

## 19

## Méthodes d'études *in silico*

Le développement d'outils informatiques fiables couplé à la croissance de la puissance informatique a permis la mise en place de techniques de simulation numérique centrées sur la biologie. Par analogie avec les expressions *in vivo* et *in vitro*, le terme « *in silico* » a été introduit pour qualifier les méthodes numériques mises en œuvre pour traiter de tels systèmes. De par son nom, ce terme fait référence au silicium, matériau principal retrouvé dans les puces informatiques de tous les ordinateurs. Le champ *in silico* regroupe un très large ensemble de méthodes numériques fondées sur les lois de la physique et de la chimie qui, utilisant les approches des mathématiques, permettent de simuler ou de modéliser un phénomène biologique à l'aide de l'outil informatique. Dans le cadre de cette expertise, deux grandes familles de méthodes seront définies. La première, plus communément connue sous l'acronyme anglais QSAR (*Quantitative Structure Activity Relationship*) se base sur les relations quantitatives entre structures et propriétés, tandis que la seconde se réfère directement aux propriétés atomistiques de l'entité étudiée. Ces deux familles de méthodes brièvement décrites par la suite sont fréquemment employées dans les études des interactions comme celles rencontrées entre les perturbateurs endocriniens et leurs récepteurs.

Les méthodes *in silico* sont complémentaires des études *in vivo* et *in vitro* sur les perturbateurs endocriniens et ne peuvent remplacer l'expérience. Elles apportent des informations sur la nature des interactions chimiques entre les perturbateurs et les récepteurs et sont nécessaires à la bonne compréhension de ces mêmes mécanismes. Par la suite, de telles études peuvent être utilisées dans le développement de nouvelles expériences ou lors de la sélection de molécules à tester. En d'autres termes, l'utilisation combinée de la théorie et de l'expérience engendre une meilleure compréhension des phénomènes traités dans un domaine de recherche.

### QSAR

Le principe des méthodes QSAR (*Quantitative Structure Activity Relationship*) consiste comme leurs noms l'indiquent à mettre en place une relation mathématique à l'aide de méthodes d'analyse de données, reliant des propriétés moléculaires microscopiques appelées descripteurs, à un effet expérimental (activité biologique, toxicité, affinité pour un récepteur), pour une série de

composés chimiques similaires. Le point de départ de telles méthodes se construit sur la définition des descripteurs moléculaires empiriques ou théoriques. Ces dernières prennent en compte des informations sur la structure et les caractéristiques physico-chimiques des molécules, comme dans ce cas présent, les perturbateurs endocriniens.

Les descripteurs théoriques se déclinent eux-mêmes en sous-classes de descripteurs : les constitutionnels, basés sur la composition chimique de la molécule, les topologiques, obtenus à partir de la structure bi-dimensionnelle (table de connectivité des atomes de la molécule), les géométriques, évalués à partir de la structure tri-dimensionnelle et les quantiques, issus de la structure électronique de la molécule. Les détails concernant les descripteurs quantiques sont disponibles dans l'ouvrage de Karelson (2010).

Le choix de la base de données expérimentales de référence est décisif dans une étude QSAR. Elle doit être composée de données expérimentales fiables obtenues en suivant un protocole expérimental unique. En effet, la robustesse du modèle dépend fortement de la base sur lequel il se fonde. Néanmoins, malgré tous les efforts mis en œuvre pour recueillir des données homogènes, une certaine incertitude ne peut être évitée, notamment pour les systèmes biologiques.

Enfin, le lien entre les descripteurs et la base de données est déterminé grâce à des outils d'analyse comme les régressions multi-linéaires (MLR), les régressions aux moindres carrés partiels (PLS), les arbres de décisions, les réseaux de neurones, et les algorithmes génétiques.

En pratique, le développement d'un modèle débute par la collecte de données expérimentales fiables et en nombre conséquent. Cette étape est suivie par le développement d'une série de descripteurs qui caractérisent les structures géométriques et électroniques des composés étudiés en vue de les relier à la propriété expérimentale étudiée. Des outils d'analyse de données sont alors employés pour orienter le choix des descripteurs adéquats et mettre en place le modèle à proprement dit.

Le modèle une fois développé, sa corrélation doit ensuite être validée sur le jeu d'entraînement. Sa robustesse, c'est-à-dire l'influence des composés du jeu d'entraînement sur le modèle, est estimée par des méthodes de validation interne. Afin d'estimer son pouvoir prédictif, des données expérimentales supplémentaires sont nécessaires afin de déterminer la capacité du modèle à prédire ces valeurs. Enfin, il est important de savoir quel type de molécule s'utilise avec quel modèle. On parle alors de domaine d'applicabilité.

La relation QSAR mise en place peut alors être employée pour la prédiction de propriétés d'un jeu de nouvelles molécules existantes physiquement ou non, pour lesquelles les valeurs expérimentales ne sont pas disponibles. Une analyse fine des descripteurs intervenant dans le modèle QSAR donne également des informations sur la nature des mécanismes et des phénomènes

moléculaires (interactions électrostatiques ou hydrophobes) mis en jeu dans la propriété d'intérêt.

La littérature relate bon nombre d'études concernant les perturbateurs endocriniens, ainsi que les modèles robustes et fiables développés et validés sur de vastes ensembles de molécules. Dans la plupart des cas, les propriétés considérées sont relatives à des mesures *in vitro* comme l'affinité de liaison au récepteur, absolue (*binding affinity*, BA) ou relative au ligand naturel E2 (*relative binding affinity*, RBA). Les propriétés issues d'études *in vivo* ont rarement été considérées.

## Méthodes de modélisation à l'échelle atomique

Si les modèles QSAR donnent des informations sur la nature des interactions entre ligands et récepteurs (électrostatiques ou hydrophobes), ils ne permettent pas de quantifier la force de liaison ni les atomes impliqués dans cette même liaison. Néanmoins, si la structure de la paire ligand/récepteur est connue (par exemple œstradiol/ER), des informations sur l'interaction peuvent être déduites par analogie. En outre, les méthodes de modélisation moléculaire permettent de déterminer la structure tri-dimensionnelle, l'énergie ainsi que d'autres propriétés chimico-physiques des molécules et des systèmes biologiques à partir de leur composition atomique.

Différentes approches de modélisation sont envisageables. Si celles concernant la mécanique moléculaire sont économiques en termes de temps de calcul et permettent de traiter des systèmes moléculaires de grande taille, les méthodes quantiques (*ab initio*, semi-empiriques, ou théorie de la fonctionnelle de la densité) sont quant à elles capables de calculer les propriétés électroniques des systèmes.

La mécanique moléculaire (MM) est fondée sur des calculs de mécanique classique qui permettent de calculer l'énergie stérique du système à partir des forces entre les atomes constituant la biomolécule. Les liaisons chimiques sont alors modélisées par des ressorts suivant l'approximation harmonique. Ces modèles sont paramétrés avec des données expérimentales (résolution de structure des protéines par les rayons X, spectre IR) ou théoriques. Malgré leur nature empirique, les approches de MM peuvent donner des informations précieuses sur la structure tri-dimensionnelle des récepteurs, sur leurs conformations et sur les interactions chimiques (électrostatiques, van der Waals, liaison hydrogène) avec le ligand. Si l'ensemble des paramètres constituant le modèle de MM (défini comme champs de force) est fiable, les résultats seront semi-quantitatifs.

La MM est à la base de méthodes de « *docking* » moléculaires, méthodes spécifiquement développées pour étudier les interactions entre une protéine et un substrat, ainsi qu'à la base de criblage *in silico*.

Les méthodes fondées sur la mécanique quantique (MQ) s'appuient sur la résolution approchée de l'équation de Schrödinger. Elles permettent la détermination de la configuration électronique de la molécule, sa structure tridimensionnelle, son énergie et ses propriétés chimico-physiques. Le plus souvent, ces propriétés comme par exemple, l'énergie d'interaction entre le substrat et le récepteur, sont déterminées de façon quantitative avec une précision qui avoisine celle de l'expérience. En revanche, les méthodes de MQ permettent seulement de traiter des systèmes dont la taille s'élève à la centaine d'atomes, taille approximative du site d'interaction avec le récepteur.

Le choix du type de calcul dépendra donc du problème étudié (taille et précision souhaitée pour le calcul) et évidemment des ressources de calcul (puissance de calcul) mises à disposition de l'utilisateur.

Dans les deux cas, que ce soit pour la MM ou la MQ, une structure tridimensionnelle du récepteur et du ligand est un pré-requis indispensable. Cette dernière peut être obtenue par la spectroscopie à rayon X, dans le cas où la cristallisation du récepteur est possible (structure en milieu solide, sans ou avec le substrat). De manière alternative, des indications sur la structure peuvent être données par la spectroscopie RMN (structure en solution) ou par modélisation par homologie. Ce dernier modèle implique la reconstruction de la structure tri-dimensionnelle à partir d'un alignement de séquence avec un autre système pour lequel la structure est connue. Les méthodes de MM permettent également d'identifier la structure tri-dimensionnelle d'une protéine à partir de la seule séquence en acides aminés, mais la présence de plusieurs conformations énergiquement stables peut compliquer la recherche.

## BIBLIOGRAPHIE

CRAMER CJ. Essentials of computational chemistry: theories and models. Wiley, 2<sup>e</sup> édition, 2004

DEVILLERS J, MARCHAND-GENESTE N, CARPY A, PORCHER JM. SAR and QSAR modeling of endocrine disruptors. *SAR and QSAR Environmental research* 2006, **17** : 393-412

GAO H, KATZENELLENBOGEN JA, GARG R, HANSCH C. Comparative QSAR analysis of estrogen receptor ligands. *Chem Rev* 1999, **99** : 723-744

JACOBS MN. In silico tools to aid risk assessment of endocrine disrupting chemicals. *Toxicology* 2004, **205** : 43-53

KARELSON M. Molecular Descriptors in Qsar/Qspr. John Wiley & Sons Inc, New York 2000

LEACH AR. Molecular modelling: principles and applications. Prentice Hall, 2<sup>e</sup> édition, 2001

LI J, GRAMATICA P. The importance of molecular structures, endpoints' values, and predictivity parameters in QSAR research: QSAR analysis of a series of estrogen receptor binders. *Mol Divers* 2010 **14** : 687-696

LIU H, PAPA E, GRAMATICA P. QSAR Prediction of estrogen activity for a large set of diverse chemicals under the guidance of OECD principles. *Chem Res Toxicol* 2006, **19** : 1540-1548

LIU H, PAPA E, WALKER JD, GRAMATICA P. In silico screening of estrogen-like chemicals based on different nonlinear classification models. *Journal of Molecular Graphics and Modelling* 2007, **26** : 135-144

RONCAGLIONI A, BENFENATI E. In silico-aided prediction of biological properties of chemicals: oestrogen receptor-mediated effects. *Chem Soc Rev* 2008, **37** : 441-450

SHI LM, FANG H, TONG W, WU J, PERKINS R, et coll. QSAR Models using a large diverse set of estrogens *J Chem Inf Comput Sci* 2001, **41** : 186-195

TODESCHINI R, CONSONNI V. Handbook of molecular descriptors. In : MANNHOLD R, KUBINYI H, TIMMERMAN H (eds). WILEY VCH, Weinheim, 2000

ZHU BT, HAN GZ, SHIM JY, WEN Y, JIANG XR. Quantitative structure-activity relationship of various endogenous estrogen metabolites for human estrogen receptor  $\alpha$  and  $\beta$  subtypes: insights into the structural determinants favoring a differential subtype binding. *Endocrinology* **147** : 4132-4150