

## La probabilité d'hydratation moléculaire : un nouveau concept pour le calcul du $\log P$ d'une molécule à partir de sa structure 3D

N.T.D. Nguyen<sup>1</sup>, E. Kummer<sup>1</sup>, J.P. Dubost<sup>1</sup>, T. Convard<sup>2</sup>, J. Barbanton<sup>2</sup> et A. Carpy<sup>3</sup>

<sup>1</sup>GERSAAC, Laboratoire de Chimie Analytique, UFR de Pharmacie, Université de Bordeaux II, 3 place de la Victoire, 33076 Bordeaux Cedex, France

<sup>2</sup>Centre de Recherche Lacassagne, Laboratoire LIPHA, Lyon, France

<sup>3</sup>LPTC, UPRES-A 5472 CNRS, UFR de Chimie, Université de Bordeaux I, 351 cours de la Libération, 33405 Talence Cedex, France

A new concept, the Molecular Hydration Probability (MHP), depending on the 3D structure of a molecule and on the composition of the first hydration shell, is presented. The knowledge of MHP allows the estimation of virtual  $\log P$  on the van der Waals surface of molecules.

Pour atteindre son lieu d'action, une molécule médicamenteuse va devoir franchir un certain nombre de barrières constituées de membranes biologiques de nature lipidique séparant des compartiments aqueux. Cette étape dépend de la lipophilie de la molécule. C'est elle qui, en imposant la concentration en drogue libre au niveau du récepteur, va moduler en partie l'activité du médicament. La lipophilie, qui intervient aussi lors de l'interaction drogue-récepteur, est donc une propriété moléculaire importante. Elle est classiquement quantifiée par le logarithme du coefficient de partage,  $\log P$ , dans le système eau-octanol. Comme la détermination expérimentale de  $\log P$  est souvent délicate, de nombreuses méthodes ont été développées pour

prédire ce paramètre. Parmi celles-ci, les plus couramment utilisées\* sont celles qui sont basées sur l'additivité des contributions lipophiles de fragments moléculaires ou atomiques [1-10]. Elles ne tiennent pas compte ni de l'aspect tridimensionnel des molécules ni de leur stéréochimie. Pour pallier ce manque, d'autres techniques dites « tridimensionnelles » ont été proposées [11-18]. Cet article décrit le calcul du  $\log P$  d'une molécule à partir d'un nouveau concept, dénommé Probabilité d'Hydratation Moléculaire (PHM), qui fait lui-même appel à la notion de Potentiel de Lipophilie Moléculaire (PLM) [19-21].

### Définition de la probabilité d'hydratation moléculaire (PHM)

Le Potentiel de Lipophilie Moléculaire (PLM) est tel qu'en tout point M de l'espace entourant une molécule, composée de N fragments atomiques :

\* Ayant donné lieu à des logiciels commerciaux ou incluses dans des logiciels de modélisation moléculaire.

$$\text{PLM} (M) = \sum f_i g(d_i) \quad (1)$$

où  $f_i$  désigne la contribution lipophile d'un fragment  $i$   $g(d_i)$  est une fonction décroissante de la distance  $d_i$  entre le point M et un fragment  $i$ , exprimée en Å.

D'autre part, les auteurs ont montré qu'il est possible de calculer à partir du PLM, la probabilité  $\text{pH}_2\text{O}$  pour qu'existe une molécule d'eau au point M :

$$\text{pH}_2\text{O} = \frac{1}{1 + 10^{\text{PLM}}} \quad (2)$$

Considérons une molécule contenant N fragments atomiques. Lorsque la structure 3D de cette molécule est connue, il est possible de calculer le nombre total (NT) de molécules d'eau que l'on peut placer côte à côte sur la surface de Van der Waals de cette molécule. Ce nombre ne dépend que des caractéristiques géométriques du conformère envisagé ; en particulier, il est indépendant des contributions lipophiles des différents fragments atomiques qui constituent la molécule. Par la suite, il est possible de calculer pour chacune de ces NT molécules, sa probabilité de présence  $\text{pH}_2\text{O}$ . Cette probabilité est un nombre compris entre 0 et 1 que l'on peut assimiler à une fraction de molécule d'eau. La somme de toutes ces fractions est alors égale au nombre de molécules d'eau (NH) présentes à la surface de Van der Waals de la molécule.

La probabilité d'hydratation moléculaire, PHM est définie comme étant le rapport  $\text{NH}/\text{NT}$  :

$$\text{PHM} = \frac{\text{NH}}{\text{NT}} \quad (3)$$

Ce nombre dépend à la fois des caractéristiques géométriques du conformère envisagé et des contributions lipophiles des fragments atomiques dans une base donnée, pour une fonction  $g(d_i)$  donnée.

### Relation entre la probabilité d'hydratation (PHM) et le log P d'une molécule

Pour un soluté quelconque, de lipophilie moyenne, les groupements hydrophiles vont avoir tendance à attirer les molécules d'eau et les groupements hydrophobes à les repousser. Plus une molécule est globalement hydrophile, plus son log P est faible et plus le nombre de molécules d'eau situées sur sa surface de Van der Waals (NH) est grand ou encore, plus la différence (NT-NH) est petite. Dès lors, il semble logique de penser qu'il existe une relation linéaire entre le log P du soluté et le logarithme du rapport (NT-NH)/NH :

$$\log P = a \log \frac{\text{NT} - \text{NH}}{\text{NH}} + b = a \log \frac{1 - \frac{\text{NH}}{\text{NT}}}{\frac{\text{NH}}{\text{NT}}} + b \quad (4)$$

$$\log P = a \log \frac{1 - \text{PHM}}{\text{PHM}} + b \quad (5)$$

Si donc on détermine au préalable les valeurs de  $a$  et  $b$ , il devient possible de calculer le log P d'une molécule à partir de sa probabilité d'hydratation.

### Détermination des coefficients $a$ et $b$

#### Principe

Si l'on dispose d'une série de N molécules dont les log P ont été mesurés expérimentalement, la détermination des coefficients  $a$  et  $b$  peut théoriquement s'obtenir par une méthode de moindres carrés, après avoir calculé pour chaque molécule sa probabilité d'hydratation. Cependant comme la probabilité d'hydratation d'une molécule dépend :

- des caractéristiques géométriques du conformère envisagé,
- de la formule du potentiel de lipophilie moléculaire (PLM) que l'on choisit d'utiliser, c'est-à-dire de l'expression mathématique de la fonction  $g(d_i)$  pour une base de  $f_i$  donnée,

il est nécessaire de choisir :

- parmi tous les conformères possibles d'une même molécule, celui sur lequel on va effectuer le calcul ; ce peut être par exemple le conformère de plus basse énergie obtenu après analyse conformationnelle ou encore celui qui satisfait au pharmacophore. Dans un premier temps, pour s'affranchir du problème du choix du conformère, une série de 59 molécules organiques simples<sup>1</sup> dont les log P expérimentaux sont connus et qui présentent des structures rigides ont été utilisées. Ainsi chaque molécule ne possède qu'un seul conformère.
- parmi les différentes expressions mathématiques de  $g(d_i)$  entrant dans la formule du PLM, celle qui permet d'obtenir la meilleure corrélation.

4 expressions différentes  $g(d_i)$  ont été envisagées (Tab. I).

Les valeurs de  $f_i$  utilisées ont été déterminées par l'un de nous [24] en décomposant 5000 molécules dont les log P expérimentaux étaient connus, en 45 fragments atomiques prenant en compte la nature de chaque atome et son environnement immédiat c'est-à-dire le nombre et la nature des liaisons qu'il forme avec ses voisins.

### Résultats

Les valeurs obtenues pour  $a$  et  $b$  ainsi que les paramètres statistiques caractéristiques montrent que la meilleure

1. La liste des molécules et leurs log P seront envoyés sur simple demande.

**Tableau I. Différentes fonctions de distances entrant dans la formule du PLM.**

Formule proposée à l'origine par Audry [19]	Formule proposée par Fauchère et al. [22]
$g(di) = \frac{1}{1+di}$	$g(di) = e^{-di}$
Formule n° 1 proposée par Gaillard et al. [23]	Formule n° 2 proposée par Gaillard et al. [23]
$g(di) = e^{-di/2}$	$g(di) = \frac{1+e^{-ab}}{1+e^{a(di-b)}}$ avec $a = 1,5$ $b = 3,15$

**Tableau II. Paramètres a et b et grandeurs statistiques de l'équation de corrélation : log P exp. = a log [(1-PHM)/PHM] + b.**

Formule	Audry	Fauchère et al.	Gaillard (1)	Gaillard (2)
a	6,053	56,780	9,996	6,548
b	-0,077	-0,388	-0,329	-0,545
R <sup>2</sup>	<b>0,974</b>	0,97	0,973	0,954
F	<b>1090</b>	909	1015	588
SCE	<b>5,09</b>	6,05	5,45	9,04

corrélation est obtenue avec la formule hyperbolique d'Audry [19] (Tab. II).

**Fiabilité de la prévision de log P à partir de la probabilité d'hydratation moléculaire**

Pour tester la fiabilité de la prévision de log P à partir de PHM, 49 nouvelles molécules organiques simples dont les

**Tableau III. Valeurs des paramètres α et β et des grandeurs statistiques caractéristiques de l'équation de corrélation : log P<sub>mes</sub> = α log P<sub>calc</sub> + β.**

Formule	Expression g(di)	α	β	R <sup>2</sup>	SCE	F	SCE/Σ(log P <sub>mes</sub> )
Audry	$g(di) = \frac{1}{1+di}$	<b>1,07</b>	<b>-0,21</b>	<b>0,98</b>	<b>6,22</b>	<b>1109</b>	<b>0,015</b>
Fauchère et al.	$g(di) = e^{-di}$	1,08	-0,05	0,96	13,83	485	0,034
Gaillard et al. (Equation n° 1)	$g(di) = e^{-di/2}$	1,08	-0,11	0,97	10,50	632	0,026
Gaillard et al. (Equation n° 2)	$g(di) = \frac{1+e^{-ab}}{1+e^{a(di-b)}}$ avec $a = 1.5$ $b = 3.15$	1,07	0,07	0,93	21,65	298	0,053

log P expérimentaux étaient connus et qui présentaient une structure rigide, ont été choisies dans la littérature [6,9]. Les log P de ces molécules, calculés à partir de leur probabilité d'hydratation moléculaire (log P calc.), ont été comparés à leurs log P expérimentaux (log P<sub>mes</sub>). Pour cela, l'équation de corrélation :

$$\log P_{mes} = \alpha \log P_{calc} + \beta .$$

les valeurs α et β et les paramètres statistiques caractéristiques de cette corrélation : R<sup>2</sup>, F et SCE ont été déterminés (la prévision du log P est d'autant plus fiable que les valeurs de R<sup>2</sup> et F sont grandes, que la valeur de SCE est faible et que le rapport SCE/Σ(log P<sub>mes</sub>) est faible).

L'examen du tableau III montre que parmi les 4 expressions différentes de g(di) et pour le jeu de molécules utilisé, c'est encore l'expression proposée par Audry [19] qui est la plus fiable. L'équation (2) de Gaillard et al. [23] conduit aux résultats les moins fiables.

**Conclusion**

Le concept de probabilité d'hydratation moléculaire est un concept simple qui permet d'estimer le log P d'une molécule en prenant en compte sa structure tridimensionnelle. Avec une base de 45 fragments atomiques et un jeu de 49 molécules organiques simples, rigides, la fiabilité de la prévision du log P par l'intermédiaire de la Probabilité d'Hydratation Moléculaire (PHM) (faisant intervenir une expression de distance hyperbolique dans le calcul du Potentiel de Lipophilie Moléculaire (PLM)) est légèrement supérieure à celle faite par simple sommation des contributions fragmentales. La valeur du rapport SCE/Σ(log P<sub>mes</sub>) obtenue par la méthode des sommations avec la même base de fi est égale à 0,020. Si le gain sur la précision des résultats peut sembler minime, d'un point de vue conceptuel cette nouvelle méthode d'estimation de log P, conformation-dépendante, est plus satisfaisante pour l'esprit que les

méthodes purement fragmentales. Comme les molécules flexibles peuvent adopter différentes conformations dans des solvants de polarités différentes, l'influence du caractère plus ou moins lipophile des molécules sur les conformations ne peut être ignorée. Les premiers résultats encourageants présentés dans cet article devront cependant être confirmés sur un ensemble plus important et plus diversifié de molécules. S'ils sont confirmés on peut penser que la méthode pourra être utilisée ultérieurement, en conjonction avec l'étude d'autres propriétés moléculaires, pour identifier les conformations actives des ligands ou tout au moins les conformations dans différents solvants.

### Références

1. Fujita, T.; Iwasa, J.; Hansch, C. *J. Am. Chem. Soc.* **1964**, *86*, 5175-5180.
2. Leo, A.; Jow, P. Y. C.; Silipo, C.; Hansch, C. *J. Med. Chem.* **1975**, *18*, 865-868.
3. Rekker, R. F. In: *The Hydrophobic Fragmental Constant*, Elsevier Scientific Publishing Co 1977.
4. Broto, P.; Moreau, G.; Vanduycke, C. *J. Med. Chem. -Chim. Ther.* **1984**, *19*, 71-78.
5. Ghose, A. K.; Crippen, G. M. *J. Comp. Chem.* **1986**, *7*, 565-577.
6. Suzuki, T.; Kudo, J. *J. Comput. Aided. Mol. Des.* **1990**, *4*, 155-198.
7. Klopman, G.; Wang, S. *J. Comp. Chem.* **1991**, *12*, 1025-1032.
8. Niemi, G. J.; Basak, S. C.; Veith, G. D.; Grunwald, G. *Environ. Toxicol. Chem.* **1992**, *11*, 893-900.
9. Rekker, R. F.; Mannhold, R. In: *Calculation of Drug Lipophilicity. The Hydrophobic Fragmental Approach*. VCH, Weinheim, 1992.
10. Meylan, W. M.; Howard, P. *J. Pharm. Sci.* **1995**, *84*, 83-92.
11. Hopfinger, A. J.; Battershell, R. D. *J. Med. Chem.* **1976**, *19*, 569-573.
12. Iwwase, K.; Komatsu, K.; Hirono, S.; Nakagawa, S.; Moriguchi, I. *Chem. Pharm. Bull.* **1985**, *33*, 2114-2121.
13. Bruijn, J.; Hermens, J. *Quant. Struct. -Act. Relat.* **1990**, *9*, 11-21.
14. Kantola, A.; Villar, H. O.; Loew, G. H. *J. Comp. Chem.* **1991**, *12*, 681-689.
15. Kellogg, G. E.; Semus, S. F.; Abraham, D. J. *J. Comput. Aided Drug. Des.* **1991**, *5*, 545-552.
16. Richard, N. G. J.; Williams, P. B. *Chem. Des. Automat. News.* **1994**, *9*, 1-26.
17. Waller, C. L. *Quant. Struct. -Act. Relat.* **1994**, *13*, 172-176.
18. Bodor, N.; Buchwald, P. *J. Phys. Chem. B.* **1997**, *101*, 3404-3412.
19. Audry, E. *Le Potentiel de Lipophilie Moléculaire, nouvelle méthode d'approche des relations structure-activité - Application à une série de molécules  $\beta$ -adrénnergiques*. Thèse de Doctorat d'État des Sciences Pharmaceutiques, Université de Bordeaux 2, 1985.
20. Audry, E.; Dubost, J. P.; Colleter, J. C.; Dallet, Ph. *Eur. J. Med. Chem.* **1986**, *21*, 71-72.
21. Audry, E.; Dubost, J. P.; Dallet, Ph.; Langlois, M. H.; Colleter, J. C. *Eur. J. Med. Chem.* **1989**, *24*, 155-161.
22. Fauchere, J. L.; Quarendon, P.; Kaetterer, L. *J. Mol. Graph.* **1988**, *6*, 202-206.
23. Gaillard, P.; Carrupt, P.; Testa, B.; Boudon, A. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 83-96.
24. Nguyen, N. T. D. *Détermination de nouvelles constantes atomiques de lipophilie utilisables pour la prévision du log P d'une molécule*. Thèse de l'Université de Bordeaux 2, Mention Sciences Biologiques et Médicales, Option Sciences Pharmaceutiques. Université Bordeaux 2, 1996.

\* Author to whom all correspondence should be addressed. E-mail: giulo@indigo.farma.unimi.it