Mediachimie
Donner matière à l'avenir

OUTILS NUMÉRIQUES EN LIGNE POUR LA PHARMACOCHIMIE

Philippe Thomas

Partie des programmes associées

Programme d'enseignement scientifique de terminale : Thème 3 – Une histoire du vivant & Thème 3.5 – Intelligence artificielle

Programme de terminale STL spécialité : S33 – Cycle cellulaire cancer et cellules souches & L4 – Mobiliser les outils numériques en biotechnologie

Programme de CPGE : Chimie du et pour le vivant / La chimie dans les processus du vivant

Programme de DUT Chimie : Chimie du vivant

Mots-clés : outils numériques, pharmaceutique, médicaments, protéines, molécules chimiques, toxicité

INTRODUCTION

La quantité de données en ligne est trop importante pour être totalement exploitée manuellement par l'Homme. Des outils numériques (algorithmes et logiciels) ont été développés pour les manipuler : ils permettent de récupérer les données pour les utiliser et éventuellement générer des hypothèses pour concevoir de nouveaux médicaments. Cela passe, entre autres, par l'étude de données relatives aux petites molécules chimiques d'intérêt pharmaceutique et aux poches de liaisons de protéines. Il est également possible d'envisager le repositionnement de médicaments et de prédire la toxicité de molécules.

UTILISER DES OUTILS NUMÉRIQUES POUR EXPLOITER DES DONNÉES EXISTANTES

Banques de données de molécules connues

Des bases de données donnent accès aux propriétés biophysiques de petites molécules chimiques à portée thérapeutique (Figure 1). Dans la *Protein Data Bank* (PDB), on trouve les petites molécules d'un certain nombre de médicaments co-cristallisés avec la cible thérapeutique. Dans la *Cambridge Structural Database* ou dans la *Crystallography Open Database*, les structures 3D de plus d'un million de petites molécules chimiques sont

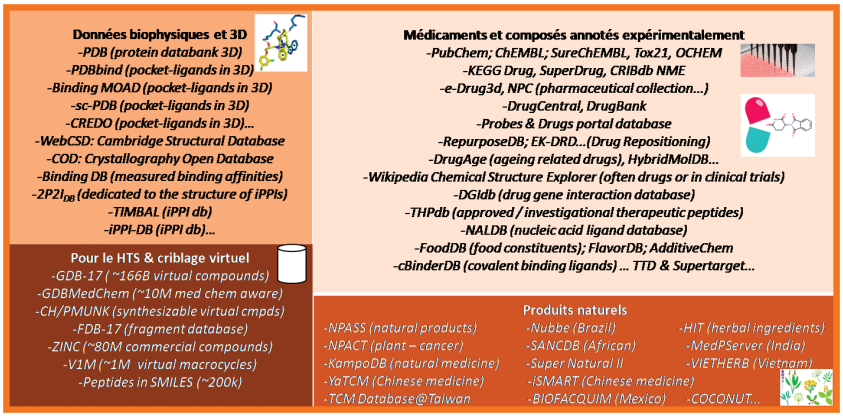

Figure 1 – De très nombreuses bases de données spécialisées et contenant des informations au format adapté sont disponibles sur Internet.

identifiées et étudiées par des approches biophysiques. La base de données ZINC répertorie à peu près 80 millions de petites molécules qui peuvent être achetées. PubChem et ChEMBL sont des données issues de deux grands projets aux États-Unis et en Europe sur lesquels on trouve environ 100 000 composés qui sont testés sur des dizaines de protéines. Le logiciel DataWarrior permet de récupérer directement de l'information sur des bases de données comme ChEMBL ou Wikipédia (Figure 2).

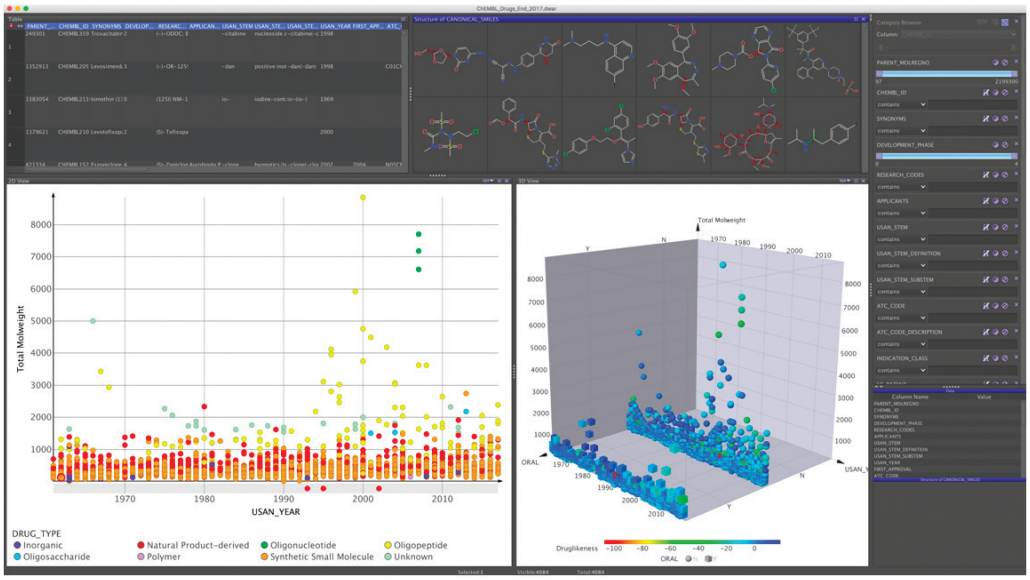

Figure 2 – Le logiciel de visualisation des données DataWarrior (www.openmolecules.org ; tutoriel : ww.radarweb.fr).

Algorithmes pour le ciblage de poches de liaisons de protéines de structures 3D connues

Ciblage par des molécules

Pour qu'une petite molécule chimique puisse cibler une protéine, il faut que celle-ci présente une poche de fixation. Des algorithmes, comme FTMap, permettent d'observer l'existence de poches lorsque l'on connait la structure 3D de la protéine. On peut alors identifier la plupart des sous-cavités impliquées dans l'amarrage de fragments chimiques ou de petites molécules. Lorsque l'on modélise la structure de l'interleukine par exemple, il est possible d'observer qu'elle possède des poches et des protubérances et de modéliser des potentielles molécules chimiques qui pourraient s'y accrocher (Figure 3). Lorsqu'il ne parait pas possible de fixer une molécule, il est possible d'utiliser des approches de simulations dynamiques, comme avec le serveur TRAPP (Allemagne) par exemple, pour observer différentes approches d'ouverture de poches.

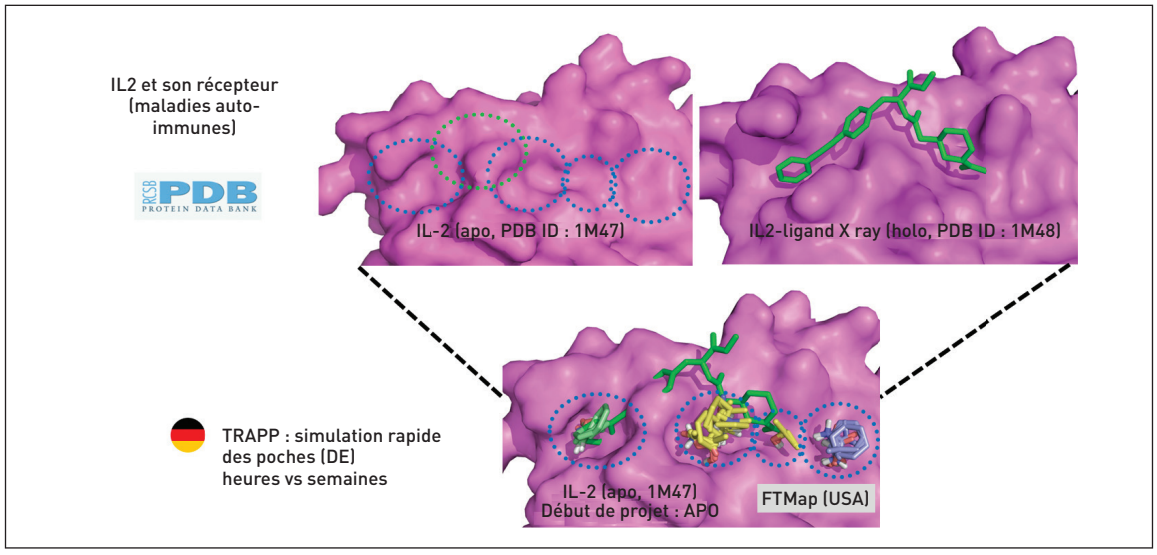

Figure 3 – À l'aide des données disponibles sur la Protein Data Base (PDB), il est possible de déterminer les propriétés de fixation d'une molécule fonctionnelle sur une protéine et d'effectuer des simulations dynamiques, avec le serveur TRAPP par exemple. Source : Arkin et coll. (2003). PNAS.

Mutations et substitutions qui se produisent dans les cibles

Par exemple, l'antithrombine (protéine anticoagulante), qui fixe le médicament fondaparinux, a été étudiée à la suite de cas de patients qui présentent des problèmes de coagulation : les localisations des mutations ponctuelles dans la structure 3D de la protéine peuvent être visualisées et récupérées dans la PDB (Figure 4A). On peut utiliser ce fichier PDB dans le serveur italien RING pour la visualisation en 2D des informations tridimensionnelles telles que des liaisons hydrogènes, des ponts salins (Figure 4B). On peut transformer cette visualisation un peu complexe en un autre type de visualisation avec le logiciel Cytoscape (Figure 4C). On peut aussi y ajouter les mutations de la protéine et voir l'impact de ces mutations sur la stabilité de la protéine et sur la fixation d'un certain nombre de médicaments.

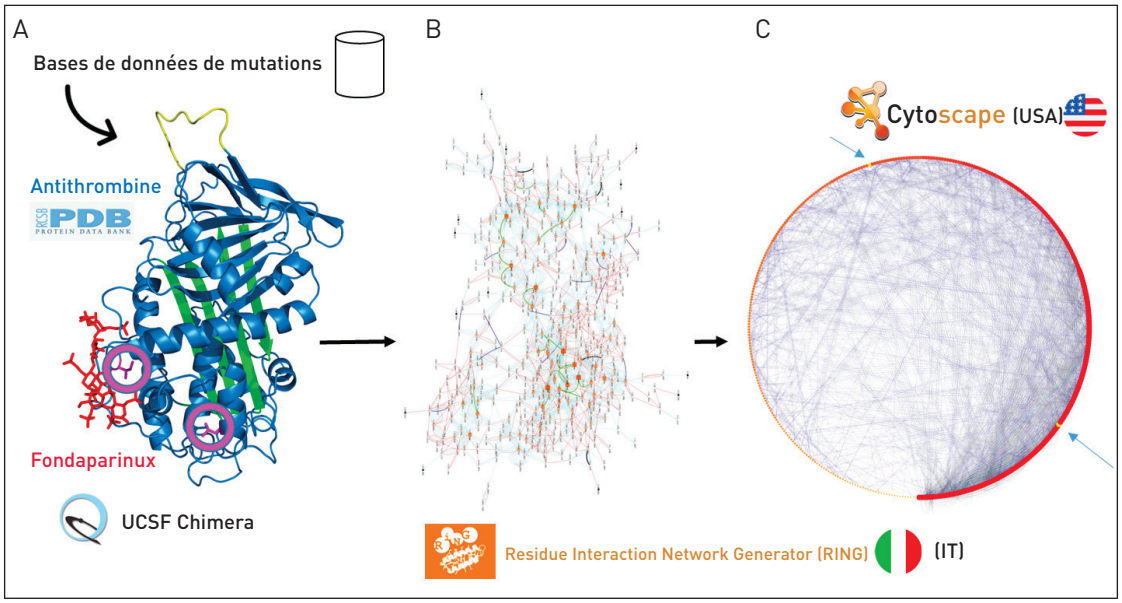

Figure 4 – Les outils de visualisation permettent de gagner en connaissance. A) Structure 3D de la protéine antithrombine avec fixation du médicament fondaparinux. B) Visualisation du système après traitement par le logiciel RING pour caractériser les interactions non covalentes. C) Traitement par le logiciel cytoscape pour caractériser la force des interactions non covalentes.
Source : Dinarvand et coll. (2018). J. Thromb Haemost.

Banques de données et serveurs pour le repositionnement de médicaments existants

L'objectif est de repositionner des médicaments qui existent déjà sur de nouvelles cibles thérapeutiques et de réduire le temps de développement. Les approches pour ce repositionnement sont basées sur les signatures transcriptomiques, les connaissances des ligands et la connaissance tridimensionnelle des cibles.

Une première approche peut être l'approche vectorielle (criblage virtuel basé sur la connaissance de ligands) : la molécule à tester est transformée en un vecteur de 1 et 0. Chacune des molécules à tester est transformée en une suite de 1 et 0. Cela permet d'identifier dans la base de données ChEMBL ou PubChem les composés qui sont proches et donc ainsi espérer toucher d'autres protéines cibles. Une autre approche est l'approche géométrique : on étudie l'amarrage de la petite molécule chimique sur chacune des cibles pour obtenir un score prédit d'affinité. Le serveur MTiOpenScreen est dédié à l'amarrage des petites molécules et au criblage virtuel utilisant des bases de données de médicaments, des bases de données de produits alimentaires et des produits naturels (Figure 5). MTiOpenScreen a, par exemple, permis d'identifier des molécules qui bloquent l'activité d'une protéine impliquée dans l'angiogenèse de certains cancers (Figure 6).

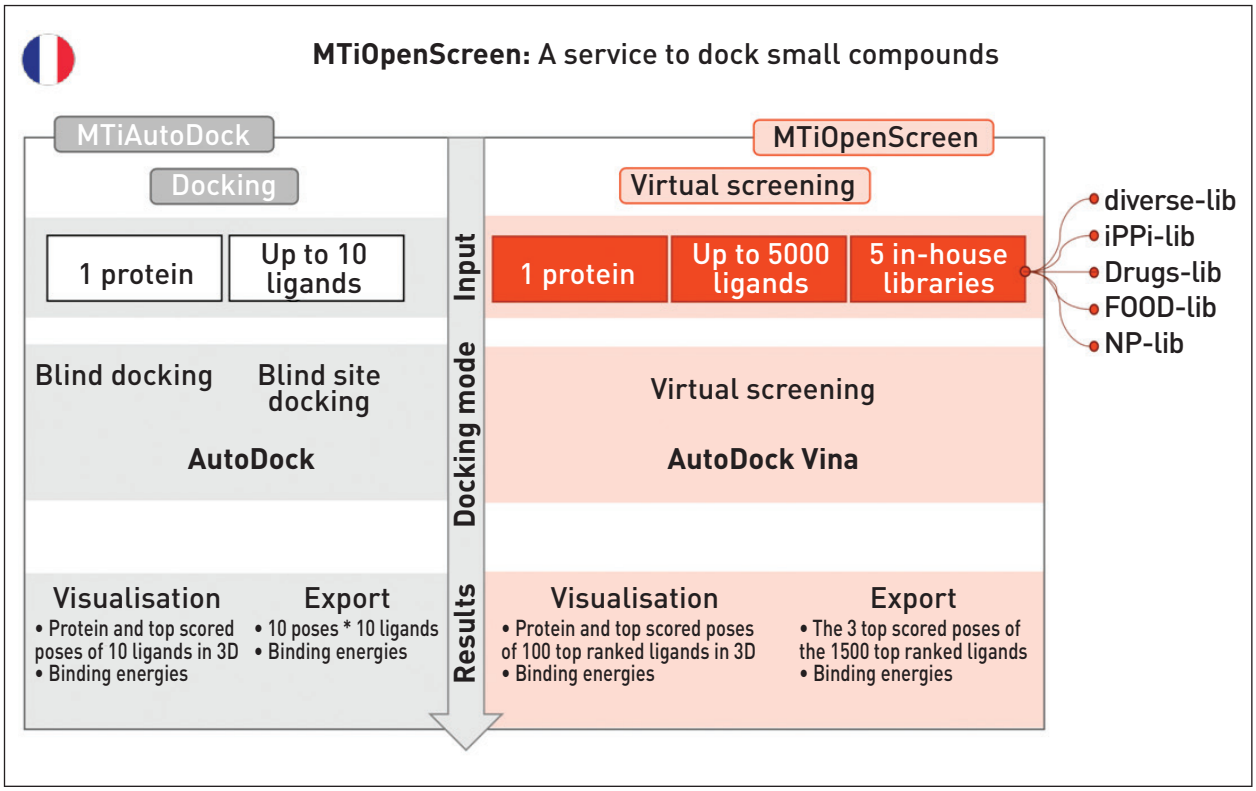

Figure 5 – MTiOpenScreen (http://drugmod.rpbs.univ-paris-diderot.fr/index.ph) est un serveur de traitement des données protéiques pour le criblage.

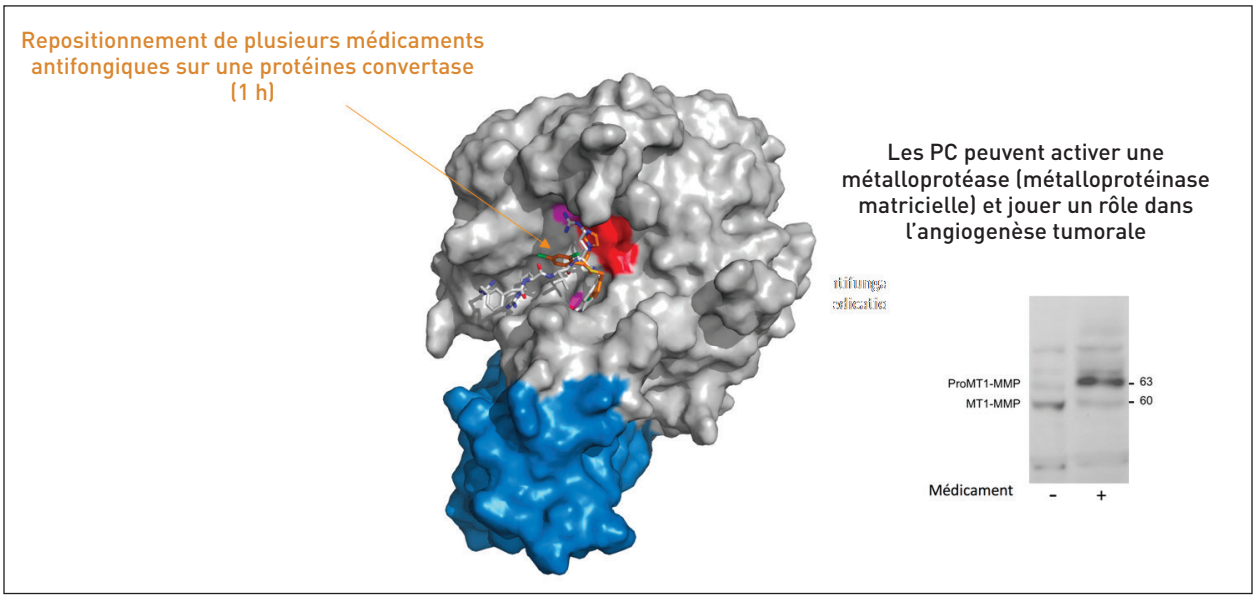

Figure 6 – Utilisation de MTiOpenscreen. Grâce à ce serveur, il a été permis d'identifier des molécules qui bloquent l'activité d'une protéine impliquée dans l'angiogenèse de certains cancers.

UTILISER DES OUTILS NUMÉRIQUES POUR FAIRE DES PRÉDICTIONS

Banques de données de molécules virtuelles

Le criblage virtuel peut même utiliser des molécules qui sont virtuelles, c'est-à-dire qui n'ont pas encore été synthétisées. C'est le cas de la base de données suisse GDB-17 qui contient 166 milliards de molécules qui n'ont pas été encore synthétisées (pour la plupart parce que c'est quasiment impossible).

Algorithmes pour le ciblage de poches de liaisons de protéines de structures 3D non connues

Il existe des algorithmes spécialisés pour prévoir des poches à la surface des protéines ou des macromolécules (Figure 7). Ces algorithmes prévisionnels sont basés soit sur la géométrie de la molécule – c'est le cas des sites indiens fpocket et PocketDepth, soit sur des calculs d'énergie. En effet lorsque l'on bombarde la surface de la protéine ou de la macromolécule qui est potentiellement impliquée dans une pathologie, on peut construire des cartes d'affinité d'interaction, car dans certaines zones appelées « hotspots » des petits fragments de molécules ou des atomes s'accrochent préférentiellement. Ces zones sont utilisées ensuite pour faire du design de molécule capable de s'y accrocher.

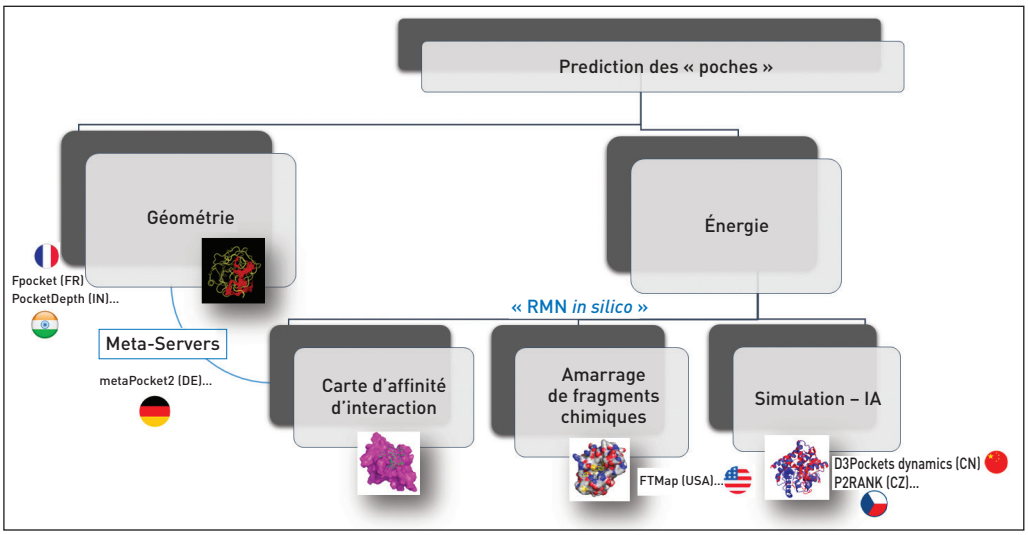

Figure 7 – Deux grands types d'approche pour la prédiction des « poches » existent : les approches géométriques et les approches énergétiques. Elles permettent ainsi d'obtenir des cartes et d'effectuer des simulations.

Intelligence artificielle et apprentissage automatique

L'utilisation de l'intelligence artificielle pour prédire est encore peu développée. L'intelligence artificielle d'un point de vue général permettra de faciliter le traitement du langage, la vision artificielle, la représentation des connaissances, mais dans le domaine du développement thérapeutique, on développe plutôt l'apprentissage automatique. Toute une série d'algorithmes permet de classer ou de construire des modèles statistiques à partir des données, notamment d'apprendre par exemple à partir de données existantes, à prédire si un composé peut être potentiellement toxique, sans expérience au laboratoire et éventuellement sans tester sur des modèles animaux. Les données sont récupérées à partir d'une chimiothèque comme ChEMBL. Les molécules sont insérées et annotées dans le système qui nettoie les données et calcule toute une série de descripteurs par rapport aux petites molécules chimiques, comme par exemple leur toxicité. Le système envoie ensuite ces informations à différentes approches statistiques pour créer des modèles. La qualité, et la performance des modèles est visualisée par des courbes qui sont récupérées par l'utilisateur. Le logiciel et le modèle statistique mathématique qui ont été généré par le système de manière automatique, peuvent ensuite être utilisés et appliqués à d'autres composés (Figure 8).

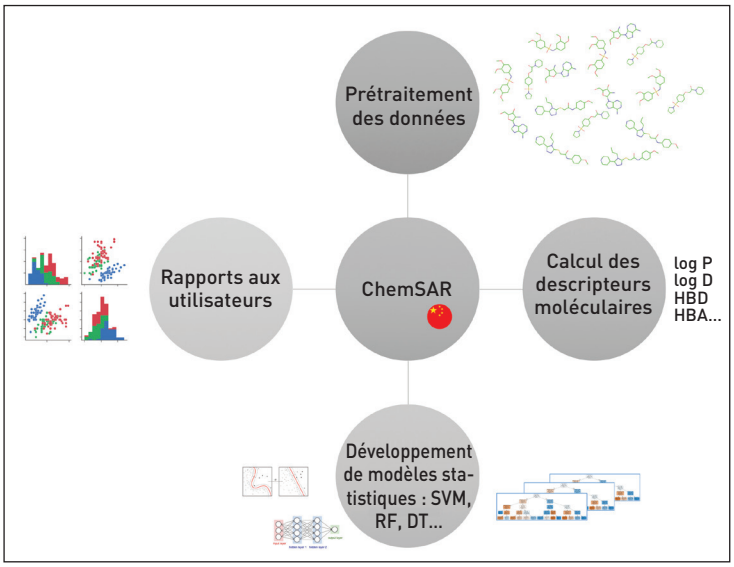

Figure 8 – Des systèmes de gestion des données permettent un traitement automatisé des informations récoltées sur les bases de données : modélisation, visualisation, performance.
Source: Dong et coll. (2017). J. Cheminform.

Prédiction de la toxicité

Dans le cadre du développement thérapeutique, les composés doivent être absorbés, distribués, métabolisés, excrétés. ADME-Tox permet de prévoir la toxicité d'une molécule tout au long de ce parcours (Figure 9). L'administration par voie orale d'un médicament fait intervenir le passage à travers toute une série de membranes biologiques. Il y a donc de nombreuses interactions avec des protéines et il faut essayer de reproduire un certain nombre de ces processus sur informatique. La première étape est le nettoyage des chimiothèques, qui s'appelle un filtrage ADME-Tox. Le logiciel FAF-Drugs permet d'introduire les molécules dans le système : les fichiers sont soumis à toute une série de calculs et de filtres à l'issue desquels elles sont soit rejetées comme potentiellement toxiques, soit acceptées pour continuer le développement thérapeutique (Figure 10).

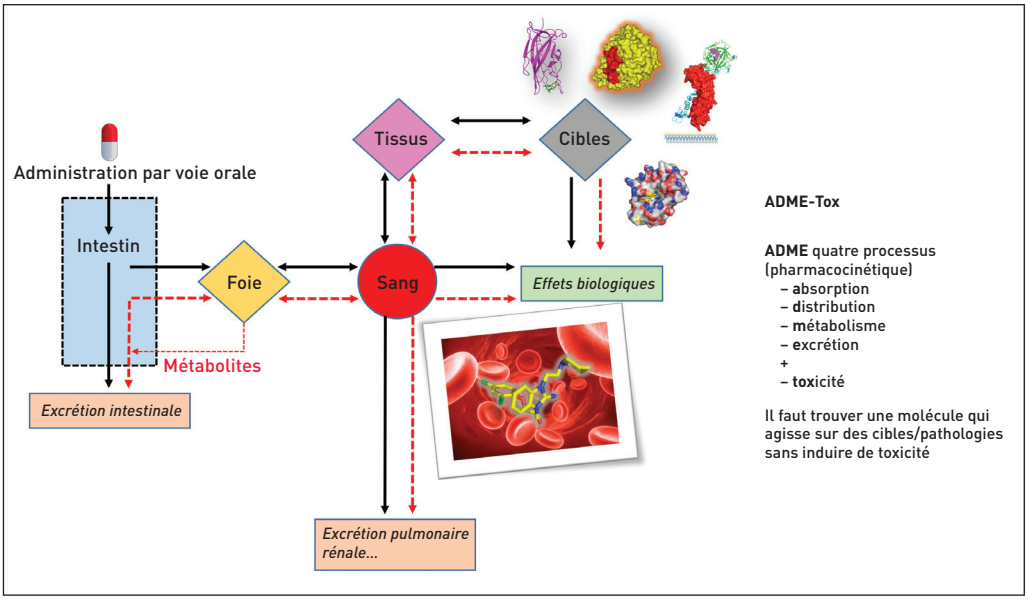

Figure 9 – ADME-Tox permet de suivre du point de vue toxicologique une molécule sur tout un système biologique au cours de son processus cinétique de métabolisation. Source : B. Testa, 2006.

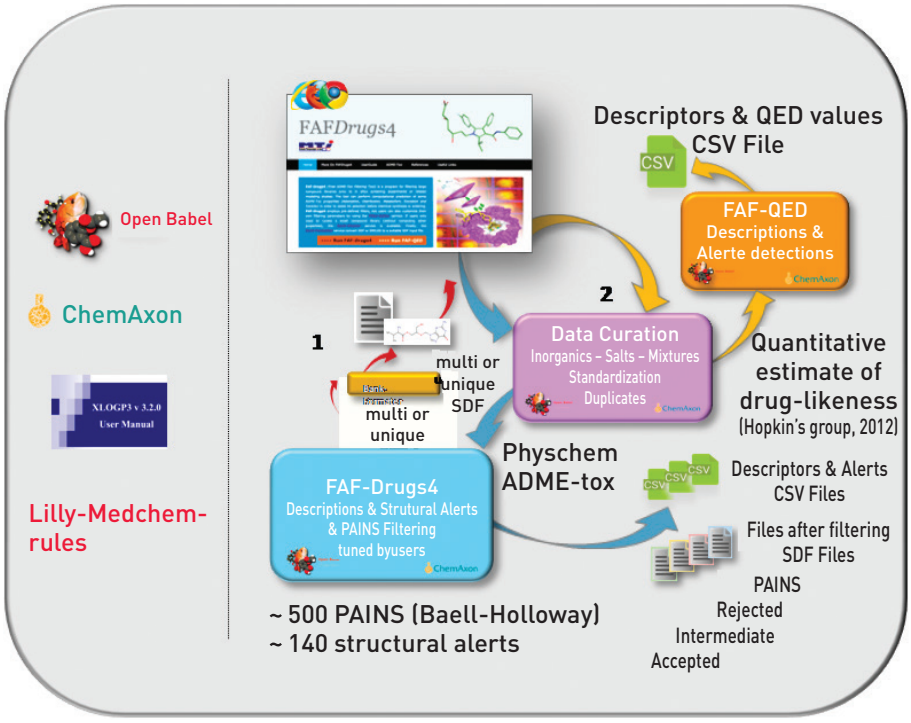

Figure 10 – FAF-Drugs est un logiciel en ligne accessible gratuitement qui permet, grâce à un traitement de données, de visualiser les caractéristiques toxicologiques du produit, notamment selon les critères de plusieurs groupes pharmaceutiques.

CONCLUSION

Le déploiement de nombreux outils numériques résulte de collaborations à l'international et de la compréhension des principes physiques auxquels répondent les processus du vivant. Le scientifique d'aujourd'hui et de demain est connecté avec ses pairs et utilise des outils numériques, en français et en anglais, qui lui permettent de canaliser le flux d'informations disponible. Ces outils sont par ailleurs utilisés pour formuler des hypothèses et prédictions pour l'identification de médicaments : cela permet de réaliser moins de tests en laboratoire (gain de temps et d'argent) et d'envisager des approches thérapeutiques avec des molécules non disponibles sur le marché.

SOURCES PRINCIPALES

Chimie et nouvelles thérapies, EDP Sciences, 2020, ISBN 978-2-7598-2469-4, « Recherche de sondes pharmacologiques et candidats-médicaments dans le cyber-espace » par Bruno Villoutreix.

Philippe Thomas est ingénieur ENSCP Chimie ParisTech

Comité éditorial : Danièle Olivier, Jean-Claude Bernier, Grégory Syoen