

NOTE DE SYNTHÈSE

État de l'art sur la modélisation QSAR de la toxicité des mélanges chimiques

Mai 2022

La caractérisation des dangers de mélanges de substances chimiques représente un défi scientifique lors de l'évaluation de leur risque (éco)toxicologique. La caractérisation expérimentale de la toxicité des mélanges chimiques est souvent associée à des coûts économiques et éthiques élevés en raison du nombre important de tests expérimentaux sur lesquels elle se base.

Dans le cadre de son programme d'appui aux politiques publiques, l'Ineris analyse la mise en place des méthodes alternatives en expérimentations animales (méthodes *in silico* et *in vitro*) pour les règlements REACH et CLP. L'objectif de cette note est de faire un état des lieux des approches de modélisation QSAR (« Quantitative Structure-Activity Relationship ») pour prédire les effets toxicologiques des mélanges chimiques. Ces approches computationnelles se présentent comme des méthodes de remplacement et une alternative rapide et avantageuse économiquement et éthiquement à l'expérimentation animale. Cette note dresse un bilan sur les potentialités, limitations et axes de progrès de ce domaine d'investigation qui constitue une priorité stratégique pour l'Ineris, avec notamment une attention accrue portée aux effets cocktails, et phénomènes de synergie/antagonisme entre substances chimiques.

La problématique de la prédiction de la toxicité des mélanges chimiques

Comme mis en évidence par les campagnes de biosurveillance nationales et européennes, les populations humaines sont exposées simultanément à de nombreux produits chimiques présents dans l'environnement (Willey *et al.* 2021). Les substances chimiques sont généralement évaluées et réglementées de façon individuelle. Or les substances présentes dans un mélange peuvent interagir entre elles et produire des effets différents de ceux attendus pour les substances seules (Sprinkle and Payne-Sturges 2021). À la complexité de ces phénomènes viennent s'ajouter une nature des mélanges pas toujours connue, une évolution temporelle de l'exposition et de la composition, et une caractérisation toxicologique basée sur des plans d'expérience nécessitant un nombre important de tests expérimentaux (Sillé *et al.* 2020).

L'effet toxicologique des substances chimiques au sein d'un mélange est usuellement décrit sous l'hypothèse d'addition des doses pour les substances chimiques ayant des mécanismes d'action similaires, ou d'action indépendante dans les autres cas. En pratique, la majorité des approches

d'évaluation des dangers de mélanges suppose que la toxicité d'un mélange suit le modèle d'addition des doses. Selon cette approche, la toxicité du mélange peut être prédite par la somme des toxicités des substances chimiques individuelles. Cependant, dans certains mélanges, des interactions peuvent se produire entre les substances chimiques ou au niveau des cibles biologiques. Ces interactions peuvent modifier la toxicité de sorte qu'elle est plus ou moins importante que celle attendue sur la base d'une addition des concentrations (Heys *et al.* 2016).

Il apparaît donc que l'évaluation toxicologique des mélanges chimiques est marquée par une inhérente complexité biologique couplée avec une démultiplication du nombre de mélanges possibles parmi toutes les substances présentes dans l'environnement. De plus, si la caractérisation toxicologique des mélanges chimiques est effectuée *in vivo*, il faut aussi envisager les aspects éthiques liés à l'expérimentation animale.

Dans ce contexte, la possibilité d'adapter la modélisation QSAR (Quantitative Structure-Activity Relationship, Relation Quantitative entre Structure chimique et Activité biologique), normalement appliquée aux substances

individuelles, à la problématique des mélanges chimiques représenterait une possibilité intéressante.

Principes de modélisation QSAR

La modélisation QSAR permet de prédire les effets biologiques de substances chimiques en fonction de leur structure moléculaire (Mombelli and Ringeissen 2009). Elle est aujourd'hui utilisée dans plusieurs domaines liés aux sciences du vivant tels que la toxicologie ou la pharmacologie.

Les prédictions élaborées par ces modèles peuvent être quantitatives (ex. une dose létale médiane) ou qualitatives (modèles QSAR de classification, ex. substance génotoxique vs. Substance non génotoxique). Cette dernière catégorie de modèles de classification est aussi identifiée par l'acronyme SAR.

Il faut aussi souligner que ces approches de modélisation permettent de répondre aux exigences en termes de remplacement de l'expérimentation animale dont les aspects éthiques sont encadrés par le principe des 3R (Russell 1995) : Remplacer le recours à l'expérimentation *in vivo*, Réduire le nombre d'animaux et Raffiner l'expérimentation afin de réduire au minimum la souffrance animale.

Substances individuelles

Les modèles QSAR sont traditionnellement développés et utilisés pour prédire les effets sur le vivant de substances chimiques considérées individuellement. D'un point de vue conceptuel, le développement d'un modèle QSAR nécessite trois composantes (Roy et al. 2015) :

- / un jeu de données mettant à disposition des mesures expérimentales d'une activité biologique (ex. une dose létale médiane pour une espèce animale donnée) permettant de calibrer le modèle QSAR. Ce jeu de données est communément appelé « jeu d'apprentissage » ;
- / une description de la structure moléculaire des substances chimiques qui composent le jeu d'apprentissage, obtenue grâce au calcul de variables numériques connues sous le nom de « descripteurs moléculaires » (ex. poids moléculaire, nombre de liaison carbone-carbone...);
- / une méthode statistique pour caractériser la relation entre l'effet biologique à prédire et les descripteurs moléculaires.

Une fois le modèle développé, celui-ci doit être validé au moyen d'une validation croisée⁷ effectuée sur le jeu d'apprentissage (validation interne) et d'une validation externe effectuée sur des données supplémentaires regroupées au sein d'un « jeu de validation externe » (Gramatica 2007). La validation externe vérifie le pouvoir prédictif d'un modèle QSAR vis-à-vis d'un ensemble de substances chimiques n'ayant pas été utilisées lors du

développement du modèle. Elle quantifie la capacité du modèle à prédire la toxicité de substances « nouvelles ».

Mélanges chimiques

Le développement de modèles QSAR pour les mélanges chimiques diffère de celui des modèles des substances individuelles essentiellement pour deux aspects : la description moléculaire doit être adaptée à un mélange chimique, et la validation des modèles QSAR doit tenir compte de la coprésence de plusieurs substances chimiques dans différents mélanges (Muratov et al., 2012).

Description moléculaire des mélanges chimiques

Les descripteurs des mélanges utilisés pour les modèles QSAR peuvent être répartis en plusieurs catégories selon les besoins de l'étude. Muratov et al. (2012) ont proposé la classification suivante :

- / **Les descripteurs basés sur le coefficient de partage² du mélange** entre une phase aqueuse et une phase hydrophobe. La particularité de ces descripteurs est de considérer la propriété du mélange dans son ensemble, plutôt que pour chaque substance chimiquement séparément. Muratov et al. (2012) ont proposé le calcul du coefficient de partage du mélange (K_{mix}) à partir des coefficients de partage des substances individuelles (K_i) :

$$K_{mix} = \frac{W}{V} x \frac{\sum_{i=1}^n \frac{Q_i}{1 + \frac{W}{V \cdot K_i}}}{\sum_{i=1}^n Q_i - \sum_{i=1}^n \frac{Q_i}{1 + \frac{W}{V \cdot K_i}}}$$

Où W le volume de la phase aqueuse, V le volume de la phase lipidique, n le nombre de composés du mélange, et Q_i le nombre total de composés i dans le système.

- / **Les descripteurs intégraux additifs** sont définis sur la base d'un schéma additif et intègrent usuellement la notion de fraction molaire des composants d'un mélange chimique (Muratov et al. 2012). Cette stratégie de description physicochimique repose sur l'hypothèse qu'un mélange peut être représenté par une combinaison de descripteurs caractérisant les substances individuelles qui le composent. La combinaison entre les descripteurs est souvent pondérée par les fractions molaires qui déterminent le mélange. Voici deux exemples de calcul de descripteur du mélange (DM) :

$$\begin{aligned} DM &= R_1 D_1 + R_2 D_2 \cdot (\text{Ajmani et al. 2006}) \\ DM &= (R_1 D_1 + R_2 D_2)^2 \cdot (\text{Qin et al. 2021}) \end{aligned}$$

Où R1 la fraction molaire de la première substance, R2 la fraction molaire de la deuxième substance, D1 le descripteur caractérisant la première substance et D2 le descripteur caractérisant la deuxième substance.

/ A l'inverse des descripteurs additifs, **les descripteurs intégraux non-additifs** considèrent le mélange dans sa globalité. Selon Muratov *et al.* (2012), cette approche plus inclusive devrait augmenter ainsi la précision des modèles concernant les mélanges. Voici deux exemples de descripteur intégral non-additif pour des mélanges binaires :

$$\Delta TE = TE_M - (TE_i - TE_j)$$

Dans cette équation ΔTE est la différence de l'énergie totale³, TE_M est l'énergie totale du mélange binaire des substances individuelles i et j , et TE_i et TE_j représentent l'énergie totale des substances i et j respectivement.

$$Q_{mM} = \sum_i^n \frac{|q_i|}{n}$$

Dans cette équation Q_{mM} est la charge absolue moyenne d'un mélange binaire, q_i est la charge de l'atome i et n est le nombre d'atomes composant le mélange.

/ **Fragments structuraux non-additifs** : cette classe de descripteurs est basée sur la définition de fragments structuraux des composants du mélange chimique qui sont simultanément pris en compte au sein d'un même descripteur. Par exemple, Muratov *et al.* (2012) ont étendu aux mélanges chimiques l'approche de description moléculaire SiRMS⁴. Selon cette approche, toute molécule peut être représentée par un système de fragments tétratomiques appelés simplex et qui prennent en compte quatre atomes appartenant à la même molécule et chimiquement liés entre eux. La stratégie mise en place par Muratov *et al.* (2012) a permis d'étendre ce concept afin de pouvoir l'utiliser pour définir un mélange dans sa globalité. Pour ainsi faire, ces auteurs ont défini des simplex spéciaux toujours formés par quatre atomes qui peuvent ne pas être tous liés entre eux. Par exemple, dans le cas d'un mélange binaire, un « simplex mélange » peut être défini par deux fragments moléculaires non liés entre eux : un fragment composé par trois atomes appartenant à la première molécule du mélange et un fragment composé par un atome appartenant à la deuxième molécule du mélange.

La validation d'un modèle QSAR pour les mélanges chimiques

La validation des modèles QSAR de mélanges diffère de celle pour les substances individuelles. Trois approches ont été répertoriées (Muratov *et al.* 2012) :

/ « Points out » : d'un point de vue des substances composant les mélanges, chaque composition des mélanges est présente à la fois dans le jeu d'apprentissage et dans le jeu de validation. La différence

des mélanges au sein des deux jeux de données ne réside que dans les proportions des substances composant les mélanges. C'est une approche de validation peu rigoureuse car le modèle QSAR a été calibré en fonction de substances chimiques qui sont aussi présentes dans le jeu de validation. Cette caractéristique ne se conforme pas aux critères de validation communément appliqués car le jeu de validation et calibration contient les mêmes substances chimiques (Gramatica 2007).

/ « Mixtures out » : dans cette approche, un mélange ne peut pas être présent à la fois dans le jeu d'apprentissage et le jeu de validation, mais les substances individuelles peuvent être présentes au sein des deux jeux de données. Cette stratégie donne lieu à une validation externe plus rigoureuse que la précédente.

/ « Compounds out » : chaque mélange dans le jeu de validation contient au moins une substance chimique qui n'est pas présente dans le jeu d'apprentissage. Il s'agit de la méthode la plus rigoureuse de validation externe pour les mélanges chimiques. Il sera ainsi possible d'évaluer le pouvoir prédictif du modèle pour des mélanges composés par des « nouvelles substances » individuelles qui n'ont pas été considérées lors de la calibration du modèle.

Identification et documentation des modèles QSAR pour des mélanges chimiques

Notre travail de synthèse prend en considération les modèles QSAR prédisant directement la toxicité des mélanges chimiques en fonction d'un ensemble de descripteurs moléculaires⁵. Les articles analysés dans ce travail ont été identifiés en juin 2021 grâce aux moteurs de recherche « Google Scholar », « PubMed » et « ISI Web of Science », sans limites pour les dates de publication avec les combinaisons suivantes de mots clés :

- / chemical mixtures + QSAR ;
- / chemical mixtures + structure-activity ;
- / joint toxicity + QSAR ;
- / joint toxicity + structure-activity ;
- / chemical co-exposure + QSAR ;
- / chemical co-exposure et structure-activity.

Après lecture des résumés des articles, cette recherche a permis d'identifier 121 articles potentiellement pertinents, le plus ancien datant de 1984 et le plus récent de 2021. Après une lecture approfondie, il a été constaté que parmi ces articles, seuls 54 présentaient le développement d'un modèle QSAR prédisant directement la toxicité en fonction d'une description moléculaire.

Les articles écartés citaient le paradigme « structure-activité » sans proposer des modèles QSAR ou décrivaient l'application de la modélisation QSAR comme outil de paramétrisation pour d'autres méthodes computationnelles.

Analyse des articles sur la modélisation QSAR des mélanges chimiques

Organismes vivants

L'analyse du corpus bibliographique met en évidence que la majorité des modèles QSAR développés (Fig. 1) concerne des espèces d'invertébrés (67 modèles) et, parmi les invertébrés, les études ayant pris en compte les bactéries sont clairement majoritaires (54 modèles). Les études prenant en compte les vertébrés sont minoritaires avec six articles uniquement : les études de Gaskill et Bruce (2016) et de Hoover *et al.* (2019) portant, respectivement, sur des hépatocytes de rat et des fibroblastes de l'amphibien *Xenopus tropicalis*, l'étude de Huang *et al.* (2003) réalisée sur des têtards (*Rana japonica*), l'étude de Kar *et al.* (2018) sur les embryons de *Danio rerio* et les études de Sheridan (2000) et Zakharov *et al.* (2016) qui analysent des effets pharmacologiques⁶ chez l'Homme.

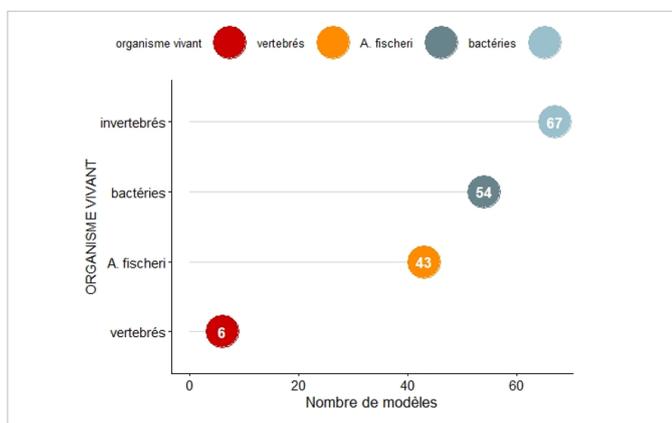


Figure 1 / Organismes vivants considérés par les articles analysés. La majorité des modèles QSAR décrits dans l'ensemble des articles analysés a pris en compte des essais conduits chez des espèces d'invertébrés et, notamment, sur les bactéries. Parmi les bactéries, *Aliivibrio fischeri* représente l'espèce la plus étudiée.

Parmi tous les organismes vivants, une place particulière est occupée par *A. fischeri*⁷. Cette bactérie marine est l'organisme vivant le plus étudié (43 articles, Fig. 2). *A. fischeri* possède une bioluminescence naturelle qui rend possible des essais de toxicité aigüe rapides (effets fréquemment exprimés après 30 minutes). Par conséquent, les essais de toxicité basés sur cette espèce permettent l'analyse de l'effet de plusieurs mélanges à un coût relativement faible.

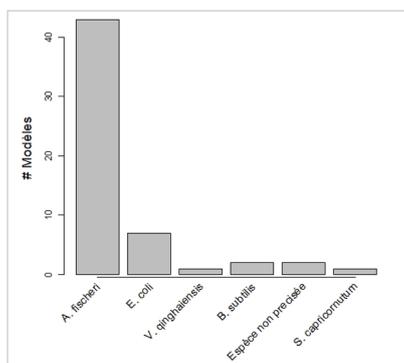


Figure 2 / Espèces bactériennes prises en compte par les articles analysés.

Parmi les autres espèces d'invertébrés, il faut signaler cinq articles détaillant des modèles QSAR chez les algues

microscopiques (Lu *et al.* 2007, Wang *et al.* 2008, Lu *et al.* 2009, Chang *et al.* 2016, Qin *et al.* 2021) et un article chez les abeilles (Carneseccchi *et al.* 2020).

En ce qui concerne les effets toxicologiques modélisés, il faut signaler que presque la totalité des essais de toxicité (ex. reproduction, bioluminescence, cytotoxicité) ont été conduits en fonction d'une courte durée d'exposition (de 15 minutes à quelques jours). En effet, ces courtes durées d'expositions réduisent le temps nécessaire pour l'analyse de plusieurs mélanges.

Cette analyse des publications conduite en fonction des organismes vivants considérés met en évidence l'absence de modélisation QSAR pour des effets relevés habituellement *in vivo* chez les rongeurs. Cette absence est très probablement due à la faible disponibilité de données caractérisant la toxicité *in vivo* des mélanges chimiques.

Familles chimiques

Les familles chimiques étudiées dans les articles analysés recouvrent plusieurs séries de composés organiques (ex. aldéhydes, benzènes halogènes, anilines...) et des métaux lourds (cadmium, plomb, cuivre, cobalt, chrome). Plusieurs catégories de substances chimiques d'intérêt environnemental ont été prises en compte pour la dérivation de modèles : pesticides (Qin *et al.* 2018, Sigurnjak Bureš *et al.* 2021), fongicides (Carneseccchi *et al.* 2020, Pizzolitto *et al.* 2020), insecticides (Carneseccchi *et al.* 2020), herbicides (Yao *et al.* 2013, Wang *et al.* 2018b, Chatterjee and Roy 2021, Toropova and Toropov 2021), composés perfluorés (Luan *et al.* 2013, Wang *et al.* 2017, Kar *et al.* 2018, Hoover *et al.* 2019), et antibiotiques (Yao *et al.* 2013, Fang *et al.* 2016, Long *et al.* 2016, Wang *et al.* 2017, Wang *et al.* 2018a, Wang *et al.* 2018b, Wang *et al.* 2020, Sigurnjak Bureš *et al.* 2021).

Descripteurs moléculaires

Du point de vue de la description physico-chimique des mélanges, la majorité des articles analysés a adopté l'approche des descripteurs intégraux additifs (Fig. 3). Ainsi les autres trois approches décrites par Muratov *et al.* (2012) restent minoritaires. Une minorité d'articles a adopté des approches non décrites par Muratov *et al.* (2012) comme par exemple, la dérivation de plusieurs modèles QSAR pour des substances organiques individuelles mais en présence de concentrations différentes de métaux lourds.

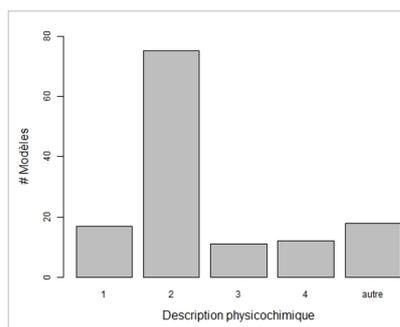


Figure 3 / Approches de description physico-chimique adoptées par les articles analysés. (1) : Descripteurs basés sur le coefficient de partage, (2) : Descripteurs intégraux additifs, (3) : Descripteurs intégraux non additifs, (4) : Fragments structuraux non additifs.

Dans ce cas de figure, la description physico-chimique adoptée par les modèles QSAR ne concerne que la composante organique, et la composante métallique est prise en compte de façon implicite. Cette prise en compte implicite a été réalisée en dérivant trois équations QSAR qui prennent en compte une description physico-chimique des substances organiques mais en présence de concentrations différentes de métaux lourds (concentration faible, moyenne et haute).

Approches de modélisation

Sur la base des approches de modélisation statistique adoptées par les auteurs des articles analysés, il apparaît que les algorithmes les plus utilisés sont linéaires (régression linéaire simple, multiple ou par les moindres carrés partiels) tandis que les algorithmes non-linéaires (ex. forêts aléatoires réseaux de neurones) représentent la stricte minorité des approches (Fig. 4). Ces algorithmes de modélisation ne présentent pas de particularité par rapport à ceux normalement utilisés lors de la modélisation QSAR des effets toxicologiques induits par les substances individuelles.

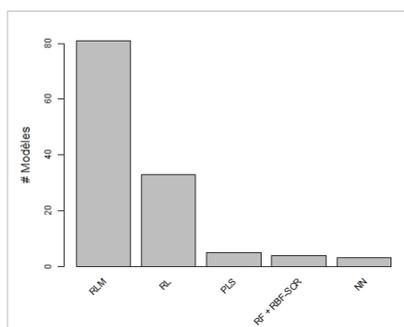


Figure 4 / Algorithme de modélisation adoptés dans les articles analysés. RLM = régression Linéaire Multiple, RL = régression Linéaire simple, PLS = régression par les moindres carrés partiels (Partial Least Squares), RF = forêt aléatoires (Random Forests), RBF-SCR = réseaux de neurones à fonctions de base radiales, NN = réseaux de neurones.

Validation des modèles

A l'égard de la précision prédictive, les modèles QSAR sont généralement considérés comme étant valides d'un point de vue interne si l'indicateur statistique Q^2 (variance expliquée estimée par validation croisée) est égal ou supérieur à 0,6 (Chirico and Gramatica 2011). La grande majorité des modèles QSAR (71 modèles sur 79) pour lesquels cet indicateur a été calculé, atteint ou dépasse ce seuil critique.

Une proportion importante des modèles (49 %) n'a pas été qualifiée en ce qui concerne la validation externe. En revanche, la grande majorité des modèles qualifiés en fonction de l'indicateur statistique⁸ Q^2_{EXT} franchisse ou dépasse le seuil critique de validité (0,7) communément adopté en littérature QSAR (Chirico and Gramatica 2012).

La stricte minorité de modèles QSAR de classification est associée avec une performance prédictive caractérisée par une concordance comprise entre 0,68 et 0,85. Pour les modèles de classification, il n'y a pas de valeurs critiques de validité préconisées par la littérature scientifique et un jugement sur leur pertinence est élaboré au cas par cas selon l'utilisation envisagée.

Il faut cependant observer que pour une proportion non négligeable d'articles la performance prédictive a été éva-

luée en fonction d'un nombre peu élevé de mélanges et substances :

- / Dans plusieurs cas, le nombre de mélanges pris en compte dans les articles analysés n'était pas élevé. La médiane du nombre de mélanges considérés est égale à 23 et un quart des modèles QSAR a été dérivé avec moins de 11 mélanges.
- / Le nombre de substances chimiques pris en compte dans les articles analysés donne lieu à un constat similaire : la médiane du nombre de substances chimiques considérées est égale à 12 et un quart des articles analysés a pris en compte moins de 8 substances chimiques.

La validation externe la plus répandue est l'approche « points out » (Fig. 5, histogramme A) qui est la moins stricte. Par conséquent, dans la majorité des cas le pouvoir prédictif externe des modèles a pu être surestimé.

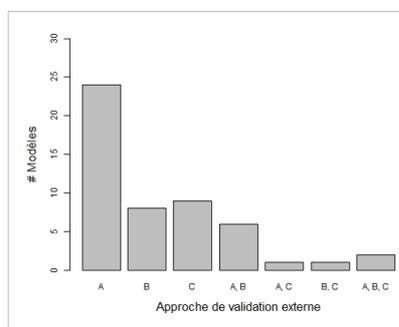


Figure 5 / Approches de validation externe adoptées dans les articles analysés. A = « points out », B = « mixtures out », C = « compounds out ».

Conclusions

L'utilisation de modèles QSAR pour prédire la toxicité liée à des mélanges de substances chimiques suscite un intérêt grandissant. Cette approche *in silico* permet d'évaluer les effets toxiques sans avoir à recourir à de nouveaux essais *in vitro* ou *in vivo*.

Notre travail bibliographique a permis d'évaluer l'avancée du développement des modèles QSAR pour la prédiction de l'(éco)toxicité des mélanges chimiques. Les modèles QSAR publiés étudient des mélanges peu complexes dont la majorité (69 %) sont des mélanges binaires. Il faut aussi souligner que la grande majorité des modèles analysés reste focalisée sur les dangers écotoxicologiques (notamment chez les bactéries) caractérisables grâce à des essais rapides et économiques, reléguant ainsi la toxicologie à une minorité de modèles. Pour déployer des modèles en toxicologie, il serait nécessaire de générer de nouvelles données mais la priorité aujourd'hui est de tester la faisabilité sur des organismes pour lesquels les coûts et exigences éthiques sont raisonnables.

La grande majorité des modèles QSAR analysés est caractérisée par une performance prédictive comparable à celle habituellement associée aux modèles QSAR traitant les substances individuelles. Ce constat semblerait rester valable même si on exclut les modèles validés selon l'approche moins rigoureuse (approche « points out ») même

si le nombre réduit de modèles restants ne permet pas de conclure avec certitude sur cet aspect. Il semblerait donc que la modélisation QSAR des mélanges demeure une problématique pouvant être résolue grâce aux techniques computationnelles communément utilisées mais que la disponibilité de données, en quantité et qualité suffisantes, serait le véritable facteur limitant un développement plus étendu.

Vis-à-vis de ce facteur limitant, il apparaît que la calibration de modèles QSAR en fonction de données *in vitro* pouvant être générées à moyen ou haut débit (la tendance actuelle en toxicologie) représenterait la perspective de travail la plus apte à répondre au besoin en informations sous-jacent à la modélisation QSAR. Afin de faciliter les applications réglementaires, cette perspective de travail devrait s'inscrire au sein du concept des Adverse Outcome Pathways (AOP, (Leist *et al.* 2017)) (« chemins de l'effet néfaste » en français) et se focaliser sur la modélisation des événements-clés composant ces derniers. En particulier, la possibilité de prendre en compte des données caractérisant des événements-clés à l'intersection parmi plusieurs AOP permettrait d'augmenter la pertinence biologique et l'utilité des modèles QSAR pouvant être développés.

D'un point de vue pratique, il n'existe pas à ce jour d'outils QSAR informatisés disponibles en ligne ou par téléchargement pouvant traiter des mélanges chimiques. L'application éventuelle des modèles publiés devrait donc s'appuyer sur une reproduction des modèles QSAR à partir des publications et cet exercice nécessite de solides connaissances en modélisation QSAR ainsi qu'un certain investissement.

Par conséquent, en l'état actuel et d'un point de vue réglementaire, l'approche prédictive non-expérimentale la plus pratique pour la caractérisation de la toxicité des mélanges chimiques reste l'approche par lecture croisée⁹ pouvant être argumentée au cas par cas sur des bases essentiellement mécanistiques (Patlewicz *et al.* 2014).

Références

- / Ajmani, S., S. C. Rogers, M. H. Barley and D. J. Livingstone (2006). «Application of QSPR to Mixtures.» *Journal of Chemical Information and Modeling* 46(5): 2043-2055.
- / Carnesecchi, E., A. A. Toropov, A. P. Toropova, N. Kramer, C. Svendsen, J. L. Dorne and E. Benfenati (2020). «Predicting acute contact toxicity of organic binary mixtures in honey bees (*A. mellifera*) through innovative QSAR models.» *Science of The Total Environment* 704: 135302.
- / Chang, C. M., Y. H. Ou, T.-C. Liu, S.-Y. Lu and M.-K. Wang (2016). «A quantitative structure–activity relationship approach for assessing toxicity of mixture of organic compounds.» *SAR and QSAR in Environmental Research* 27(6): 441-453.
- / Chatterjee, M. and K. Roy (2021). «Prediction of aquatic toxicity of chemical mixtures by the QSAR approach using 2D structural descriptors.» *Journal of Hazardous Materials* 408: 124936.
- / Chirico, N. and P. Gramatica (2011). «Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient.» *Journal of Chemical Information and Modeling* 51(9): 2320-2335.
- / Chirico, N. and P. Gramatica (2012). «Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection.» *J Chem Inf Model* 52(8): 2044-2058.
- / Fang, S., D. Wang, X. Zhang, X. Long, M. Qin, Z. Lin and Y. Liu (2016). «Similarities and differences in combined toxicity of sulfonamides and other antibiotics towards bacteria for environmental risk assessment.» *Environ Monit Assess* 188(7): 429.
- / Gaskill, S. J. and E. D. Bruce (2016). «Binary Mixtures of Polycyclic Aromatic Hydrocarbons Display Nonadditive Mixture Interactions in an *In Vitro* Liver Cell Model.» *Risk Anal* 36:968-991.
- / Gramatica, P. (2007). «Principles of QSAR models validation: internal and external.» *Qsar & Combinatorial Science* 26(5): 694-701.
- / Heys, K. A., R. F. Shore, M. G. Pereira, K. C. Jones and F. L. Martin (2016). «Risk assessment of environmental mixture effects.» *RSC Advances* 6(53): 47844-47857.
- / Hoover, G., S. Kar, S. Guffey, J. Leszczynski and M. S. Sepúlveda (2019). «In vitro and in silico modeling of perfluoroalkyl substances mixture toxicity in an amphibian fibroblast cell line.» *Chemosphere* 233: 25-33.
- / Huang, H. Wang, X. Shao, Y. Chen, Dai, D. Wang, L. (2003). «QSAR for prediction of joint toxicity of substituted phenols to tadpoles (*Rana japonica*)» *Bull Environ Contam Toxicol* 71:1124-1130.
- / Kar, S., S. Ghosh and J. Leszczynski (2018). «Single or mixture halogenated chemicals? Risk assessment and developmental toxicity prediction on zebrafish embryos based on weighted descriptors approach.» *Chemosphere* 210: 588-596.
- / Leist, M., A. Ghallab, R. Graepel, R. Marchan, R. Hassan, S. H. Bennekou, A. Limonciel, M. Vinken, S. Schildknecht, T. Waldmann, E. Danen, B. van Ravenzwaay, H. Kamp, I. Gardner, P. Godoy, F. Y. Bois, A. Braeuning, R. Reif, F. Oesch, D. Drasdo, S. Hohme, M. Schwarz, T. Hartung, T. Braunbeck, J. Beltman, H. Vrieling, F. Sanz, A. Forsby, D. Gadaleta, C. Fisher, J.

- Kelm, D. Fluri, G. Ecker, B. Zdrzil, A. Terron, P. Jennings, B. van der Burg, S. Dooley, A. H. Meijer, E. Willighagen, M. Martens, C. Evelo, E. Mombelli, O. Taboureau, A. Mantovani, B. Hardy, B. Koch, S. Escher, C. van Thriel, C. Cadenas, D. Kroese, B. van de Water and J. G. Hengstler (2017). «Adverse outcome pathways: opportunities, limitations and open questions.» *Arch Toxicol* 91(11): 3477-3505.
- / Long, X., D. Wang, Z. Lin, M. Qin, C. Song and Y. Liu (2016). «The mixture toxicity of environmental contaminants containing sulfonamides and other antibiotics in *Escherichia coli*: Differences in both the special target proteins of individual chemicals and their effective combined concentration.» *Chemosphere* 158: 193-203.
- / Lu, G., C. Wang, Z. Tang and X. Guo (2007). «Joint toxicity of aromatic compounds to algae and QSAR study.» *Ecotoxicology* 16(7): 485-490.
- / Lu, G. H., C. Wang, P. F. Wang and C. Z. Yang (2009). «Predicting toxicity of aromatic ternary mixtures to algae.» *Chinese Sci Bul* 54: 3521-3527.
- / Luan, F., X. Xu, H. Liu and M. N. D. S. Cordeiro (2013). «Prediction of the baseline toxicity of non-polar narcotic chemical mixtures by QSAR approach.» *Chemosphere* 90(6): 1980-1986.
- / Mombelli, E. and S. Ringeissen (2009). «The computational prediction of toxicological effects in regulatory contexts Current use and future potential of (Q)SAR tools.» *Actualite Chimique*(335): 52-59.
- / Muratov, E. N., E. V. Varlamova, A. G. Artemenko, P. G. Polishchuk and V. E. Kuz'min (2012). «Existing and Developing Approaches for QSAR Analysis of Mixtures.» *Mol Inform* 31(3-4): 202-221.
- / Patlewicz, G., N. Ball, R. A. Becker, E. D. Booth, M. T. Cronin, D. Kroese, D. Steup, B. van Ravenzwaay and T. Hartung (2014). «Read-across approaches - misconceptions, promises and challenges ahead.» *Altex* 31(4): 387-396.
- / Pizzolitto, R. P., A. G. Jacquat, V. L. Usseglio, F. Achimón, A. E. Cuello, J. A. Zygodlo and J. S. Dambolena (2020). «Quantitative-structure-activity relationship study to predict the antifungal activity of essential oils against *Fusarium verticillioides*.» *Food Control* 108: 106836.
- / Qin, L.-T., Y.-H. Chen, X. Zhang, L.-Y. Mo, H.-H. Zeng and Y.-P. Liang (2018). «QSAR prediction of additive and non-additive mixture toxicities of antibiotics and pesticide.» *Chemosphere* 198: 122-129.
- / Qin, L.-T., M. Liu, X. Zhang, L.-Y. Mo, H.-H. Zeng and Y.-P. Liang (2021). «Concentration Addition, Independent Action, and Quantitative Structure-Activity Relationships for Chemical Mixture Toxicities of the Disinfection By products of Haloacetic Acids on the Green Alga *Raphidocelis subcapitata*.» *Environmental toxicology and chemistry* 40(5): 1431-1442.
- / Roy, K., S. Kar and R. N. Das (2015). Chapter 1 - Background of QSAR and Historical Developments. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. K. Roy, S. Kar and R. N. Das. Boston, Academic Press: 1-46.
- / Russell, W. M. S. (1995). «The Development of the Three Rs Concept.» *Alternatives to Laboratory Animals* 23(3): 298-304.
- / Sheridan, R. P. (2000). «The Centroid Approximation for Mixtures: Calculating Similarity and Deriving Structure-Activity Relationships.» *Journal of Chemical Information and Computer Sciences* 40(6): 1456-1469.
- / Sigurnjak Bureš, M., Š. Ukić, M. Cvetnić, V. Prevarić, M. Markić, M. Rogošić, H. Kušić and T. Bolanča (2021). «Toxicity of binary mixtures of pesticides and pharmaceuticals toward *Vibrio fischeri*: Assessment by quantitative structure-activity relationships.» *Environmental Pollution* 275: 115885.
- / Sillé, F. C. M., S. Karakitsios, A. Kleensang, K. Koehler, A. Maertens, G. W. Miller, C. Prasse, L. Quiros-Alcala, G. Ramachandran, S. M. Rappaport, A. M. Rule, D. Sarigiannis, L. Smirnova and T. Hartung (2020). «The exposome - a new approach for risk assessment.» *ALTEX* 37(1): 3-23.
- / Sprinkle, R. H. and D. C. Payne-Sturges (2021). «Mixture toxicity, cumulative risk, and environmental justice in United States federal policy, 1980-2016.» *Environmental Health* 20(1): 104.
- / Toropova, A. P. and A. A. Toropov (2021). «Can the Monte Carlo method predict the toxicity of binary mixtures?» *Environmental Science and Pollution Research*.
- / Wang, C., G. Lu, Z. Tang and X. Guo (2008). «Quantitative structure-activity relationships for joint toxicity of substituted phenols and anilines to *Scenedesmus obliquus*.» *Journal of Environmental Sciences* 20(1): 115-119.
- / Wang, D., Y. Gu, M. Zheng, W. Zhang, Z. Lin and Y. Liu (2017). «A Mechanism-based QSTR Model for Acute to Chronic Toxicity Extrapolation: A Case Study of Antibiotics on Luminous Bacteria.» *Scientific Reports* 7(1): 6022.
- / Wang, D., Q. Ning, J. Dong, B. W. Brooks and J. You (2020). «Predicting mixture toxicity and antibiotic resistance of fluoroquinolones and their

photodegradation products in *Escherichia coli*.»
Environ Pollut 262: 114275.

- / Wang, D., J. Shi, Y. Xiong, J. Hu, Z. Lin, Y. Qiu and J. Cheng (2018a). «A QSAR-based mechanistic study on the combined toxicity of antibiotics and quorum sensing inhibitors against *Escherichia coli*.» *Journal of Hazardous Materials* 341: 438-447.
- / Wang, T., L. Tang, F. Luan and M. N. D. S. Cordeiro (2018b). «Prediction of the Toxicity of Binary Mixtures by QSAR Approach Using the Hypothetical Descriptors.» *International Journal of Molecular Sciences* 19(11): 3423.
- / Willey, J. B., T. Pollock, E. M. Thomson, C. L. Liang, A. Maquiling, M. Walker and A. St-Amand (2021).

«Exposure Load: Using biomonitoring data to quantify multi-chemical exposure burden in a population.»
International Journal of Hygiene and Environmental Health 234: 113704.

- / Yao, Z., Z. Lin, T. Wang, D. Tian, X. Zou, Y. Gao and D. Yin (2013). «Using molecular docking-based binding energy to predict toxicity of binary mixture with different binding sites.» *Chemosphere* 92(9): 1169-1176.
- / Zakharov, A. V., E. V. Varlamova, A. A. Lagunin, A. V. Dmitriev, E. N. Muratov, D. Fourches, V. E. Kuz'min, V. V. Poroikov, A. Tropsha and M. C. Nicklaus (2016). «QSAR Modeling and Prediction of Drug-Drug Interactions.» *Molecular Pharmaceutics* 13(2): 545-556.

¹ La validation croisée estime la fiabilité d'un modèle en fonction d'un échantillonnage du jeu d'apprentissage. Lors de cette validation, ce jeu de données est divisé en k blocs et k-1 blocs sont utilisés pour recalibrer le modèle. Après cette recalibration, la performance prédictive du modèle recalibré est évaluée en fonction du bloc exclu. Cette procédure de subdivision du jeu d'apprentissage est répétée plusieurs fois et la valeur de la performance prédictive moyenne estimée sur les blocs exclus est retenue comme indicateur de la fiabilité « interne » du modèle.

² Le coefficient de partage est une propriété sans dimension, qui décrit comment une substance chimique se répartie entre deux solvants (ex. l'eau et l'octanol, Log P ou Log K_{ow}). Cette propriété physico-chimique figure souvent parmi les descripteurs moléculaires adoptés par les modèles QSAR.

³ L'énergie totale d'un système est subdivisée en énergies cinétiques de tous les atomes et en énergies potentielles provenant des interactions entre ces derniers.

⁴ Simplex representation of molecular structure.

⁵ Les méthodes QCAR (quantitative composition-activity relationships) n'ont pas été considérées dans le cadre de ce rapport parce qu'elles ne sont pas basées sur l'utilisation de descripteurs moléculaires.

⁶ L'article de Sheridan (2004) est axé autour de l'analyse d'une base de données répertoriant des catégories thérapeutiques pharmacologiques variées comme, par exemple, les antagonistes de la vasopressine ou de la gastrine. L'article de Zakharov et al. (2016) propose des modèles QSAR pour la prédiction d'interactions médicamenteuses.

⁷ Les articles analysés reportent plusieurs noms pour cette bactérie : *Vibrio fischeri*, *Allivibrio fischeri* et *Photobacterium phosphoreum* Il s'agit cependant de la même espèce désormais appelée *A. fischeri*.

⁸ Cet indicateur statistique transpose l'erreur de prédiction sur une échelle allant de $-\infty$ à 1 (modèle parfait).

⁹ L'approche prédictive par lecture croisée (« read-across » en anglais) se base sur la même hypothèse que les modèles QSAR (des substances chimiques similaires induisent des effets biologiques similaires). Cependant, les analogies structurelles se limitent à un nombre restreint de substances (ou mélanges chimiques). On peut voir la lecture croisée comme un modèle QSAR limité à un espace chimique très restreint dont la pertinence et logique prédictive sont argumentés par un avis d'expert plutôt que par une validation statistique conduite sur un jeu de données plus étendu. C'est une approche prédictive qui doit être argumentée au cas par cas sur la base d'une description détaillée des mécanismes toxicologiques et profils physico-chimiques sous-jacents à la problématique à résoudre. Cette approche repose davantage sur des compétences en toxicologie plutôt que sur des compétences en chémoinformatique.