

La modélisation mathématique pour l'extrapolation à l'Homme en évaluation des risques

Depuis quelques années, les investigations toxicologiques se sont tournées vers une approche fondée sur des preuves, plutôt que descriptive, notamment pour contrer les coûts élevés, la durée des études *in vivo*, ainsi que les strictes réglementations concernant le bien-être animal (Beilmann, 2018). Pour des raisons éthiques, les tests de toxicité ne peuvent être réalisés directement chez les humains. Différents outils *in vitro*, *in vivo* et *in silico* ainsi que des protocoles d'expérimentations animales ont donc dû être développés afin d'étudier indirectement la toxicité humaine des composés chimiques qui nous entourent. Cette partie a pour objectif de présenter les modèles pharmacocinétiques et pharmacodynamiques utilisés en évaluation des risques pour étudier la toxicité humaine.

L'évaluation des composés repose sur des modèles de devenir, d'exposition, de toxicité, de risque et d'impact. Pour la gestion des risques, le développement conjoint de ces modèles, visant à améliorer l'interprétation et l'analyse des données, permet d'apporter un soutien scientifique pour la prise de décisions réglementaires (McCarty et al., 2018). Un modèle mathématique pertinent est une simplification conceptuellement plausible d'un processus ou d'une activité basée sur des données expérimentales. Ces modèles sont construits de sorte à représenter le degré minimum de complexité des processus physiques, chimiques et biologiques nécessaires pour quantifier la toxicité d'un composé et atteindre le pouvoir prédictif requis selon le contexte décisionnel (Margiotta-Casaluci et al., 2016). Les modèles *in silico* présentent un certain nombre de limites freinant leur acceptation comme outils indispensable pour

l'évaluation de risques. En effet, leur validation et leur développement reposent sur des données dont la disponibilité, la quantité et la qualité ne sont pas toujours de mise. De plus, leur domaine de validité est limité et défini approximativement. Finalement, la limite principale concerne le fondement prédictif et suggestif, rendant les conclusions hypothétiques. Cependant, ils offrent de nombreux avantages comparés aux tests *in vivo* et *in vitro*. Ils sont moins onéreux, hautement reproductibles et sans cesse optimisés, avec pour ambition de remplacer l'utilisation d'animaux (Erhirhie et al., 2018; Grech et al., 2017a; Leist et al., 2017).

1.1. La modélisation pharmacocinétique

La pharmacocinétique est définie par l'action d'un organisme sur un xénobiotique dont le devenir est régi par quatre processus ; l'absorption, la distribution, le métabolisme et l'excrétion (ADME). Elle permet de décrire l'évolution des concentrations d'un composé en fonction du temps selon la dose d'exposition (Grech et al., 2017a). Différents paramètres pharmacocinétiques, comme la biodisponibilité, le volume de distribution, la clairance et le temps de demi-vie, qui caractérisent la pharmacocinétique d'un xénobiotique donné, peuvent être déterminés par différentes approches. La première est l'analyse non-compartmentale permettant de déterminer les paramètres pharmacocinétiques à partir de points expérimentaux, tels que la concentration et le temps maximal, ou d'équations mathématiques simples afin de décrire les processus ADME (Gabrielsson and Weiner, 2012). La deuxième approche, dont les principes sont détaillés ci-après, modélise un organisme ou un système d'étude par un ou plusieurs compartiments, afin de quantifier les processus biologiques par une approche semi-mécaniste. Une signification physiologique et anatomique peut également être considérée pour ajouter une composante mécaniste au modèle (Dixit et al., 2003; Jusko, 2013). Grâce à la modélisation, de nouveaux paramètres pharmacocinétiques peuvent être définis et utilisés pour élaborer les relations entre le profil cinétique d'un composé et la physiologie. Par exemple, les modèles compartimentaux permettent de définir la clairance spécifique de différents processus d'élimination (Urso et al., 2002).

1.1.1. Les modèles compartimentaux classiques

Les modèles pharmacocinétiques classiques suivent une approche compartimentale considérant que le système d'étude, par exemple un organisme entier ou des cellules en culture, est réduit à un ou plusieurs compartiments pour décrire la cinétique d'un composé, comme présenté par la Figure 1.

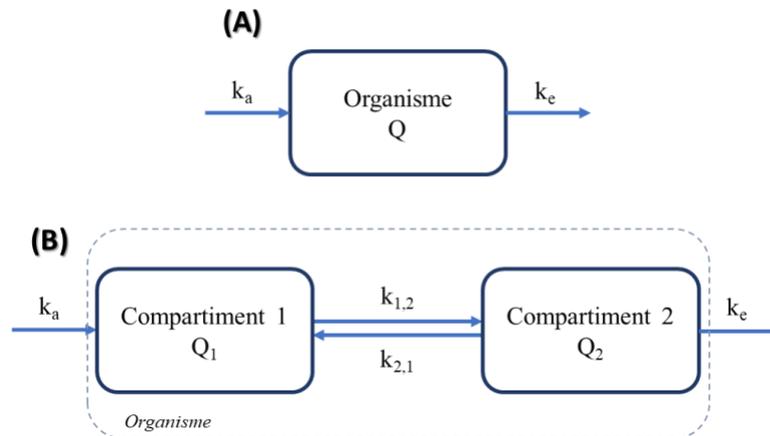


Figure 1. Schéma d'un modèle pharmacocinétique mono (A) et bi-compartmental (B), avec Q , Q_1 et Q_2 les quantités du composé dans les compartiments, k_a et k_e respectivement les constantes de vitesse d'absorption et d'élimination, $k_{1,2}$ et $k_{2,1}$ les constantes de vitesse de transfert.

Les modèles mono-compartmentaux considèrent que la concentration est homogène au sein du système. Les modèles multi-compartmentaux supposent que les concentrations peuvent différer au sein de différents compartiments caractérisant le système (Grech et al., 2017a).

Les modèles pharmacocinétiques sont définis par un système d'équations différentielles typiquement linéaire :

$$\frac{dC_i}{dt} = \sum_i (k_i \times C_i^{n_i}) \quad (1)$$

Avec dC_i/dt la variation de concentration dans le compartiment i au cours du temps, k_i une constante de vitesse et n_i un exposant déterminant l'ordre de la cinétique.

Un processus d'ordre zéro ($n = 0$) peut être dû à des phénomènes de saturation métabolique ou des systèmes de transport. La variation de concentration n'est pas dépendante de la quantité en composé restante. Un processus de premier ordre ($n = 1$) spécifie un modèle

cinétique linéaire avec des concentrations de composé dans les compartiments proportionnelles à la dose d'exposition. A faibles doses d'exposition, la majorité des composés étudiés en santé – environnement suivent approximativement cette cinétique, caractérisée par une variation de concentration proportionnelle à la quantité en composé restante et des paramètres pharmacocinétiques indépendants de la dose d'exposition (Borowy and Ashurst, 2020). Ainsi, la plupart du temps précédés d'une analyse non-compartimentale des paramètres cinétiques, les modèles compartimentaux de premier ordre utilisés en évaluation des risques permettent de quantifier les caractéristiques ADME des composés dont les processus toxiques restent à établir, suivant différentes voies d'exposition (Mackay and Fraser, 2000; Scholz et al., 2013a; Shin et al., 2019).

1.1.2. Les modèles physiologiques

Au cours des années 1970-1980, la modélisation pharmacocinétique basée sur la physiologie (PBPK) a été développée pour simuler les concentrations internes des toxiques, reconnues plus prédictives des réponses biologiques que les doses d'exposition. Les modèles PBPK considèrent l'anatomie, la physiologie, la biochimie et la physicochimie de l'organisme étudié pour simuler le devenir d'une substance dans un organisme. Ils mettent en évidence les organes ayant une action sur la substance (métabolisme, transport, dégradation, stockage...), regroupent ceux ayant peu d'importance sur les processus ou possédant grossièrement des caractéristiques similaires en un seul compartiment, puis les relient par la circulation sanguine artérielle et veineuse si l'organisme d'étude en est pourvue, comme présenté par la Figure 2 (Brochot et al., 2014).

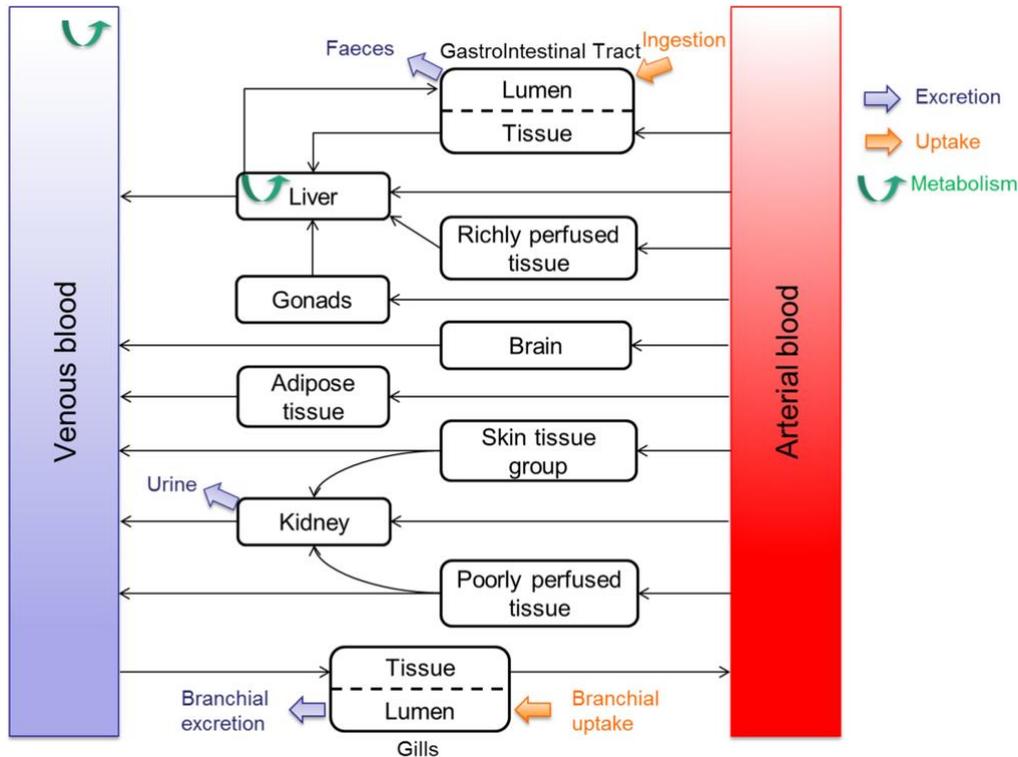


Figure 2. Modèle pharmacocinétique basé sur la physiologie du poisson adulte. L'absorption, l'élimination et le métabolisme sont représentés par les flèches orange, bleue et verte, respectivement (Grech et al., 2019).

Les échanges entre les organes, matérialisés par des compartiments, se basent sur des transports de masse, retranscrits en équations différentielles. La modélisation PBPK permet de décrire quantitativement les processus ADME des composés en corrélant les déterminants critiques que sont les volumes tissulaires, les débits physiologiques, les taux d'absorption, la diffusion membranaire, les taux et coefficients de partage tissu – sang et les réactions biochimiques (Krishnan and Peyret, 2009; Munir et al., 2018; Pearce et al., 2017). Ils peuvent également considérer des facteurs impactant ces déterminants, tels que l'âge ou les conditions d'exposition à un composé, pour étudier des populations spécifiques (Beaudouin et al., 2010; Lipscomb et al., 2012). En évaluation des risques, la modélisation PBPK est de plus en plus performante et concerne divers champs d'applications. Elle se revendique pertinente pour analyser des données sur les mécanismes pharmacocinétiques, orienter les protocoles d'études expérimentales, caractériser la variabilité et l'incertitude physiologiques et pharmacocinétiques, extrapoler les données *in vivo* et *in vitro* entre les espèces, les stades biologiques, les voies d'exposition et le temps, et finalement étayer l'évaluation globale de l'exposition à un composé (Chou and Lin, 2019; Jaroch et al., 2018; Mumtaz et al., 2012; Tan et al., 2018; WHO, 2010).

1.2. La modélisation pharmacodynamique

La pharmacodynamie est définie par l'action d'un médicament sur un système biologique. Elle permet d'étudier la relation entre la concentration interne relative à l'interaction d'un composé avec sa cible et les effets biochimiques, moléculaires et physiologiques résultants. Un effet peut être décrit par son intensité, son délai d'apparition ainsi que sa durée, et quantifié selon différentes approches : par description des effets en fonction du temps, de la dose d'exposition et par élaboration de la relation entre la concentration – effet en fonction du temps (PK/PD) (Gabrielsson and Weiner, 2006).

1.2.1. Les modèles empiriques

Les modèles de relation quantitative structure – activité (QSAR)

Le principe de la relation structure – activité repose sur la supposition que l'activité biologique des molécules est le reflet de leur structure et que des molécules similaires ont des activités similaires. Les modèles de relation quantitative structure – activité (QSAR) permettent de prédire les propriétés physico-chimiques, biologiques et le devenir dans l'environnement des composés à partir de leur structure moléculaire (Lomba et al., 2019; Papa et al., 2014). Historiquement, l'un des premiers modèles QSAR a été développé en 1962 par (Hansch et al., 1962) pour estimer la concentration chimique (C) en fonction du coefficient de partage octanol – eau (π) et la constante de Hammett (σ) relative aux effets électroniques d'un substituant (Hammett, 1937), selon l'équation suivante (Raies and Bajic, 2016) :

$$\log\left(\frac{1}{C}\right) = 4,08\pi - 2,14\pi^2 + 2,78\sigma + 3,36 \quad (2)$$

L'objectif de la modélisation QSAR est de déterminer les facteurs impactant l'activité mesurée d'un composé, toxique ou non, pour un système biologique particulier, et d'avoir un aperçu du mécanisme et du comportement du système étudié (Kar et al., 2018). Les méthodes de modélisation utilisées sont de plusieurs types : corrélation, régression et classification. En toxicologie, les analyses de régression sont les plus courantes car elles apportent un caractère quantitatif et une base mécaniste pour la toxicité modélisée contrairement aux techniques de classifications qualitatives (Schultz et al., 2003).

En évaluation des risques, les modèles QSAR permettent de fournir des informations sur les dangers des composés en réduisant le temps, le coût et le recours aux animaux pour les tests de toxicité. Cependant, peu de modèles mécanistes existent (Benigni et al., 2007; Ringeissen et al., 2011) et leur nature empirique réduit leur application (par les organismes de réglementation) à la hiérarchisation des composés existants et nouveaux pour une évaluation ultérieure de leur toxicité (Russom et al., 2003; Schmieder et al., 2003; Walker et al., 2003). De plus, les modèles QSAR ne prédisent l'activité que d'une famille ou classe de composés alors que la réglementation, par exemple des pesticides et produits biocides, peut demander l'évaluation de formulations complexes (Scholz et al., 2013a). Dans l'optique scientifique actuelle de proposer des méthodes alternatives aux expérimentations animales, l'Organisation de Coopération et de Développement Economiques (OCDE) propose une « boîte à outils » QSAR pour rendre la technologie accessible, transparente et moins exigeante en termes de coûts d'infrastructure, dans le but d'accroître l'acceptation réglementaire des méthodes QSAR. Cette « boîte à outils » est un logiciel destiné à être utilisé par les gouvernements, l'industrie chimique et d'autres parties prenantes pour combler les lacunes des données (éco)toxiques nécessaires à l'évaluation des dangers des produits chimiques (OECD, 2007).

Les modèles dose-réponse

D'après la définition de l'Organisation Mondiale de la Santé (OMS), l'évaluation de la dose-réponse permet de déterminer la relation quantitative entre des doses d'exposition ou internes et les réponses biologiques. La réponse peut être une incidence mesurée, observée ou une modification du niveau de réponse, un pourcentage de réponse dans des groupes de sujets ou une population, la probabilité d'occurrence ou de changement du niveau de réponse au sein d'une population (WHO, 2010). En évaluation des risques, simultanément à l'analyse descriptive des données, les modèles empiriques dose-réponse sont utilisés pour quantifier statistiquement la relation entre la dose d'exposition d'un composé et les réponses biologiques ou toxicologiques au sein de populations cibles, en considérant la variabilité et l'incertitude (Mushak, 2013). Les plus couramment rencontrés pour prédire une probabilité de réponse suite à l'exposition d'un composé sont les modèles de Weibull, probit et logit. Ils modélisent une courbe dose-réponse dont l'équation, par dérivation, permet d'établir les concentrations effectives (EC) responsables de 10 ou 50% de variation de la réponse initiale (Christensen et al., 2009; Christensen and Chen, 1985).

1.2.2. Les modèles mécanistes

Les modèles PBPK/PD

Au cours des dernières années, les analyses PK/PD initialement basées sur l'élaboration de la relation dose-réponse se sont transformées en des modèles mathématiques plus sophistiqués pour comprendre le mécanisme d'action des composés. Cette évolution résulte principalement de l'amélioration des méthodes analytiques, des matériels informatiques et des logiciels, de l'intérêt réglementaire et académique accru et du raffinement continu des modèles pharmacodynamiques basés sur des mécanismes physiologiques (PBPK/PD) (Csajka and Verotta, 2006). En évaluation des risques, les premières études menées pour l'identification des dangers facilitent la détermination des organes, des tissus et des systèmes affectés par l'exposition à un composé. Pour l'évaluation dose-réponse, les données décrivant les réponses sont généralement interprétées à partir de la dose externe d'exposition (Lipscomb et al., 2012).

Les modèles PBPK/PD proposent, par couplage avec un modèle PBPK, une approche mécaniste pour quantifier l'effet pharmacodynamique d'un composé par intégration de l'évolution des concentrations externes et internes au cours du temps. Ils permettent de décrire simultanément les processus ADME et l'effet des composés à différents niveaux biologiques, cellulaire ou tissulaire, par représentation mécaniste et physiologique des interactions dans le système biologique d'étude, et ainsi de comprendre les relations entre concentration d'exposition, temps et effet (Ducrot et al., 2016; Kuepfer et al., 2016), comme présenté par la Figure 3, illustrant le modèle PBPK/PD développé par (Sharma et al., 2017) pour étudier la neurotoxicité de l'acide perfluorooctanesulfonique (PFOS) chez l'Homme.

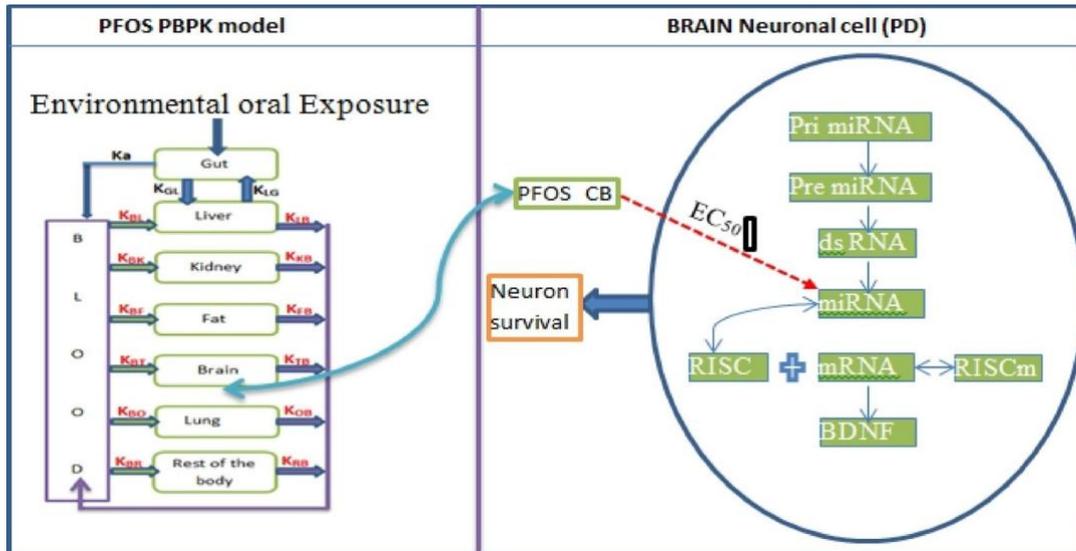


Figure 3. Schéma d'un modèle PBPK/PD de l'acide perfluorooctanesulfonique (Sharma et al., 2017).

Les processus ADME peuvent varier entre les espèces animales et l'Homme. Appréhender et déterminer les différences inter-espèces est essentiel pour extrapoler de façon pertinente et argumentée les données de réponses biologiques animales à l'humain (Lipscomb et al., 2012). Par ses considérations mécanistes, la modélisation PBPK/PD est un outil utile en évaluation des risques pour extrapoler les concentrations toxiques entre différentes espèces, mais également entre les études *in vitro* et *in vivo*, les études précliniques et cliniques, de la dose externe d'exposition à la dose systémique, pour relier la physiologie et la génétique à la disposition des composés dans les populations animales et humaines, et enfin, pour simuler la temporalité des processus toxiques (Ashauer and Escher, 2010; Dixit et al., 2003).

Les modèles de toxicologie systémique

Les voies de signalisation cellulaires ou métaboliques expliquent comment les gènes et protéines interagissent entre eux pour déclencher les fonctionnalités cellulaires ainsi que la génération de molécules ou de métabolites. La Figure 4 présente schématiquement la cascade cellulaire impliquée dans le métabolisme du glutathion, acteur essentiel des réactions de détoxification et de défense contre les espèces réactives de l'oxygène (Reed et al., 2008). La production de données génomiques a permis d'accroître les connaissances sur les mécanismes génétiques d'un grand nombre de processus toxiques. De plus, l'imagerie cellulaire permet la compréhension de la corrélation entre les paramètres, les gènes, les protéines et les voies pour

guider l'intégration de la dynamique moléculaire et cellulaire à l'aide d'approches de modélisation des systèmes (Sipes et al., 2011a).

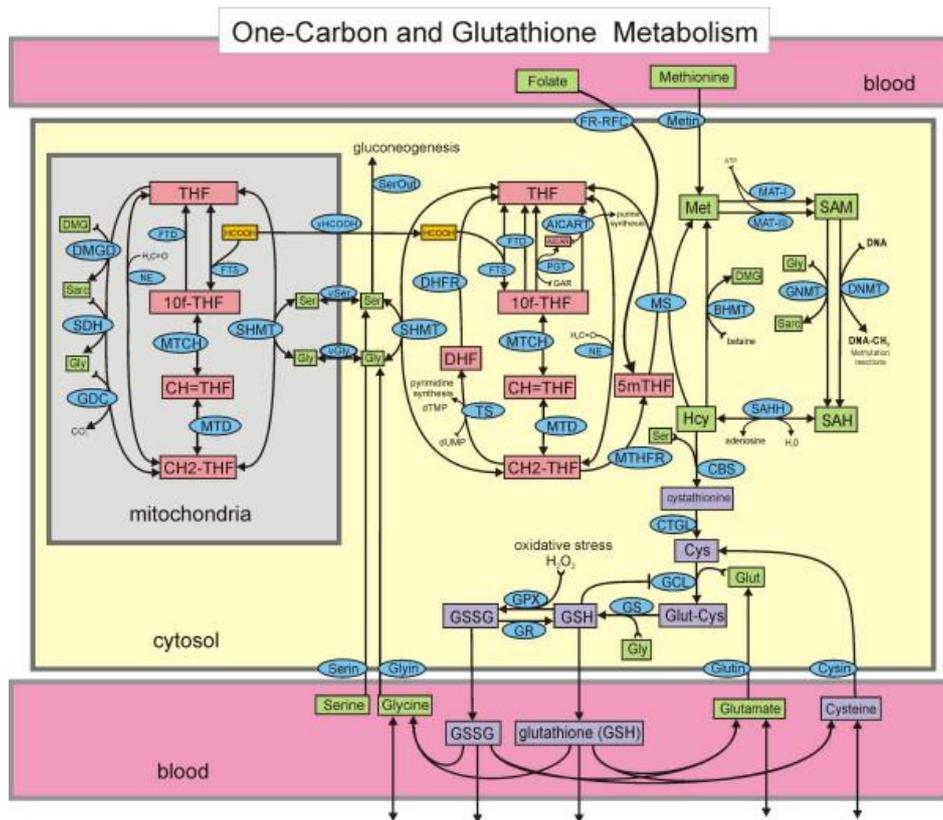


Figure 4. Schéma du modèle de la voie du métabolisme du GSH par Reed et al. (2008)

Des modèles mathématiques de type différentiel, ou discret (Booléen par exemple), parfois stochastiques, de l'activité de ces voies décrivent le comportement des cellules pour étudier les mécanismes moléculaires d'action selon des conditions d'exposition particulières et prédire les conséquences potentielles des perturbations (mutations ou changements de l'expression) des protéines *etc.* Cette approche holistique offre une alternative aux approches centrées uniquement sur les gènes pour contribuer à l'interprétation de la variabilité génomique des phénomènes toxicologiques complexes. En d'autres termes, cela signifie que l'exposition à des composés peut conduire à des variations dans la séquence d'ADN (Amadoz et al., 2015; Peña-Chilet et al., 2019; Rahman et al., 2018).

1.2.3. Les modèles intermédiaires entre empiriques et mécanistesLes modèles quantitatifs de chemins d'effets néfastes (qAOP)

Les chemins d'effets néfastes (AOP : adverse outcome pathways) modélisent le processus toxique d'un composé en reliant un évènement initiateur moléculaire (MIE : molecular initiating event) à l'effet toxique observé (AO : adverse outcome) par l'intermédiaire d'une chaîne d'évènements clés (KE : key event) à des niveaux croissants d'organisation biologique ; des réponses cellulaires jusqu'à la dynamique des populations, selon la Figure 5. Construits à partir d'informations mécanistes obtenues *in vitro*, ils sont axés sur l'identification des cibles moléculaires potentielles d'un composé et les effets *in vivo* à différents niveaux biologiques qui en résultent.

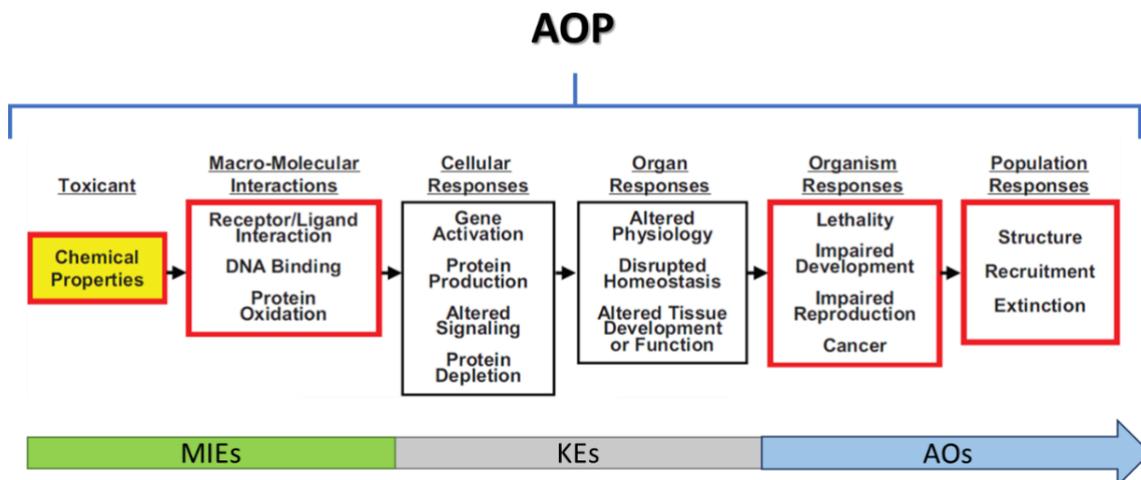


Figure 5. Représentation schématique des différents niveaux organisationnels d'un AOP, adapté à partir de (Ankley et al., 2010).

L'intégration de la pharmacocinétique et de la pharmacodynamie ajoute à la nature fondamentalement qualitative des AOP une approche quantitative (qAOP) pour la prédiction de la toxicité des composés. La considération de tous les niveaux d'organisation biologique permet d'évaluer le risque sur la partie du réseau (niveau moléculaire ou populationnel) pour différentes espèces et sur la voie d'exposition, rendant ainsi pertinent l'utilisation de qAOP pour la gestion des risques et la prise de décisions en santé – environnement (Leist et al., 2017; Margiotta-Casaluci et al., 2016), selon la Figure 6. En fonction des données disponibles, les AOP peuvent être élaborés à partir de modèles statistiques, bayésiens, de régression, d'équations différentielles ordinaires (ODE) et individuels (Groh et al., 2015; Spinu et al.,

2020), afin de répondre aux questions posées au cours des quatre étapes constitutives de l'évaluation des risques : évaluation des dangers, évaluation de la relation dose-réponse, évaluation de l'exposition et caractérisation des risques (Perkins et al., 2019).

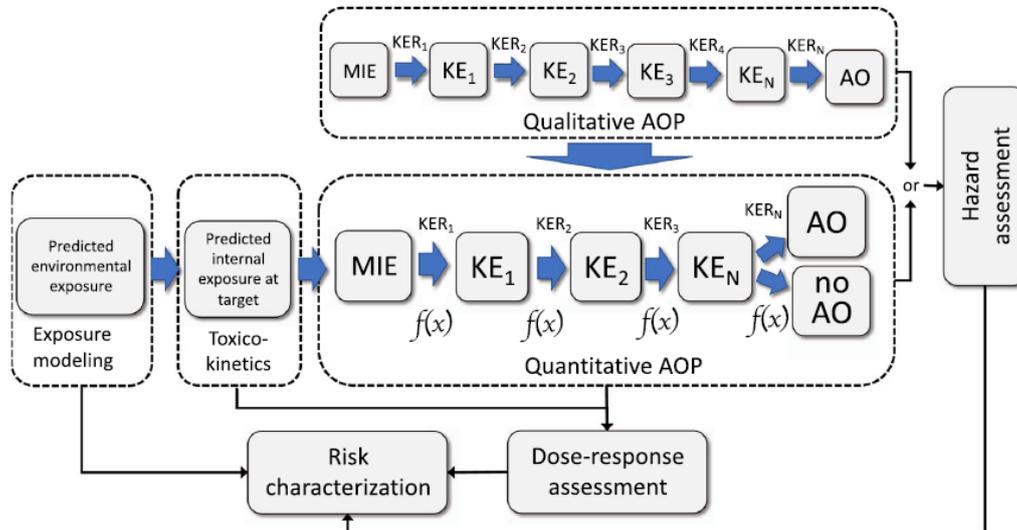


Figure 6. Organisation et utilisation des chemins de l'effet néfaste (AOP) pour l'évaluation des risques (Perkins et al., 2019).

Les modèles de survie

Décrire l'apparition de phénomènes dans le temps pour déterminer les facteurs causaux représente une composante essentielle de l'évaluation des risques. Dans le domaine santé – environnement, l'occurrence des effets observés suite à l'exposition à un composé est relevée à des temps précis, prédéterminés en fonction du protocole de l'étude. Par exemple, pour les tests de toxicité animale, l'OECD recommande d'analyser séparément les doses-réponses aux différentes concentrations d'exposition étudiées, pour chaque temps d'observation fixé (OECD, 2013). Cependant, les réponses, appelées évènements en analyse de survie, peuvent se produire à n'importe quel instant au cours de l'étude, biaisant les temps d'occurrence. Ces processus d'observation mènent à des données censurées, pouvant être de trois types :

- Censure à droite : les évènements se produisent après le temps de la dernière observation ;
- Censure à gauche : les évènements se produisent avant le temps de la première observation ;
- Censure par intervalle : les évènements se produisent entre deux temps d'observation (Moore, 2016; Philippe Saint Pierre, 2015).

Dans certaines études, une partie des individus initialement inclus n'est pas observable et les analyses se font à partir de sous-échantillons. Il s'agit d'un processus de troncation dont trois types sont également possibles : troncations à droite, à gauche ou par intervalle (Saint Pierre, 2015).

Les évènements sont en partie conditionnés par les phénomènes passés. Séparer les observations par groupes de temps réduit donc le nombre de données par analyse et ignore cette dépendance temporelle, entraînant une perte d'information importante. Les modèles de survie ont été élaborés afin de répondre à cette problématique en proposant de modéliser la distribution du temps d'apparition d'un évènement, à partir d'un temps initial fixé et de données observées chez le système biologique d'étude. Ce sont donc des modèles stochastiques car ils prédisent des probabilités d'occurrence. Ils permettent de décrire des données univariées (par exemple, l'âge du décès) ou multivariées (dose d'exposition et effet) (Bradburn et al., 2003a, 2003b; Clark et al., 2003a, 2003b; Jiang and Fine, 2007; Lee and Go, 1997; Leung et al., 1997; Mahnken et al., 2008). Ils sont communément utilisés au cours des études cliniques et des analyses de survie (Gould et al., 2015; Karmen et al., 2019). Le modèle de survie le plus simple décrit le passage de l'état 1 « vivant », à l'état 2 « mort », comme illustré par la Figure 7. Le taux de transition $q_{1,2}$ peut relier l'exposition instantanée aux médicaments ou aux substances toxiques et l'évolution temporelle de la mortalité observée chez les individus exposés. Il est à noter qu'un état est par définition statique, à la différence de l'évènement, dynamique au cours du temps. Un individu change d'état à la suite de l'occurrence d'un évènement.

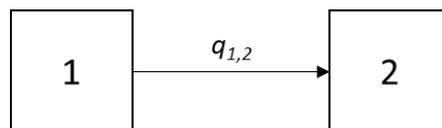


Figure 7. Schéma du processus basique d'analyse de survie.

Les analyses de survie font l'hypothèse que le temps de survie T (temps passé dans l'état 1) est une variable aléatoire. La fonction de densité de probabilité, $f(t)$, et la fonction de distribution cumulative, $F(t)$, caractérisent la distribution de cette variable aléatoire. La fonction de survie est définie par :

$$S(t) = 1 - F(t) = P(T > t), \quad t \geq 0 \quad (3)$$

La probabilité de passer d'un état initial i ($X(0)$) à un état aléatoire j ($X(t)$) au cours du temps t est définie comme suit :

$$p_{i,j}(t) = P\{X(t) = j | X(0) = i\} \quad (4)$$

Les modèles multi-états

Les modèles multi-états sont des modèles de survie plus complexes, considérant que plusieurs évènements peuvent découler de l'état initial d'observation, comme présenté par la Figure 8. Dans ce cas, il est possible que les observations soient indirectement liées, puisque si la transition de l'état 1 à l'état 2 se produit, l'individu ne pourra plus se trouver dans l'état 3. Cela conduit à un biais d'observation de l'état 3. Par exemple, des embryons observés morts ne développent pas de malformations cardiaques ultérieures. Ainsi, pour des concentrations létales, la prévalence de malformations cardiaques diminue car les embryons meurent avant le développement du cœur. Il s'agit d'un processus des risques concurrents. Pour résoudre ce genre de problèmes au cours des analyses de données de survie, les modèles multi-états markoviens peuvent décrire la distribution des temps de séjour des sujets dans les différents états, avec des probabilités d'évènement dans un état donné non dépendantes de la façon dont cet état a été atteint (Andersen et al., 2012; Beyersmann et al., 2012; Meira-Machado et al., 2009).

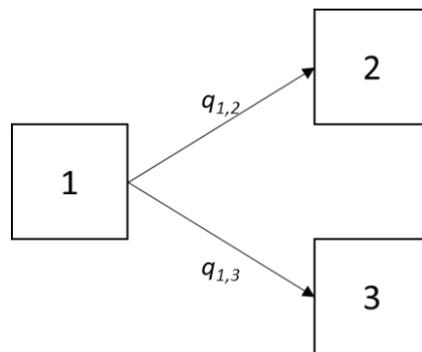


Figure 8. Schéma d'un processus simple des risques concurrents.

Les modèles multi-états markoviens sont très généraux, et peuvent considérer des évènements intermédiaires ainsi que des retours possibles entre états (Farewell and Tom, 2014; Putter et al., 2007), comme illustré par la Figure 9.

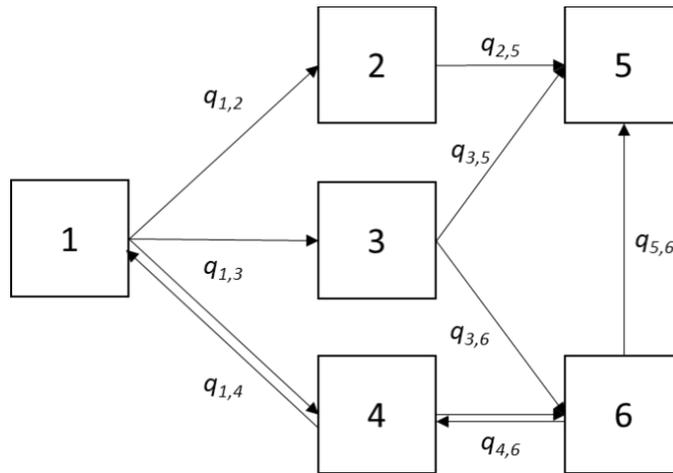


Figure 9. Schéma d'un modèle multi-états, avec évènements intermédiaires et possibilité de retour entre états.

Comme dans les modèles simples de survie, une série de taux de transitions $q_{i,j}(t,z(t))$ détermine le risque de passer de l'état i à l'état j , en fonction du temps et de variables individuelles $z(t)$, pouvant être également dépendante du temps (Jackson, 2011).

Deux simplifications sont possibles à partir du modèle général markovien :

- Les modèles multi-états markoviens homogènes dans le temps considèrent que les taux de transition dépendent seulement de l'état initial et sont indépendants du temps (Putter et al., 2007; Zare et al., 2014) :

$$q_{ij}(t, z(t)) = q_{ij}(z) \quad (5)$$

- Les modèles semi-markoviens considèrent que les taux de transition ne dépendent que de l'état initial i et du temps d'entrée t_j dans l'état suivant j (Meira-Machado et al., 2009) :

$$q_{ij}(t, z(t)) = q_{ij}(t - t_j, z) \quad (6)$$

Le modèle peut également être étendu aux processus de Markov à état cachés si l'appartenance aux états ne peut être donnée directement par les observations (Lavielle, 2018).

1.3. L'inférence statistique sur les paramètres des modèles

1.3.1. L'inférence bayésienne

On entend généralement par inférence statistique l'estimation des valeurs des paramètres d'un modèle mathématique sur la base d'un ensemble d'observations (« données »). Il y a actuellement deux grandes classes de méthodes d'inférence statistique : les approches bayésiennes et les approches fréquentistes. L'inférence bayésienne considère les paramètres à estimer comme des variables aléatoires et les données comme des valeurs fixes, alors que l'inférence fréquentiste fait juste le contraire.

L'inférence fréquentiste se fonde généralement sur la considération d'ensembles hypothétiques de données répliquées. Traditionnellement, les méthodes et outils fréquentistes comprennent des tests d'hypothèse, des intervalles de confiance et des estimateurs de paramètres. Une caractéristique fondamentale de l'inférence fréquentiste est l'évaluation des probabilités d'erreur, *i.e.* des probabilités de mauvaises décisions (Gelman and Hennig, 2017).

L'avantage de l'approche bayésienne est d'intégrer facilement des connaissances *a priori*, (disponibles indépendamment du jeu de données en cours d'analyse) ou des données obtenues séquentiellement (Gelman and Hennig, 2017). Cependant, l'approche bayésienne requiert de définir une distribution *a priori* des paramètres à estimer ; ce choix n'est pas toujours facile, et peut être perçu comme subjectif, d'où certaines controverses et critiques sur le choix le plus approprié en l'absence de connaissances *a priori* (O'Hagan, 2008). Le théorème de Bayes permet de dériver la distribution *a posteriori* de paramètres θ en fonction d'un *a priori* et de nouvelles données selon la formule suivante :

$$f(\theta|\mathbf{y}) \propto f(\theta) \times f(\mathbf{y}|\theta) \quad (7)$$

Avec $f(\theta|\mathbf{y})$ la densité *a posteriori* de θ , $f(\theta)$ la densité *a priori* et $f(\mathbf{y}|\theta)$ la distribution de probabilité des données, conditionnellement à θ . Cette dernière distribution est la seule considérée dans l'approche fréquentiste, elle est aussi appelée fonction de vraisemblance. La densité *a posteriori* a par définition une intégrale égale à 1. Cependant, le produit de la densité *a priori* par la vraisemblance n'a généralement pas une intégrale égale à 1. La constante de proportionnalité entre ce produit et la densité *a posteriori* est égale à la probabilité totale des données (intégrale du produit par rapport à θ). Cette constante n'est en général pas connue, sauf

dans des cas relativement simples (O'Hagan, 2008). Ceci a limité pendant longtemps l'application des méthodes bayésiennes.

Dans l'inférence bayésienne, toutes les décisions à prendre sont dérivées de la distribution *a posteriori*, puisqu'elle contient toute l'information disponible sur les paramètres après observation des données. Divers critères de décision peuvent être appliqués pour obtenir une estimation ponctuelle de chaque paramètre θ .

1.3.2. Les modèles multi-niveaux

Dans les situations de recherche les plus courantes, l'évaluation directe de l'ensemble d'une population d'intérêt n'est pas possible. Ainsi, les études expérimentales se portent généralement sur un échantillon pour déterminer les paramètres de la population générale. En évaluation des risques, il est important d'évaluer l'incertitude et la variabilité affectant les résultats expérimentaux afin d'accorder confiance aux prédictions des modèles. L'incertitude est principalement le résultat d'un manque de connaissance nourri par la précision des instruments de mesure, l'utilisation de données pour ajuster les modèles PK et PD causant du « bruit » sur les valeurs des paramètres résultantes, le réductionnisme de l'échantillon comparé à l'étude d'une population entière, les simplifications des modèles et les extrapolations *in vitro* – *in vivo* ou inter-espèces. La variabilité considère les différences de caractéristiques épigénétiques, génétiques, environnementales et leurs interactions, entre les individus (variabilité interindividuelle) ou dans le temps pour un individu donné (variabilité intra-individuelle). Si l'incertitude peut être réduite par enrichissement des connaissances des processus toxiques par exemple, la variabilité est inhérente aux populations. Ainsi, pour considérer l'incertitude impactant l'évaluation de la variabilité, des modèles statistiques de populations ont été développés, initialement pour les études pharmacocinétiques puis élargis aux modèles dose-réponse. L'objectif des modèles de population est d'extraire, à partir de données sur des unités individuelles (par exemple les lots d'embryons de poisson zèbre spécifiques à chaque temps d'observation du protocole expérimental de l'OCDE), une description quantitative de la variabilité de la cinétique d'un composé (PK) ou de l'occurrence d'un effet (PD) au sein d'une population. Conceptuellement, le même modèle décrit les données pour chaque unité et ses paramètres varient aléatoirement entre les unités ou les occasions d'observation d'une même unité. La variabilité, n'ayant pas de fondement mécaniste, est décrite par une distribution probabiliste. Par exemple, dans le modèle PK-PD multi-états développé

pour les travaux de cette thèse, le taux d'éclosion des embryons se distribue log-normalement autour d'une moyenne de population, avec une variance de population relative à la variabilité populationnelle. Les moyennes et les variances de la population sont appelées paramètres de population. Ils sont estimés par ajustement des données, avec les paramètres du modèle pour chaque unité. D'autres paramètres de population peuvent être définis, comme par exemple les covariances des paramètres si certains paramètres semblent corrélés au niveau populationnel (Bois, 2001; Wakefield and Rahman, 2000).

1.3.3. La calibration des modèles développés pour les travaux de thèse

Les modèles développés pour les travaux de cette thèse suivent des méthodes de Monte Carlo par chaînes de Markov (MCMC) pour calculer des approximations numériques d'intégrales multidimensionnelles, selon une approche bayésienne. Dans les statistiques bayésiennes, le développement récent des méthodes Monte Carlo avec chaînes de Markov simulées – par exemple, l'échantillonnage de Gibbs (Jones et al., 2017) – permet de calculer des modèles multi-niveaux avec de nombreux paramètres inconnus. Ces méthodes créent des échantillons d'une variable aléatoire continue multidimensionnelle, avec une densité de probabilité proportionnelle à une fonction de densité connue. Plusieurs chaînes sont typiquement générées à partir d'un ensemble de points choisis arbitrairement et suffisamment éloignés les uns des autres (Banerjee et al., 2015). Dans un cadre bayésien, la densité cible est la distribution *a posteriori* des paramètres du modèle, étant donné le modèle statistique, les distributions de paramètres *a priori* et les données pour lesquelles une fonction de vraisemblance peut être calculée (Bois, 2009; Smith and Roberts, 1993).

Les modèles pharmacocinétiques et pharmacodynamiques suivent des approches fréquentistes ou bayésiennes. Pour les modèle PBPK, l'approche bayésienne est plus appropriée puisque leurs paramètres ont en général une signification physique directe et bénéficient d'estimation *a priori* indépendantes des données à analyser. C'est moins souvent le cas pour les modèles multi-états, également moins utilisés et moins spécialisés. Cependant, même pour ces modèles l'analyse bayésienne est de plus en plus proposée (Williams et al., 2020). Nous adopterons cette approche car elle rend assez facile l'application des modèles multi-niveaux.

1.4. Les intérêts et limites de la modélisation pour l'extrapolation à l'Homme

Pour extrapoler les données de toxicité animale et *in vitro* d'un composé à l'Homme, il est primordial de déterminer sa pharmacocinétique, relative à son comportement au sein du corps humain, et sa pharmacodynamie, relative à ses effets au niveau des cibles biologiques. Malgré une distinction courante des deux processus, ils forment un continuum puisque l'occurrence et l'intensité des effets dépendent de l'évolution de la concentration du composé au cours du temps, au sein d'un système d'étude. En raison des effets pharmacocinétiques, les sensibilités tissulaires et organiques observées chez l'animal ou *in vitro* peuvent être différentes chez l'Homme. Ainsi, la première étape pour extrapoler de façon pertinente des données de toxicité est de quantifier la pharmacocinétique des composés. Pour ce faire, les modèles mécanistes PBPK représentent actuellement la meilleure alternative pour extrapoler les données à l'Homme, avec un large domaine d'applicabilité et permettre de prédire la pharmacocinétique d'un composé au-delà du temps d'observation et des doses testées. La Figure 10 illustre les principales données essentielles pour informer sur la structure des modèles et les valeurs des paramètres. La deuxième étape consiste à prédire la réponse pharmacodynamique ou toxique chez l'Homme. Comme une même concentration d'exposition *in vitro* ou chez l'animal peut conduire à des effets différents chez l'Homme, il est important de développer des modèles mécanistes des réponses cellulaires caractérisés par une structure hiérarchique (Nadia Quignot et al., 2014).

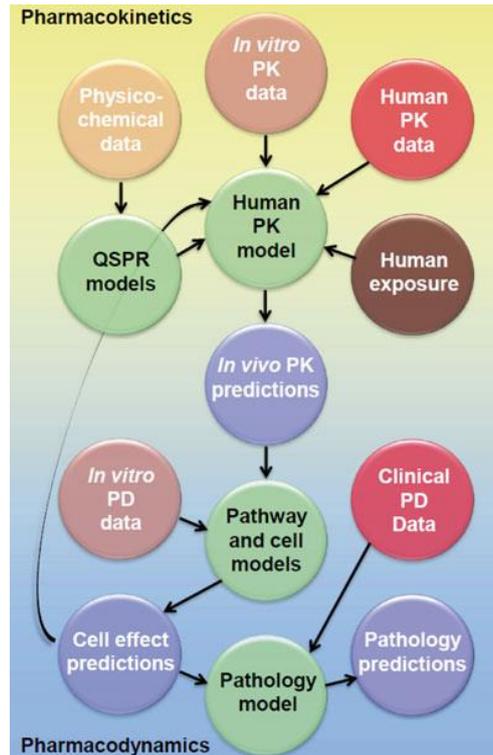


Figure 10. Principales données nécessaires à l'extrapolation quantitative des données à l'Homme et de la prédiction de la toxicité. Les prédictions (en bleu) de la pharmacocinétique *in vivo*, des effets cellulaires *in vivo*, ainsi que des paramètres pathologiques (toxicité), sont obtenues par différents modèles (en vert) capables d'intégrer des données (en rouge) provenant de diverses sources (Nadia Quignot et al., 2014).

Prédire la toxicité chez l'Homme implique également de considérer la variabilité des caractéristiques physiologiques, du métabolisme, etc. au sein de la population humaine. Si l'impact de la PK sur la PD est considéré, l'inverse l'est rarement. Cependant, l'action de la PD sur la PK est susceptible de se produire et ce avec une certaine latence et des mécanismes complexes. Les modèles prédictifs de tels effets, notamment des effets cumulatifs, de la toxicité développementale ou de la carcinogenèse, permettraient de couvrir le continuum PK – PD (Nadia Quignot et al., 2014).

Les modèles mathématiques représentent un réel atout pour valoriser les expérimentations animales et les modèles *in vitro*. Ils fournissent des informations complémentaires fondamentales pour permettre une juste extrapolation des données de toxicité des composés à l'Homme, par modélisation des processus toxiques à tout niveau de considération biologique. Cependant, le critère prédictif et représentatif caractérisant les modèles *in silico* conduit à des limites à considérer lors de l'interprétation de résultats.