



HAL
open science

La modélisation QSAR

Enrico Mombelli

► **To cite this version:**

Enrico Mombelli. La modélisation QSAR. Rapport Scientifique INERIS, 2008, 2007-2008, pp.36-37.
ineris-01869194

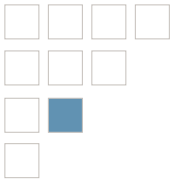
HAL Id: ineris-01869194

<https://ineris.hal.science/ineris-01869194>

Submitted on 6 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



RECh, LES ENJEUX DE LA RECHERCHE

La modélisation QSAR

> ENRICO MOMBELLI

#1

> Lors d'une analyse QSAR les structures moléculaires des composés chimiques sont traitées mathématiquement afin d'obtenir des descripteurs moléculaires qui codent les divers éléments structuraux des substances en cours d'évaluation (exemple : topologie, géométrie, propriétés quantiques). Les valeurs numériques de ces descripteurs sont utilisées pour établir une équation qui permet de modéliser la corrélation statistique entre la variation de la structure moléculaire et l'effet toxique investigué.



FORMALISATION
MATHÉMATIQUE



DE LA STRUCTURE
MOLÉCULAIRE

- ➔ Propriétés Physico-chimiques
- ➔ Descripteurs Topologiques
- ➔ Champs moléculaires stériques et électrostatiques
- ➔ Descripteurs géométriques
- ➔ Descripteurs quantiques

La modélisation semi-empirique QSAR [(Quantitative) Structure Activity Relationship] a comme objectif la prédiction des effets d'une variation de la structure moléculaire sur l'activité biologique (exemple : relation entre structure moléculaire et propriétés mutagéniques). Cette modélisation peut être quantitative (QSAR) ou qualitative (SAR). La première stratégie de modélisation prévoit trois éléments :

- a) un ou plusieurs descripteurs de la structure moléculaire (figure 1) ;
- b) un effet à prédire ;
- c) une relation mathématique permettant de décrire la corrélation statistique entre les descripteurs moléculaires et l'effet biologique à modéliser. La relation mathématique est d'habitude établie grâce à des méthodes statistiques multivariées.

LA PREMIÈRE FORMALISATION

La première formalisation d'une activité biologique en fonction de la structure chimique est historiquement attribuée à Hansch. Dans un travail ayant fait école^[1], Hansch proposa une équation permettant de modéliser l'activité biologique en fonction de l'hydrophobicité et les caractéristiques électroniques du benzène grâce à cette relation :

$$\log(1/C) = k_1 \log P - k_2 (\log P)^2 + k_3 \sigma + k_4$$

où C est la concentration de la substance qui est nécessaire pour produire un niveau défini de réponse biologique, Log P est le logarithme du coefficient de partage entre l'n-octanol et l'eau, σ est le paramètre d'Hammett et k_1 - k_4 sont des constantes. Le paramètre d'Hammett est une mesure de l'effet électronique des substituants électro-capteurs ou électro-donneurs et Log P modélise la capacité d'un composé à diffuser dans les membranes lipidiques. En effet, Hansch remarqua qu'il y a une valeur optimale pour Log P : si elle est trop grande, le composé chimique restera à l'intérieur de la membrane ; si elle est

trop petite, le composé chimique restera dans la phase aqueuse.

En revanche, les modèles qualitatifs sont basés sur la reconnaissance d'alertes structurelles qui sont d'habitude associées à l'effet toxique investigué.

Un exemple d'alertes structurelles est donné sur la figure 2.

« SYSTÈMES EXPERTS »

Il existe également plusieurs logiciels commerciaux, qui sont communément qualifiés de « Systèmes Experts », et qui peuvent être regroupés en quatre catégories : QSAR prêts à l'emploi, systèmes basés sur l'implémentation informatisée de l'expertise toxicologique (reconnaissance de toxicophores), systèmes à même de définir et reconnaître de façon automatique les toxicophores, systèmes basés sur des arbres décisionnels^[2].

Dans leur globalité, les QSAR constituent une méthode alternative à l'expérimentation animale reconnue par le règlement européen RECh^[3] qui présente les avantages de réduire soit le nombre d'animaux nécessaires à l'expérimentation toxicologique soit les coûts relatifs à l'implémentation de stratégies de test^[4,5].

Cette possibilité d'utilisation réglementaire pose soit le problème de la fiabilité de ces modèles soit le problème relatif à une interprétation correcte des prédictions. Pour y répondre, l'OCDE a publié, en février 2007, un guide technique pour la validation des modèles QSAR dans un contexte réglementaire. Ce guide détaille cinq principes qui visent à établir une base commune pour une utilisation correcte des modèles QSAR lors de l'estimation de propriétés toxicologiques :

- 1/ L'effet toxicologique modélisé doit être bien défini.
- 2/ L'algorithme qui a généré les prédictions doit être transparent afin que la logique à la base de la modélisation puisse être facilement reproductible.

- 3/ L'utilisation du modèle doit être limitée à son domaine d'application. Autrement dit, les substances chimiques que le modèle peut analyser doivent avoir une structure similaire aux substances utilisées lors de sa paramétrisation.
- 4/ La prédiction doit être accompagnée par des mesures statistiques appropriées qui décrivent la robustesse, l'ajustement et la prédictivité du modèle.
- 5/ Une interprétation mécaniste du modèle doit être donnée si une telle rationalisation est possible.

Les quatre premiers critères sont impératifs pour l'acceptation des prédictions QSAR. Le cinquième reste facultatif mais son interprétation correcte permet une consolidation épistémologique de la pertinence du modèle^[6-9] qui pourrait se révéler cruciale dans un contexte réglementaire.

DES MODÈLES SEMI-EMPIRIQUES

Les QSAR sont des modèles semi-empiriques. Leur pouvoir prédictif et leur domaine d'application dépendent de la composition du jeu d'apprentissage utilisé pour calibrer le modèle. Un modèle QSAR parfait ne pourra donc pas permettre une prédiction pour toute sorte de substance et avec un niveau de précision supérieure à celui de la méthode expérimentale qui a généré les données du jeu d'apprentissage. La nécessité d'avoir accès à des bases de données de qualité, qui contiennent une proportion comparable de substances toxiques et inoffensives, est donc primordiale pour le développement de modèles QSAR ayant une bonne spécificité et sensibilité.

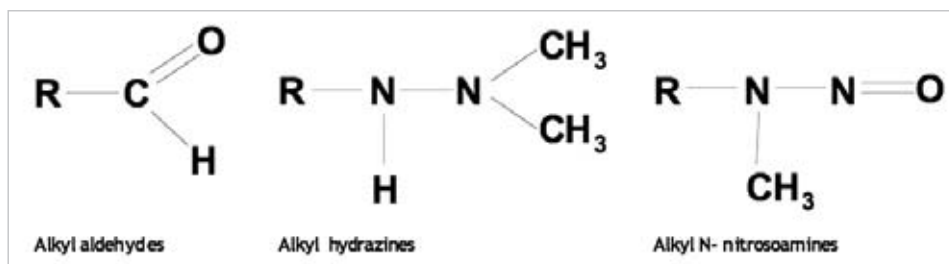
L'expertise QSAR développée à l'INERIS^[9] a mis en évidence les limites du pouvoir prédictif des

logiciels QSAR qui sont censés permettre une prédiction pour plusieurs familles chimiques. En effet, la généralisation d'un modèle QSAR pour la modélisation de catégories moléculaires qui sont très hétérogènes en termes de structure correspond à une perte de spécificité par rapport à la capacité à discriminer les molécules toxiques et les molécules inoffensives. Parallèlement à ce constat, ces études ont également révélé deux autres aspects importants : premièrement, la pleine conformité aux critères de l'OCDE n'est possible que pour une minorité des modèles et, deuxièmement, la définition rigoureuse du domaine d'application du modèle est primordiale pour une interprétation correcte des résultats.

D'un point de vue plus général, il est important de noter que la quasi-totalité des modèles QSAR disponibles en littérature ne peuvent pas modéliser la toxicité des mélanges chimiques. Pour cette raison, les efforts de recherche dans le domaine de la modélisation Structure-Activité sont en train d'étendre leur domaine d'application vers l'analyse des mélanges chimiques afin de prendre en compte les phénomènes de synergie ou d'antagonisme parmi les composants toxiques d'un mélange chimique^[10-12].

Dans un futur proche, les modèles de toxicologie prédictive *in silico* connaîtront un développement important dont les jalons les plus importants seront l'implémentation de modèles avec une prédictivité rigoureusement vérifiée pour tous les effets toxicologiques et une analyse exhaustive de ces effets pour les principales familles chimiques. La réalisation de ces objectifs suppose une collaboration étroite et efficace entre le secteur industriel, les instances réglementaires et le milieu académique.

#2



> Trois alertes structurelles pour la mutagénicité d'après Ashby et Tennant^[13]. La présence de ces alertes à l'intérieur d'une molécule est souvent associée avec des propriétés mutagéniques. Leur identification lors d'un criblage virtuel d'une base de données permet d'identifier les molécules potentiellement toxiques de telle sorte qu'elles puissent servir de base à une investigation toxicologique plus approfondie (optimisation des ressources expérimentales).

[RÉFÉRENCES

- [1] Hansch C., 1969. *A quantitative approach to biochemical structure-activity relationships*. Accounts of chemical research, 2, 232-239.
- [2] Dearden J. C., 2003. *In silico prediction of drug toxicity*. Journal of Computer-Aided Molecular Design 2003, 17, 119-27.
- [3] Barratt M., 2003. *QSAR, Read-across and REACH*. Alternatives to laboratory animals, 31, 463-5.
- [4] Pedersen F., de Bruijn J., Munn S. and Van Leeuwen K., 2003. *Assessment of additional testing needs under REACH*. (http://ihcp.jrc.ec.europa.eu/docs/ecb/reach_testing_needs.pdf)
- [5] ECVAM, *REACH and the need for Intelligent Testing Strategies*. (ihcp.jrc.ec.europa.eu/docs/20051107its.pdf) Issued by the Institute for Health and Consumer Protection, 2005.
- [6] Serafimova R., Todorov M., Nedelcheva D., Pavlov T., Akahori Y., Nakai M. and Mekenyan O., 2007. *QSAR and mechanistic interpretation of estrogen receptor binding*. SAR QSAR Environ Res, 18, 389-421.
- [7] Roberts D. W., Patlewicz G., Kern P. S., Gerberick F., Kimber I., Dearman R. J., Ryan C. A., Basketter D. A. and Aptula A. O., 2007. *Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization*. Chem Res Toxicol, 20, 1019-30.
- [8] Benigni R., Bossa C., Netzeva T. and Worth A., 2007. *Collection and Evaluation of QSAR Models for Mutagenicity and Carcinogenicity*. EUR 22623 EN. (<http://ecb.jrc.it/qsar/publications/>)
- [9] Mombelli E., 2008. *An evaluation of the predictive ability of the QSAR software packages*. DEREK, HAZARDEXPERT and TOPKAT, to describe chemically-induced skin irritation. Alternatives to laboratory animals, 36, 15-24.
- [10] Riviere J. E. and Brooks J. D., 2007. *Prediction of dermal absorption from complex chemical mixtures: incorporation of vehicle effects and interactions into a QSPR framework*. SAR QSAR Environ Res, 18, 31-44.
- [11] Wang B., Yu G., Hu H. and Wang L., 2007. *Quantitative structure-activity relationships and mixture toxicity of substituted benzaldehydes to Photobacterium phosphoreum*. Bull Environ Contam Toxicol, 78, 503-9.
- [12] Zhang L., Zhou P. J., Yang F. and Wang Z. D., 2007. *Computer-based QSARs for predicting mixture toxicity of benzene and its derivatives*. Chemosphere, 67, 396-401.
- [13] Ashby J. and Tennant R. W., 1991. *Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP*. Mutat Res, 257, 229-306.