

## CALCULATION OF COEFFICIENT OF SHARING OCTANOL-WATER OF ORGANIC COMPOUNDS USING MOLECULAR DESCRIPTORS

B. Souyei

VPRS, Department of Chemical Engineering, Kasdi Merbah University, PO Box 511,  
Ouargla, Algeria

Received: 25 October 2010 / Accepted: 06 December 2010 / Published online: 31 December 2010

### ABSTRACT

A quantitative structure-property relationship (QSPR) study is carried out to develop correlations that relate the molecular structures of organic compounds to their Octanol-Water partition coefficients,  $K_{ow}$ , using molecular descriptors. The correlations are simple in application with good accuracy, which provide an easy, direct and relatively accurate way to calculate  $K_{ow}$ . Such calculation gives us a model that gives results in remarkable correlation with the descriptors of blocks fragments of the atom-centered and functional groups ( $R^2 = 0.949$ ,  $r = 0.974$ ) ( $R^2 = 0.926$ ,  $r = 0.962$ ) respectively.

**Keywords:** QSPR, Model, Activity coefficient, Infinite dilution, aqueous solution, Molecular descriptors.

### 1. INTRODUCTION

Aujourd'hui, l'industrie chimique joue un rôle majeur et divers dans notre vie, ce rôle se diffère suivant la nature des composés chimiques utilisés. Il existe plusieurs variétés de composés chimiques classifiés comme produits dangereux (Toxiques), qui ont un impact sérieux sur la santé de l'homme et de l'environnement. Parmi ces composés, on cite: les organochlorés comme les PCB (polychlorobiphényles) sont des composés aromatiques possédant une ossature biphényle, substituée par des atomes de chlore en position et nombre variables [1]

---

Author Correspondence, e-mail: [Belgacem3963\\_alg@yahoo.fr](mailto:Belgacem3963_alg@yahoo.fr)

[ICID: 1020798](#)

Le polychlorure de venyle (PVC), les hydrocarbures aromatiques polycycliques (PAHs) sont souvent résistants à la dégradation biologique et ne sont pas efficacement éliminés par des techniques classiques de séparation [2]. les pesticides comme le(DDT) (dichlorodiphényltrichloroethane) , les éthers de glycol et les organophosphorés [3].

Pour l'évaluation du risque environnemental, la législation européenne sur les produits chimiques (substances existantes, substances nouvelles, biocides) impose une évaluation du risque chimique afin d'assurer la protection de la santé humaine et celle de l'environnement.

La caractérisation du risque est basée sur la valeur du rapport PEC/PNEC, Tel que:

**PEC** : Concentration Environnementale Prévisible.

**PNEC** : Concentration Prévisible sans Effet [4].

D'après tout cela, on comprend qu'il y a une nécessité urgente et sérieuse qui demande et exige le traitement avec une grande prudence et exactitude de ces composés chimiques. Cela ne peut se faire qu'après avoir élaboré une manière pour l'étude et le calcul avec précision des propriétés physico-chimiques qui ont une influence sur l'environnement.

Le but de ce travail, est centré sur le développement d'un modèle mathématique basé sur les données expérimentales disponibles et en fonction de la structure moléculaire, permettant ainsi d'estimer, d'étudier et d'identifier la propriété physico-chimiques notamment le coefficient de partage octanol/eau (log<sub>kw</sub>), qui est un paramètre clé pour l'évaluation de l'impact des composés chimiques sur l'environnement, sans recours aux activités expérimentales.

Pour la réalisation de notre travail, nous avons construits une base de données qui a été arrange a partir des variétés de familles chimiques et biochimiques, suivi du choix des blocs des descripteurs moléculaires qui sont (Descripteurs constitutionnels, Descripteurs topologiques, Indices de connectivité, Indices de charge topologiques , Descripteurs Propriétés moléculaires , Descripteurs du groupe fonctionnel , Descripteurs des fragments d'atome centré). Ces blocs et cette base de donnée subits a un calcul statistique de la régression linéaire multiple par la méthode « stepwise » a l' aide du logiciel XLSTAT. Et finalement les résultats du travail.

## 2. MATERIEL ET METHODES

### 2.1. Base de données

Cette base de données contient environ 1311 constituants chimiques et biochimiques. Parmi ces composées il existe des substances qui n'ont pas de valeurs expérimentales (vide), et d'autres qui ont des valeurs sans recommandation, et un nombre de composées que nous avons calculées leurs valeurs proches de la moyenne (M). Et reste 4 composées sans valeurs proches de la moyenne, et finalement nous avons élaboré une nouvelle base de données contient 936 composées avec des valeurs recommandées. Cette base de données qui sera utilisées dans notre travail.

Le nombre total des composées = 1311.

Le nombres des composées sans valeurs (vides) = 271.

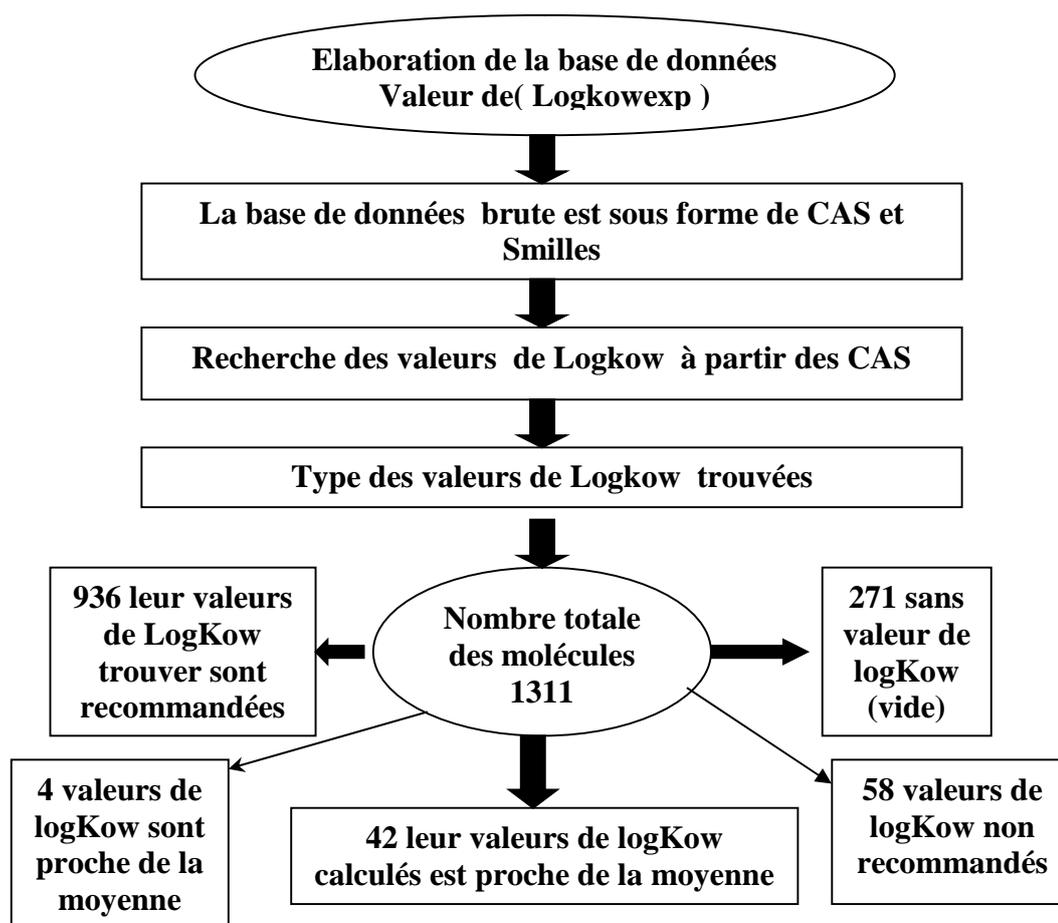
Le nombre des composées qui ont des valeurs sans vérification = 58.

Le nombre des composées qui ont des valeurs proches de la moyenne(M) = 42

Le nombre des composées sans valeurs proches de la moyenne = 4 qui sont (51-67-2), (729-99-7), (1689-84-5), (144-83-2).

Le nombre total des composées de valeurs recommandées trouvées (R) = 936 .

Ce dernier (R) représente les valeurs fiables qui sont recommandées et convenables d'être utilisées comme base de données pour notre étude.



**Fig.1.** Montre l'élaboration de la base données valeur de (logKow exp)

## 2.2. Génération des descripteurs moléculaires

Les blocs des descripteurs moléculaires choisis et les résultats de calcul des descripteurs (paramètres) par le Logiciel DRAGON sont résumés dans le tableau suivant III-1

**Tableau 1.** Montre les résultats de calcul des descripteurs

Les blocs des descripteurs	Le nombre de descripteurs dans le bloc .	Le nombre de descripteurs après la réduction.
Descripteurs du groupe fonctionnel	154	90
Descripteurs des fragments de d'atome centré	120	83

### 2.3. La régression multiple

Cette analyse est divisée à l'analyse de régression simple (training set 70%) et la validation du modèle par 280 molécules qui présentent 30% de l'ensemble total. Nous avons fait les calculs avec XLSTAT software et les résultats pour ces descripteurs sont rapportés avec référence aux valeurs expérimentales. Les modèles sont calculés par la méthode pas à pas de la sélection des variables pour la régression linéaire multiple (stepwise). Avec la probabilité d'entrée est  $= 0.05$ .

### 2.4. Méthodologie

Nous a permis de faire le calcul et l'analyse statistique par la technique QSPR qui est couramment utilisé dans les études linéaire et aussi comme un excellent outil pour la sélection des descripteur qui précède la construction d'un modèle linéaire ou non linéaire [5]. et la méthode régression multiple « stepwise » avec les sept blocs des descripteurs moléculaires. Cette analyse se divise en analyse statistique simple et analyse des résidus et validation du modèle par 280 molécules qui présentent 30% du nombre globale. Pour l'évaluation de la corrélation entre les descripteurs 2D du bloc et les valeurs expérimentales de log<sub>kow</sub> qui ont été enregistrés durant la préparation de la base de données. Nous avons réalisé les calculs à l'aide du logiciel de modélisation moléculaires XLSTAT et rapporté les résultats obtenus pour ces descripteurs en nous référant à la prévision de la propriété physico-chimique log<sub>kow</sub>. dont les valeurs expérimentales ont été prélevées dans la littérature.

Les modèles ont été calculés à l'aide de la méthode de sélection pas à pas des variables de la régression linéaire multiple «STEPWISE». Avec la probabilité pour l'entrée est  $= 0.05$

## 3. RESULTATS ET DISCUSSION

**Tableau 2.** Coefficient de corrélation  $R^2$  Nombre de variables et de molécules non corrélées par bloc de descripteurs.

Blocs des descripteurs	Groupes fonctionnels	Fragments de l'atome central
$R^2$	0.926	0.949
Nombre de variables	49	43
Nombre de molécules non corrélées (hors 95%)	65	32

**Tableau 3.** Montre le nombre de constituants, Type de descripteurs

Nombre de constituants	Type de descripteurs	R <sup>2</sup>	AAD	S(std)	Ref
936(Plusieurs familles chimiques)	Groupe fonctionnel	0.926	0.139	0.548	This work
	Fragments de l'atome central	0.949	0.248	0.477	

#### 4. CONCLUSION

Dans ce travail nous avons abordé le problème de prédiction du coefficient de partage octanol/eau (log<sub>ow</sub>), pour cela nous avons fait la corrélation de différent types de paramètres (descripteurs) et la contribution pour le calcul de (log<sub>ow</sub>) par la méthode QSPR. Après le calcul des descripteurs 2D (constitutionnel, indices de connectivite, propriétés moléculaires, groupes fonctionnelles, fragments de l'atome centré, topologique, indices de charge topologique) par le logiciel Dragon. Ce travail a été effectué sur 295 paramètres de 7 différents blocs et une base de données contienne 936 composés a été arrangé à partir de 1311 composés de différentes familles chimiques et biochimiques. Pendant la prédiction nous avons trouvés une relation mathématique.

$$\log_{ow} = f(\text{paramètres structuraux}), \log_{ow} = f(b + (a_i + x_i)).$$

Ce modèle donne des résultats de corrélation remarquables avec les descripteurs des blocs fragments de l'atome centré et groupes fonctionnels, (R<sup>2</sup> = 0,949, = 0.477), (R<sup>2</sup> = 0,926, = 0.548) respectivement.

Après la comparaison des résultants de notre travail avec des autres on peut conclure que le pouvoir prédictif des méthodes dépend du type et du nombre de composés dans la base de données. Pour l'amélioration des résultats on doit: exclure les molécules non corrélées (rejetées) et on fait la combinaison entre les deux (1ère ordre groupes fonctionnels + 2<sup>ème</sup> ordre fragments de l'atome centré) tant qu'ils sont les meilleurs blocs.

#### 5. REFERENCES

- [1] Barret M. (2009). Devenir des perturbateurs endocriniens HAPs/NP/PCBs au cours de la digestion anaérobie de boues contaminées: rôle de la biodisponibilité et du cométabolisme, Thèse Doctorat, Université Montpellier II.
- [2] Crisafullyetal. 2008. page 821

Prediction of environmental parameters of polycyclic aromatic hydrocarbons with COSMO-RS Bernd Schroder , Luis M.N.B.F. Santos , Marisa A.A. Rocha , Mariana B. Oliveira , Isabel M. Marrucho , Joao A.P. Coutinho.

journal homepage: [www.elsevier.com/locate/chemosphere](http://www.elsevier.com/locate/chemosphere)

[3] Dr : Khalil Hanna. Etude de la faisabilité de l'utilisation de molécule "cage" dans la dépollution des sols : Solubilisation et Extraction de Polluants Organiques par les Cyclodextrines . Thèse page 51

[4] Marchand M. and Tissier C. Cellule Ifremer/Ineris d'Analyses des Risques Chimiques en Milieu Marin (ARC), Pages 7.52.79. Centre Ifremer de Nantes, BP 21105, 44311 Nantes Cedex 3, France.

[5] Xia B., Liu K., Gong Z., Zheng B., Zhang X. et Fan B. *Ecotoxicol. Environ. Saf.* 2009, 72, 787-794.

## CALCUL DE COEFFICIENT DE PARTAGE OCTANOL-EAU DES COMPOSES ORGANIQUES EN UTILISANT DES DESCRIPTEURS MOLECULAIRES

### RESUME

L'étude de la relation quantitative structure-propriété (QSPR) est réalisée pour établir des corrélations qui relient les structures moléculaires des composés organiques à leur coefficients de partage Octanol-eaux,  $K_{ow}$ , en utilisant les descripteurs moléculaires.

Les corrélations sont simples en application avec une précision, qui fournissent une manière simple, directe et relativement précise pour calculer  $K_{ow}$ . Ces calcul nous donnent un modèle qui nous amène aux résultats remarquables en corrélation avec les descripteurs des blocs fragments de l'atome centré et groupes fonctionnels respectivement. ( $R^2 = 0.949$ ,  $n = 0477$ ) ( $R^2 = 0.926$ ,  $n = 0,548$ ).

**Mots clés:** QSPR; modèle; Coefficient d'activité; Dilution infinie; Solution aqueuse; Descripteurs moléculaires.

### How to cite this article

Souyei B. Calculation of coefficient of sharing octanol-water of organic compounds using molecular descriptors. J Fundam Appl Sci. 2010, 2(2), 217-224.