neuf chimiothèques nationales en une chimiothèque européenne.

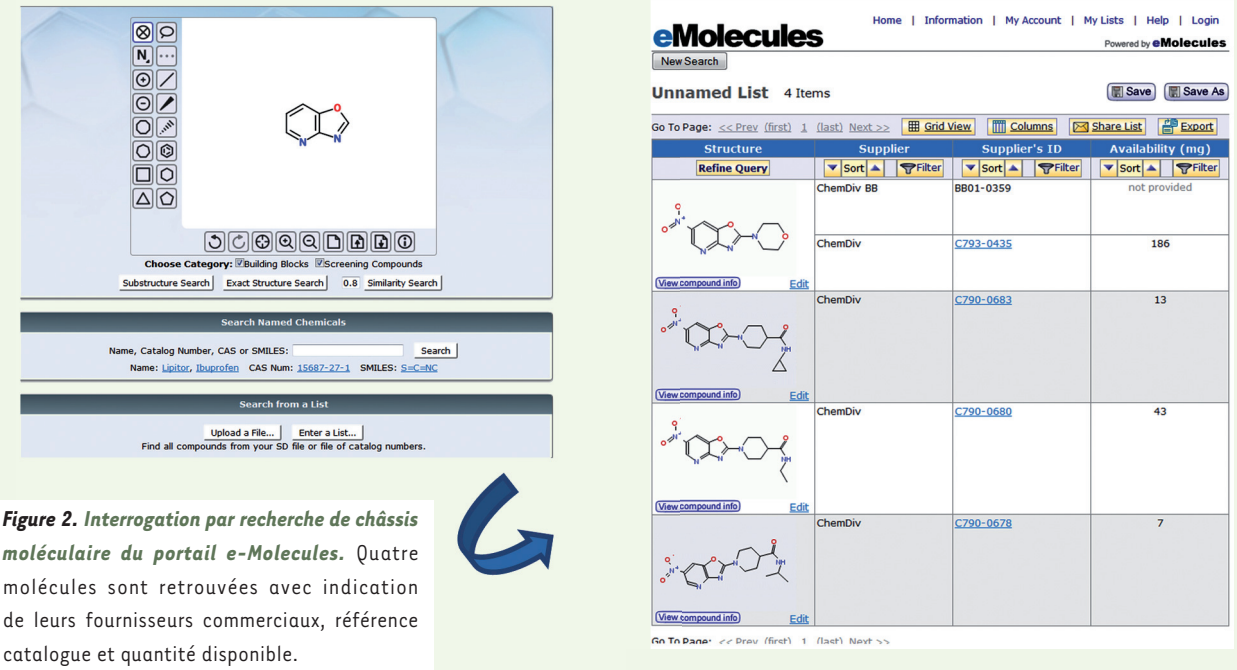

Figure 2. ***Interrogation par recherche de châssis moléculaire du portail e-Molecules.*** Quatre molécules sont retrouvées avec indication de leurs fournisseurs commerciaux, référence catalogue et quantité disponible.

Nom	Nombre de molécules	Disponibilité	Commentaires	Site internet
DrugBank	1 527	Électronique	Approuvé (FDA)	www.drugbank.ca
e-Drug3D	1 632	Électronique	Approuvé (FDA)	http://chemoinfo.ipmc.cnrs.fr/MOLDB/
Integrity	8 348	Électronique	Payant	https ://integrity.thomson-pharma.com/integrity/xmlxsl/
Prestwick	1 200	Électronique, poudre, microplaques	Approuvé (FDA, EMEA)	www.prestwickchemical.com/
Sigma-Aldrich	1 280	Électronique, poudre, microplaques	Sondes pharmacologiques	www.sigmaaldrich.com/LOPAC

Tableau III. *Chimiothèques de médicaments.*

Chimiothèques de médicaments, candidats cliniques et sondes pharmacologiques

Une des applications les plus prometteuses du criblage en général (aussi bien réel que virtuel) consiste à repositionner des molécules très bien caractérisées (médicaments approuvés, molécules en phase d'études cliniques) pour une nouvelle indication thérapeutique. Il s'agit d'identifier des protéines nouvelles auxquelles une molécule connue pourrait se lier, ce qui pourrait induire des effets thérapeutiques nouveaux. Cette stratégie a deux avantages majeurs : (1) elle se limite à des molécules excessivement bien caractérisées dont les études précliniques initiales ne sont plus à faire, (2) elle permet de prolonger la vie d'une molécule déjà développée au-delà de son brevet d'application initial. Le repositionnement nécessite une chimiothèque de médicaments approuvés ou en passe de l'être. Parmi les principales sources possibles *(Tableau III)*, on peut citer les bases de données DrugBank [13] et Integrity [14], qui ne sont disponibles que sous forme électronique (payante pour cette dernière), et qui présentent le désavantage de ne pas proposer de sources commerciales pour obtenir et tester les molécules éventuellement retenues. Prestwick Chemical [15] et Sigma-Aldrich [16] fournissent des chimiothèques de médicaments connus ou de sondes pharmacologiques de référence, disponibles sous formes physique aussi bien qu'électronique, particulièrement bien adaptées au criblage virtuel de repositionnement.

Bases de données de bioactivité

Le développement considérable à la fois des essais biologiques/pharmacologiques miniaturisés et des

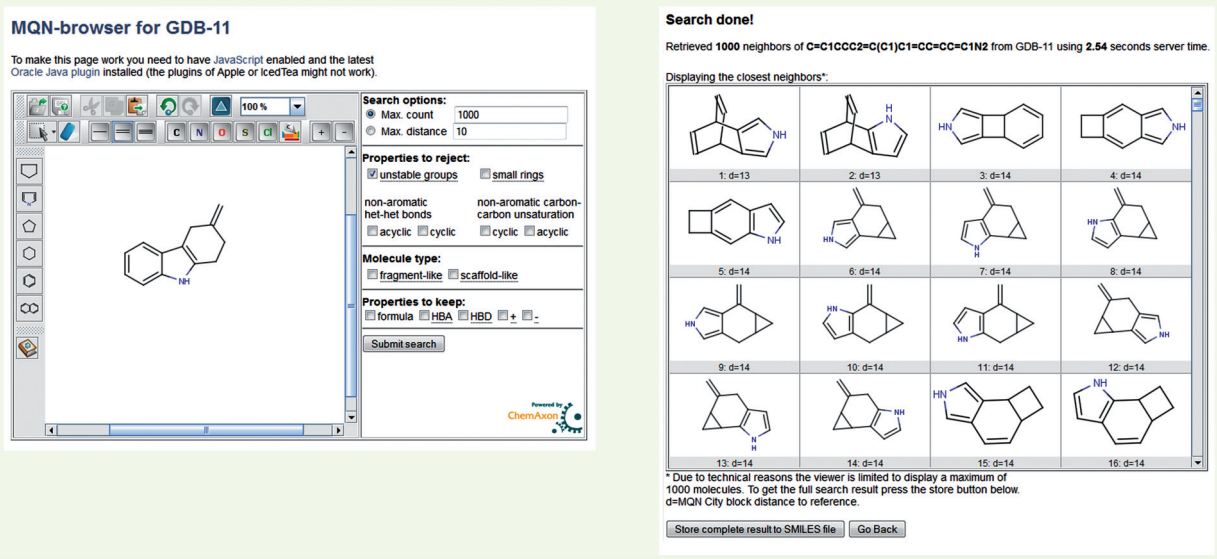

Figure 3. ***Interrogation de la base GDB-11.*** Une requête structurale simple (panel de gauche) combinée à deux propriétés (< 1 000 réponses, exclusion de groupements réactifs) permet l'obtention des structures possibles (panel de droite) les plus proches de la requête initiale.

Nom	Contenu	Site internet	Publique
BindingDB	427 000 molécules 6 589 cibles > 1 million données	http://www.bindingdb.org	Oui
ChemBank	1 266 759 de molécules 2 900 essais	http://chembank.broadinstitute.org	Oui
ChemBioBase	> 2 millions de molécules 1 500 cibles	http://www.jubilantbiosys.com	Non
ChEMBL	1,5 million de molécules 9 400 cibles > 12 millions de données	http://www.ebi/ac.uk/chembl	Oui
GOSTAR	6 millions de molécules 16 millions données	http://www.gostardb.com	Non
KKB	270 000 inhibiteurs 500 kinases 1,3 million de données	http://www.eidogen-sertanty.com	Non
IUPHAR-DB	2 200 molécules 1 150 récepteurs	http://www.iuphar-db.org	Oui
MDDR	> 150 000 molécules	http://www.symyx.com	Non
PDSP	1 500 molécules 55 000 données (Ki)	http://pdsp.med.unc.edu/pdsp.php	Oui
PubChem BioAssay	2 millions de molécules 8 000 cibles 200 millions de données	http://pubchem.ncbi.nlm.nih.gov	Oui
Wombat	330 000 molécules 1966 cibles 900 000 données d'activité	http://www.sunsetmolecular.com	non

Tableau IV. ***Différentes bases de bioactivité.***

chimiothèques disponibles a permis la mise à disposition d'une quantité phénoménale, mais hétérogène, d'informations sur les molécules bioactives, leurs cibles protéiques et leurs activités biologiques. Ces données sont rassemblées dans des bases de données de bioactivité (parfois appelées bases chémogénomiques [2]), dont beaucoup sont accessibles gratuitement *(Tableau IV)*. Ces bases de données sont actuellement très importantes dans l'industrie pharmaceutique, car elles permettent une approche globale et non plus locale (une série chimique, une cible) de la conception de molécules actives. Ce n'est donc pas tant l'information contenue, mais celle susceptible d'être prédite pour de nouvelles molécules, qui est très importante dans ce type d'approche.

Le principal écueil de l'utilisation de ces bases de données est la difficulté pratique d'obtenir les molécules, qui, pour la plupart, ont été décrites dans des publications ou des brevets, mais ne sont pas disponibles commercialement.

Chimiothèques virtuelles

La disponibilité physique des molécules, bien que conseillée, n'est pas une condition absolue au criblage virtuel. Il est donc théoriquement possible d'imaginer n'importe quelle chimiothèque virtuelle composée de molécules jamais encore synthétisées, mais répondant à des propriétés d'intérêt. Nous faisons ici la distinction entre chimiothèque électronique (molécules disponibles dans un format moléculaire électronique) et chimiothèque virtuelle (molécules existant uniquement sous forme électronique mais non disponibles physiquement).

Parmi les principales chimiothèques virtuelles, il convient de citer la famille de bases de données GDB [17-19] parmi lesquelles on répertorie :

- la base GDB-11, composée de 24 millions de molécules nouvelles à partir d'un maximum de 11 atomes lourds (C, N, O, et F) ;
- la base GDB-13, de 370 millions de molécules « *drug-like* » à partir d'un maximum de 13 atomes lourds (C, N, O, S et Cl) ;
- la base GDB-17, de 166 milliards de molécules « *drug-like* » à partir d'un maximum de 17 atomes lourds (C, N, O, S, F, Cl, Br, I).

Toutes ces molécules sont susceptibles d'être stables chimiquement et ont été conçues à partir de règles simples d'élaboration de graphes moléculaires. Les chimiothèques sont interrogeables au moyen de requêtes structurales et contextuelles simples *(Figure 3)*.

L'atout majeur de ce type de chimiothèque est bien évidemment l'exploration d'un espace chimique bien supérieur à celui couvert par les chimiothèques réelles. La contrepartie est la nécessité de synthétiser les molécules sélectionnées lors du criblage, ce qui impose bien souvent l'élimination de molécules potentiellement intéressantes, mais dont les voies d'accès synthétiques sont complexes, longues et coûteuses. Le criblage virtuel de ces chimiothèques a néanmoins conduit à des touches confirmées expérimentalement mais souvent d'affinité modeste [20]. Ce paradoxe est identique à celui rencontré lors de criblage par la méthode de *design de novo* [21]. Quitte à obtenir des molécules

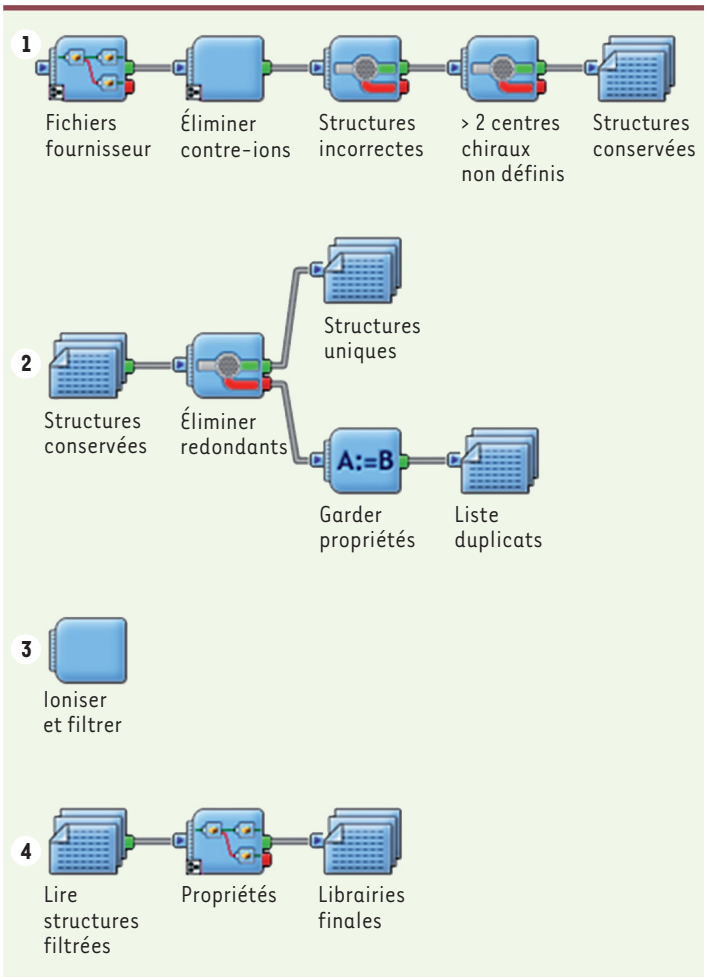

Figure 4. ***Exemple de protocole automatisé (Pipeline Pilot) de traitement de chimiothèques.***

d'affinité modeste (le plus souvent de l'ordre du micromolaire), vaut-il mieux cribler des molécules disponibles mais potentiellement peu originales, ou des molécules originales mais potentiellement difficiles d'accès ?

Traitement des chimiothèques pour le criblage virtuel

Une fois la ou les chimiothèques téléchargées, il est souvent indispensable d'en filtrer le contenu [22] afin de ne garder que les molécules réellement d'intérêt *(Figure 4)*.

Le niveau de filtrage dépend du projet de criblage. Il est néanmoins obligatoire de s'assurer qu'un certain nombre d'opérations de base ont bien été effectuées, notamment :

– l'élimination des espèces moléculaires annexes (ions, solvant) à la molécule bioactive principale ;

– l'élimination ou la correction des structures incorrectes (ex. : carbones pentavalents) ;

– la gestion des centres asymétriques (*voir Glossaire*) non définis de manière explicite. Il est courant de supprimer toute molécule possédant plus de deux centres chiraux non définis explicitement (il faudrait générer au moins huit stéréo-isomères pour chaque cas de figure) ;

– la gestion de la redondance des structures (plusieurs fournisseurs décrivant la même molécule) ;

– la gestion éventuelle des espèces ionisées les plus probables à pH physiologique (ex. : un acide carboxylique aliphatique sera stocké sous forme de carboxylate déprotoné) ;

– la gestion des formes tautomériques possibles (*voir glossaire*) ;

– l'élimination des molécules chimiquement réactives, potentiellement toxiques, ou susceptibles d'interférer avec un test de criblage biochimique (ex. : molécules à fort potentiel d'agrégation) ;

– le calcul de propriétés physicochimiques simples (poids moléculaire, nombre de donneurs/accepteurs de liaison hydrogène, surface polaire accessible au solvant, nombre de violations des règles de Lipinski [23], nombre de liaisons de rotation, nombre de cycles, etc.) qui permettront de faire des requêtes simples amenant à la sélection de sous-ensembles de la chimiothèque complète.

De nombreux logiciels permettent de réaliser ces opérations de manière automatisée mais souple, parmi lesquels PipelinePilot [24] pour les logiciels payants, certains logiciels ChemAxon [25] (gratuit pour le secteur académique uniquement) ou RDKit [26] implémentés dans Knime [27], et FAF-Drugs2 [22] ou ScreeningAssistant [28] pour les logiciels gratuits.

Les deux étapes importantes dans la préparation des molécules pour un criblage sont le choix de l'espace chimique à considérer et la sélection des molécules pertinentes dans cet espace. ScreeningAssistant (SA) [28] est un logiciel *open-source* [29] permettant l'analyse, la comparaison, la visualisation et la gestion complète et efficace de grands ensembles de molécules dédiées au criblage *(Figure 5)*. SA peut être considéré comme une plate-forme regroupant un ensemble de méthodes chémoinformatiques afin de répondre au mieux aux problématiques liées à l'analyse et à la comparaison de chimiothèques qui peuvent être virtuelles ou provenir de fournisseurs commerciaux *(Tableau I)*. Ce logiciel est donc utilisé en amont du criblage. Parmi les différentes tâches possibles et non exhaustives, SA permet de calculer des propriétés physicochimiques et des descripteurs moléculaires, de filtrer les structures redondantes, d'éliminer les molécules potentiellement problématiques, de prédire les caractères « *drug-like* » et « *lead-like* » des composés, et, enfin, de générer un sous-ensemble de composés divers. En effet, SA possède un module de diversité globale, qui permet à l'utilisateur de sélectionner un sous-ensemble de

Laboratoire	Code Unité	Responsable	Ville
GRIIOT	EA 4481	Philippe Chavatte	Lille
Laboratoire génomique, bioinformatique et applications	EA 4627	Matthieu Montes	CNAM Paris
Institut de biologie systémique et synthétique	FRE 3561	Jean-Loup Faulon	Évry
Centre de bioinformatique Mines ParisTech	U900	Jean-Philippe Vert	Paris
Institut Curie	UMR 176	Nicolas Saettel	Orsay
Centre d'études et de recherche sur le médicament de Normandie	UMR 3038	Ronan Bureau	Caen
Centre de biologie structurale	UMR 5048	William Bourguet	Montpellier
Institut de pharmacologie et de biologie structurale	UMR 5089	Laurent Maveyraud	Toulouse
Centre d'études d'agents pathogènes et biotechnologie pour la santé	UMR 5236	Laurent Chaloin	Montpellier
Laboratoire d'ingéniérie des systèmes biologiques et des procédés	UMR 5504	Magali Remaud-Siméon	Toulouse
Institut de chimie organique et analytique	UMR 6005	Pascal Bonnet	Orléans
Institut de pharmacologie moléculaire et cellulaire	UMR 6097	Dominique Douguet	Nice
Chimie et interdisciplinarité : synthèse, analyse, modélisation	UMR 6230	Jean-Yves Le Questel	Nantes
Laboratoire de signalisation et récepteurs matriciels	UMR 6237	Manuel Dauchez	Reims
Centre des science du goût de de l'alimentation	UMR 6265	Élisabeth Guichard	Dijon
Laboratoire de chémoinformatique	UMR 7140	Alexandre Varnek	Strasbourg
Chémogénomique structurale	UMR 7200	Didier Rognan	Strasbourg
Centre de recherche en cancérologie	UMR 7258	Xavier Morelli	Aix-Marseille
Institut de chimie moléculaire	UMR 7312	Jean-Marc Nuzillard	Reims
Laboratoire ICube	UMR 7357	Nicolas Lachiche	Strasbourg
Molécules thérapeutiques *in silico*	UMRS 973	Bruno Villoutreix	Paris

Tableau V. ***Ressources académiques dans le domaine du criblage virtuel.***

molécules les plus diverses à cribler parmi plusieurs librairies de fournisseurs commerciaux, et donc, *in fine*, à limiter l'achat de molécules ayant des châssis moléculaires proches. Outre les recherches simples effectuées à partir de critères de filtres sur les propriétés moléculaires, SA propose la recherche par nom, par structure exacte, par sous-structure, par similitudes et par sous-structures génériques (SMARTS). Afin de comparer visuellement des chimiothèques entre elles ou à des sous-espaces chimiques, des enveloppes convexes délimitant différents types de sous-espaces de références (HTS [*high throughput screening*], « *drug-like* », « *lead-like* », etc.) ont été calculées à partir d'une base de produits contenant plus de six millions de molécules. Un module innovant implémenté dans SA calcule l'enveloppe convexe (*voir Glossaire*) d'un espace chimique en utilisant la méthodologie des DRCS (*delimited reference chemical space*) [30]. Il est alors possible de calculer d'autres enveloppes convexes une fois l'espace créé, sur l'ensemble de la base ou sur les molécules associées à une librairie. Ces enveloppes permettent de visualiser rapidement l'espace chimique occupé par différentes chimiothèques ainsi que la

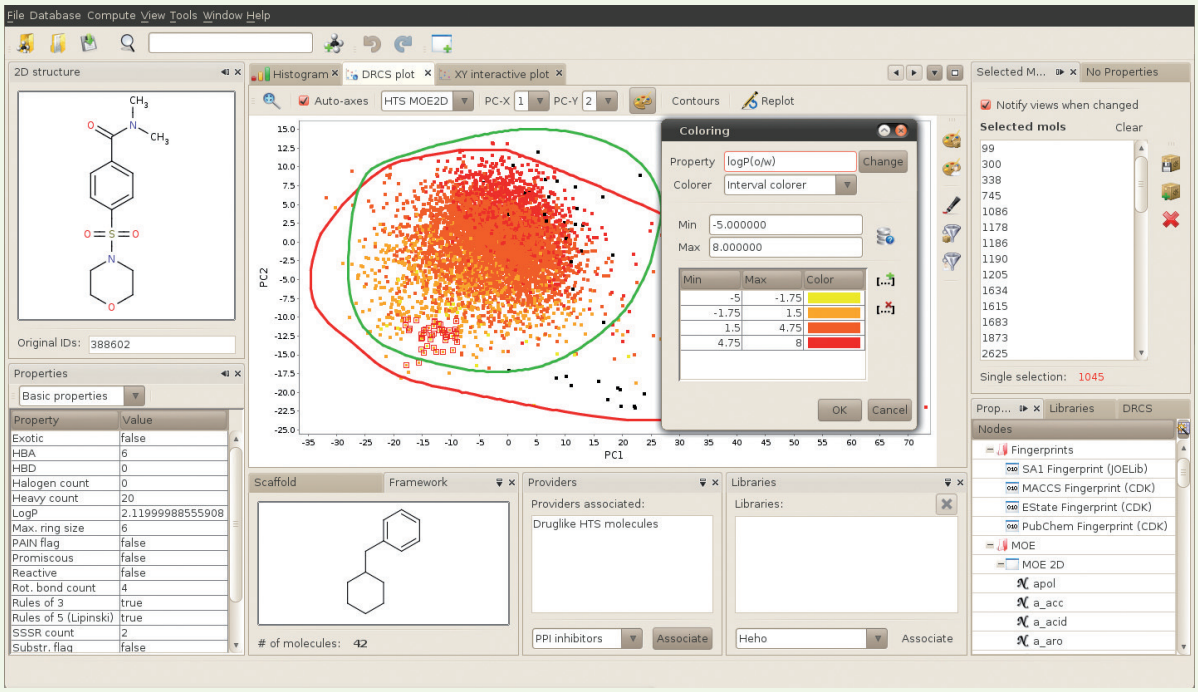

Figure 5. *Capture d'écran du logiciel ScreeningAssistant.*

GLOSSAIRE

Chimiothèque : collection de molécules (existantes ou non) dont les structures et éventuellement les propriétés sont archivées sous un format électronique. La plupart des chimiothèques couramment utilisées existent à la fois physiquement (molécules disponibles en poudre) et électroniquement.

Touche : molécule active dont certaines propriétés (physicochimiques, pharmacologiques, pharmacocinétiques) doivent être optimisées pour qu'elle devienne un candidat-médicament.

Contre-ion : ion accompagnant une espèce ionique (médicament) permettant de maintenir sa neutralité. Par example, l'ion chlorure Cl^- est le contre-ion de molécules synthétisées sous forme de chlorhydrates.

Tautomère : isomère structural d'une molécule, issu d'un réarrangement de la position d'un atome d'hydrogène et d'une liaison double adjacente.

Centre asymétrique/chiral : atome conférant une chiralité à la molécule correspondante, c'est-à-dire donnant lieu à deux molécules (stéréoisomères) images dans un miroir.

Enveloppe convexe : l'enveloppe convexe d'un objet ou d'un regroupement d'objets géométriques est l'ensemble le plus petit parmi ceux qui le contiennent.

densité d'occupation de cet espace. En outre, le module de visualisation permet de sélectionner les molécules de manière interactive et de les représenter sous la forme de tables de propriétés moléculaires (diversité, doublons, pourcentage de composés « *drug-like* », etc.), ou de les visionner individuellement sous forme structurale 2D. Il permet aussi une visualisation des données projetées dans des espaces chimiques sous la forme d'un nuage de points représenté en deux dimensions. Finalement, ce logiciel s'appuie sur un système de gestion de bases de données offrant la possibilité de créer et de maintenir à jour des chimiothèques de plusieurs millions de molécules uniques provenant de fournisseurs différents.

Le niveau de filtrage est un paramètre très important nécessitant à la fois pragmatisme et expérience. Dans le cas où le criblage est supposé fournir des sondes pharmacologiques, nous conseillons un filtre relativement léger afin d'éliminer principalement les molécules susceptibles d'interférer avec l'essai biochimique/pharmacologique de validation expérimentale (ex. : agrégants, composés fluorescents, réactifs chimiques) [31]. De même, la recherche des tout premiers inhibiteurs d'une cible jusque-là orpheline (de ligands) ne doit pas être handicapée par une élimination prématurée de molécules certes imparfaites mais présentant l'avantage de fournir des points de départ potentiels pour une pharmacomodulation. En revanche, l'identification de molécules répondant à un cahier des charges bien précis (ex. : fragments, profil polypharmacologique particulier, sélectivité fine vis-à-vis de sous-types du même récepteur, molécules brevetables) requiert un premier élagage plus discriminant de manière à réduire l'espace chimique à cribler. Une erreur typique

de débutant est de cribler le maximum de molécules de manière à augmenter le taux de touches. Dans ce domaine, la qualité et la diversité des molécules criblées priment toujours sur leur nombre.

Conclusions

De nombreuses chimiothèques sont disponibles à des fins de criblage virtuel mais répondent à des besoins divers. Le plus souvent, le point de départ reste une, voire plusieurs librairies commercialement disponibles qui permettront une validation expérimentale rapide du criblage par l'achat de touches d'intérêt. De nombreux portails internet regroupent plusieurs de ces catalogues et donnent ainsi accès à plusieurs millions de molécules dont la diversité ne cesse de s'accroître au fil des ans. Dans plusieurs pays se constituent, en appui des chimiothèques commerciales, des librairies de molécules issues de laboratoires académiques dont le rapport diversité/taille est particulièrement intéressant. Enfin, des collections de médicaments et de sondes pharmacologiques sont disponibles dans une optique de repositionnement.
Il est conseillé de confier la gestion pratique de ces chimiothèques à un chémoinformaticien qui saura, en collaboration avec des chimistes médicinaux, des biologistes et des pharmacologues, faire fructifier le projet de criblage initial. Un tableau *(Tableau V)* inventorie les principaux laboratoires académiques actifs dans le domaine du criblage virtuel, regroupés au sein de la Société française de chémoinformatique (http://*chidept.enscm.fr/SFCI/wordpress/)* ainsi que d'un groupement de recherche (GDR chémoinformatique, http://infochim.u-strasbg.fr/gdrchemoinfo/). ◊

SUMMARY

Chemical databases and virtual screening

A prerequisite to any virtual screening is the definition of compound libraries to be screened. As we describe here, various sources are available. The selection of the proper library is usually project-dependent but at least as important as the screening method itself. This review details the main compound libraries that are available for virtual screening and guide the reader to the best possible selection according to its needs. ◊

LIENS D'INTÉRÊT

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

RÉFÉRENCES

1. Vayer P, Arrault A, Lesur B, *et al.* Apports de la chémoinformatique dans la recherche et l'optimisation des molécules d'intérêt thérapeutique. *Med Sci (Paris)* 2009 ; 25 : 871-7.
2. Rognan D. Chemogenomic approaches to rational drug design. *Br J Pharmacol* 2007 ; 152 : 38-52.
3. Scior T, Bender A, Tresadern G, *et al.* Recognizing pitfalls in virtual screening : a critical review. *J Chem Inf Model* 2012 ; 52 : 867-81.
4. Krier M, Bret G, Rognan D. Assessing the scaffold diversity of screening libraries. *J Chem Inf Model* 2006 ; 46 : 512-4.
5. Clark DE, Pickett SD. Computational methods for the prediction of drug-likeness. *Drug Discov Today* 2000 ; 5 : 49-58.
6. Hann MM, Oprea TI. Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin Chem Biol* 2004 ; 8 : 255-63.
7. Congreve M, Chessari G, Tisi D, Woodhead AJ. Recent developments in fragment-based drug discovery. *J Med Chem* 2008 ; 51 : 3661-80.
8. Irwin JJ, Sterling T, Mysinger MM, *et al.* ZINC : a free tool to discover chemistry for biology. *J Chem Inf Model* 2012 ; 52 : 1757-68.
9. Chemaxon. Disponible à : http://www.chemaxon.com/marvin/help/applications/molconvert.html.
10. Hibert MF. French/European academic compound library initiative. *Drug Discov Today* 2009 ; 14 : 723-5.
11. Chimiothèque nationale. Disponible à : http://chimiotheque-nationale.org
12. Projet Eu_openscreen. Disponible à : http://www.eu-openscreen.de
13. Law V, Knox C, Djoumbou Y, *et al.* DrugBank 4.0 : shedding new light on drug metabolism. *Nucleic Acids Res* 2014 ; 42 : D1091-7.
14. Integrity. Disponible à : https ://integrity.thomson-pharma.com/integrity/xmlxsl
15. Prestwick Chemical. Disponible à : http://www.prestwickchemical.com/index.php?pa=26
16. Sigma-Aldrich. Disponible à : http://www.sigmaaldrich.com/chemistry/drug-discovery/validation-libraries/lopac1280-navigator.html
17. Ruddigkeit L, van Deursen R, Blum LC, Reymond JL. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J Chem Inf Model* 2012 ; 52 : 2864-75.
18. Blum LC, Reymond JL. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J Am Chem Soc* 2009 ; 131 : 8732-3.
19. Fink T, Reymond JL. Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F : assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *J Chem Inf Model* 2007 ; 47 : 342-53.
20. Blum LC, van Deursen R, Bertrand S, *et al.* Discovery of alpha7-nicotinic receptor ligands by virtual screening of the chemical universe database GDB-13. *J Chem Inf Model* 2011 ; 51 : 3105-12.
21. Schneider G. *De novo* design - hop(p)ing against hope. *Drug Discov Today Technol* 2013 ; 10 : e453-460.
22. Lagorce D, Maupetit J, Baell J, *et al.* The FAF-Drugs2 server : a multistep engine to prepare electronic chemical compound collections. *Bioinformatics* 2011 ; 27 : 2018-20.
23. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 2001 ; 46 : 3-26.
24. Pipeline Pilot, version 8.5. Accelrys, Inc. : San Diego, CA 92121.
25. Chemaxon. Disponible à : http://www.chemaxon.com
26. RDKit. Open-source cheminformatics. Available at : http://www.rdkit.org
27. Knime. The Konstanz Information Miner. Available at : http://www.knime.org/
28. Le Guilloux V, Arrault A, Colliandre L, *et al.* Mining collections of compounds with Screening Assistant 2. *J Cheminformatics* 2012 ; 4 : 20.
29. http://sa2.sourceforge.net
30. Le Guilloux V, Colliandre L, Bourg S, *et al.* Visual characterization and diversity quantification of chemical libraries : 1. creation of delimited reference chemical subspaces. *J Chem Inf Model* 2011 ; 51 : 1762-74.
31. Baell JA, Walters MA. Chemical con artists foil drug discovery. *Nature* 2014 ; 513 : 481-3.

TIRÉS À PART

D. Rognan