

Importance de la lipophilie en modélisation moléculaire

A. Carpy

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

In the same way as the shape of a molecule, its lipophilicity is an important molecular property. Lipophilicity is globally measured by the logarithm of the partition coefficient in the *n*-octanol/water system ($\log P$). This parameter is used in numerous Quantitative Structure-Activity Relationships (QSARs) studies which are part of Molecular Modelling. In addition, among estimation methods of $\log P$, more and more integrate a three dimensional (3D) parameter and require molecular models. Lipophilicity is as important for researchers involved in human health as for those involved in environmental sciences.

Propriétés de partage

Le partage d'une molécule entre une phase aqueuse et une phase lipidique conditionne en partie ses propriétés biologiques telles que le transport, le passage à travers les membranes, la biodisponibilité (distribution et accumulation), l'affinité pour un récepteur et la fixation par une protéine, l'activité pharmacologique ou encore la toxicité. S'agissant de contaminants, ce même partage conditionne leur devenir dans notre environnement en particulier leur accumulation dans les organismes aquatiques.

Depuis les travaux de Collander à la fin des années 1950 [1], puis ceux du groupe de Hansch quelques années plus tard [2,3], le coefficient de partage P^1 d'une molécule dans un système biphasique constitué de deux solvants non-miscibles (le plus souvent le système *n*-octanol/eau), est reconnu pour sa faculté à mimer le passage de cette molécule à tra-

1. Selon les auteurs, on parle de coefficient ou de constante de partage ou de distribution ou encore de rapport de distribution. Outre P , plusieurs autres symboles sont utilisés tels K_D , K_p , K_{ow} ou encore D . On parle également de P' ou de D' , le coefficient de partage apparent, apparent signifiant qu'il n'est valable que sous certaines conditions expérimentales utilisées lors de sa détermination. En effet il arrive que le soluté soit présent dans chaque phase sous plusieurs formes dont une seule intervient dans le partage. C'est le cas par exemple pour les acides faibles AH. Dans la phase organique l'acide faible existe sous sa forme non-dissociée AH alors que dans l'eau, il se trouve à la fois sous forme AH et sous forme ionisée A^- . Le coefficient de partage apparent dépend alors du pH. La connaissance du pK_a permet le calcul du coefficient de partage vrai.

vers les membranes biologiques. Pour des solutions diluées, ce coefficient de partage *n*-octanol/eau est le rapport de la concentration d'une molécule de soluté dans le *n*-octanol sur sa concentration dans l'eau lorsque le système biphasique est en équilibre. En fait, comme les valeurs de P mesurées s'étendent sur au moins douze unités de grandeur ($10^{-4} - 10^8$), on utilise de préférence les logarithmes décimaux. Le partage est donc une propriété physico-chimique importante qui peut être utilisée pour représenter la nature lipophile ou hydrophile d'une molécule².

Quant au $\log P$, il constitue un paramètre unique qui regroupe plusieurs effets : tous les types d'interactions non-covalentes, la solvation ainsi qu'une composante d'entropie. Il est très largement utilisé dans des études de relations structure-activité quantitatives (QSARs) dans les sciences pharmaceutiques, biochimiques, toxicologiques et dans les sciences de l'environnement. La lipophilie intéresse donc tout autant la communauté qui étudie les problèmes de santé humaine que celle qui est impliquée dans les problèmes de l'environnement.

Détermination expérimentale des $\log P$

Chaque fois que cela sera possible, le $\log P$ d'une molécule fera l'objet d'une détermination expérimentale. La méthode dite des flacons agités ou « shake-flask », est la plus ancienne. La molécule étudiée, interagit avec les deux phases en équilibre d'une manière qui mime la façon dont par exemple, un ligand se fixe sur le site actif d'un récepteur. Malgré ses imperfections (problèmes d'adsorption par le verre, formation d'émulsions lors de l'agitation, domaine de mesure étroit de l'ordre de -3 à $+4$, durée...), elle reste cependant la méthode de choix pour des molécules organiques originales et de ce fait, elle est préconisée comme procédure standard de caractérisation par l'OCDE [4]. Toutefois, elle tend à être supplantée par les méthodes chromatographiques. En particulier, la chromatographie liquide haute performance à polarité de phase inversée, adaptée aux études de criblage, est elle aussi préconisée par l'OCDE [5]. Dans ce cas on utilise comme indice de lipophilie, une valeur déduite de la mesure des temps de rétention, $\log k_w$. Une revue des différentes méthodes de détermination des coefficients de partage et de leurs différents domaines d'application, a été publiée récemment [6].

2. Dans cet article on parlera de lipophilie ou d'hydrophilie (Grand Larousse Universel). Certains auteurs utilisent les termes lipophilité et hydrophilité.

Méthodes d'estimation des log *P*

La plupart des méthodes expérimentales de détermination de log *P* souffrent du même inconvénient, à savoir que leur domaine d'application est relativement étroit. D'autre part, du fait de la nature intrinsèque de certaines molécules, leurs log *P* sont inaccessibles à l'expérience. C'est le cas en particulier des surfactants qui ont tendance à s'accumuler à l'interface du système biphasique au lieu de se disperser dans les deux phases. Enfin, dans le domaine de la conception assistée par ordinateur ou dans le domaine de la chimie combinatoire, les chercheurs travaillent sur des modèles moléculaires avant même que les molécules aient été synthétisées. Ceci explique le succès des nombreuses méthodes d'estimation de log *P* qui ont été décrites dans la littérature depuis plus de trente ans. Les plus anciennes sont des méthodes fragmentales dans lesquelles une molécule est divisée en fragments prédéfinis³ et les contributions correspondantes sont sommées pour conduire à une valeur estimée du log *P* [7-8 & Réfs. citées].

Il peut être intéressant de remarquer qu'une propriété moléculaire aussi complexe que la lipophilie est, encore aujourd'hui, calculée principalement sur la base de valeurs de fragments ou d'atomes déterminées de façon empirique et en faisant abstraction des processus impliqués dans le phénomène de solvation et des propriétés moléculaires qui les affectent. D'ailleurs, on rencontre un certain nombre de difficultés lorsqu'on veut établir des relations entre le comportement plus ou moins lipophile des molécules et certaines de leurs propriétés fondamentales telles que leur miscibilité avec l'eau ou avec d'autres substances, leur chaleur de vaporisation ou encore leur réactivité chimique.

Dans la première partie de leur article, **M.-H. Van Eyck et al.** brossent l'évolution du concept de solvation et son approche expérimentale par l'intermédiaire de la lipophilie. Classiquement, l'énergie de solvation d'une molécule est associée à son énergie de transfert d'un milieu hydrophobe vers un milieu hydrophile, mesurée par le coefficient de partage. Les auteurs abordent les précautions qu'il convient de prendre lorsqu'on veut modéliser une grandeur microscopique à partir d'une grandeur macroscopique.

D'autre part, comme certaines méthodes d'estimation de log *P* n'ont pas de bases scientifiques rigoureuses, elles nécessitent des bases de fragments de plus en plus importantes ainsi que l'introduction de facteurs de correction. Les méthodes prévisionnelles de log *P* continuent toutefois à mobiliser les chercheurs puisque cinq nouvelles viennent encore de voir le jour [9-13]. Dans ce dossier, **J. Devillers** présente AUTOLOGPTM, un nouveau programme puissant développé à partir d'un jeu de 7200 molécules et qui utilise la méthode d'autocorrélation et les réseaux neuro-mimétiques pour estimer les log *P*. Parmi les méthodes récentes, un certain nombre d'entre elles introduisent un paramètre tridimensionnel [13 & Réfs. citées].

3. Fragment est pris ici au sens large ; il englobe également les contributions atomiques où chaque atome, carbone ou hétéroatome est défini par son environnement, plus ou moins lointain selon les méthodes.

Paramètres tridimensionnels impliqués dans le calcul des log *P*

Le paramètre tridimensionnel le plus simple qui caractérise une molécule est lié à sa taille. Celle-ci joue un rôle important dans la détermination de la solubilité et dans les propriétés de partage. Même si dans une description de chimie quantique le nuage électronique relatif à la molécule n'a pas de surface définie, le concept de rayon de van der Waals (rayon défini en terme de distance maximale d'approche d'atomes ou de molécules non liés) est encore largement utilisé et il fournit une bonne base de départ pour un certain nombre de calculs. Chaque atome d'une molécule peut être représenté par une sphère centrée sur la position d'équilibre du noyau et ayant un rayon égal au rayon de van der Waals de l'atome correspondant. La surface extérieure de toutes ces sphères définit la surface de van der Waals. La surface d'accessibilité au solvant est elle-même définie comme étant le lieu du centre d'une sphère censée représenter le solvant, roulant sur la surface de van der Waals [14]⁴. Il existe un grand nombre de méthodes numériques [15,16 & Réfs. citées], analytiques [17,18 & Réfs. citées] ou mixtes [13] pour calculer ces différents types de surfaces. Il est alors possible de décrire le log *P* d'une molécule à l'aide d'équations faisant intervenir les surfaces ou les volumes délimités correspondants.

« Potentiel de lipophilie moléculaire »

Définitions

Une autre façon de décrire la lipophilie d'une molécule dans l'espace, consiste à calculer le « potentiel de lipophilie moléculaire », terme introduit pour la première fois dans le langage scientifique par Audry et al. [19]. La lipophilie d'une molécule dépend de la nature des groupements exposés au solvant et de ce fait, dépend de la conformation de cette molécule. Le « potentiel de lipophilie moléculaire » traduit l'influence des contributions fragmentales de lipophilie d'une molécule sur son environnement. Cette grandeur dépend de la distance à laquelle elle est calculée :

$$MLP = \sum_i f_i \cdot g(d_i) \quad (5)$$

où f_i est la constante lipophile du fragment i telle qu'elle a été définie dans le calcul de log *P* et $g(d_i)$ est une fonction de la distance d_i entre le fragment i et un point de l'espace environnant la molécule.

Ce « potentiel » peut être considéré comme une extension naturelle du « moment dipolaire hydrophobe » μ introduit par Eisenberg & McLachlan par analogie avec le moment dipolaire électrostatique [20] :

4. Certains auteurs calculent la surface de contact entre la sphère de solvant et la surface de van der Waals de la molécule lorsque la première se déplace en roulant sur la seconde.

5. Le sigle anglo-saxon MLP pour « Molecular Lipophilicity Potential » est ici utilisé.

$$\mu = \sum_i f_i \cdot r_i$$

où r_i est le vecteur joignant l'origine à un atome du fragment i .

Ces mêmes auteurs ont montré que l'énergie libre de transfert d'une molécule ou énergie libre de solvatation, était une fonction linéaire de la surface d'accessibilité au solvant.

Bien que dépourvu d'une base physique, (le « potentiel » de lipophilie n'a pas la dimension d'une énergie), le terme et sa traduction anglaise ont été adoptés par une majorité de chercheurs travaillant dans le domaine [21-23 & Réfs. citées]. Pour tourner la difficulté, d'autres auteurs, tout en conservant le sigle MLP préfèrent parler de « Molecular Lipophilicity Pattern » [24].

Plusieurs méthodes de calcul qui diffèrent soit par la base de constantes fragmentales [25-32], soit par l'expression de la fonction $g(d_i)$ [21-24,33-34], sont actuellement utilisées. Dans leur article, **N.T.D. Nguyen et al.** testent d'ailleurs quelques expressions de $g(d_i)$.

Applications en modélisation moléculaire

Depuis l'introduction du concept de « potentiel de lipophilie moléculaire », de nombreux articles ont démontré son utilité pour décrire de façon qualitative ou quantitative, la distribution dans l'espace de la lipophilie, soit sur une grille, soit sur une surface moléculaire.

J.A.W. Brickmann et R. Jäger illustreront un concept différent, celui de « densité de surface d'énergie libre ». Dans cette approche, la lipophilie globale de la molécule est toujours considérée comme étant la somme de contributions fragmentales mais elle est quantifiée à l'aide de l'énergie libre de solvatation.

L'introduction d'un champ de force lipophile dans le programme CoMFA (Comparative Molecular Field Analysis) [35] s'est révélée être d'une grande utilité pour la conception rationnelle de nouvelles molécules actives [33]. **G.E. Kellogg & D.J. Abraham** aborderont ce problème dans leur article.

N.T.D. Nguyen et al. montreront à l'aide d'un nouveau concept, la Probabilité d'Hydratation Moléculaire (PHM), qu'il est possible de régénérer le log P à partir du « potentiel de lipophilie moléculaire ».

L'utilisation dans des études de relations structure-activité quantitatives (QSARs), de paramètres dérivés du calcul et de la représentation du « potentiel de lipophilie moléculaire » (valeurs extrêmes, distances entre extrema) permet d'affiner les prédictions de pharmacophores [36]. Un exemple de relation entre propriétés conformationnelles et lipophiles d'une série de modulateurs allostériques des récepteurs muscariniques M_2 , mettant en jeu un index de lipophilie (« Molecular Hydrophobicity Index ») projeté sur la surface moléculaire, est présenté par **G. Vistoli et al.**

Depuis les travaux de Kauzmann [37] il est admis que l'énergie de stabilisation des protéines provient des interactions entre groupes non polaires pour former le coeur hydrophobe des molécules. Le rôle de la lipophilie dans la formation et la stabilisation de la conformation d'une protéine est unanimement reconnu. Différentes méthodes de calcul des énergies de solvatation sont utilisées dans la simulation du repliement des protéines bien que les règles qui régissent ces repliements ne soient toujours pas bien connues. Ce sera l'objet de la seconde partie de l'exposé de **M.-H. Van Eyck et al.** Les interactions non-covalentes responsables du repliement des protéines peuvent être analysées à la suite d'un calcul de « potentiel de lipophilie moléculaire ». Des programmes comme HINT (Hydrophobic Interactions) [33] présenté dans ce dossier par **G.E. Kellogg et D.J. Abraham**, ou MLPP (Molecular Lipophilicity Potential in Proteins) [38] permettent de générer des modèles de lipophilie de structures cristallines de macromolécules et en outre, ils aident à la prédiction de modèles de structures inconnues. Ils sont donc particulièrement utiles pour la prédiction de structures de protéines membranaires, par exemple des récepteurs couplés aux protéines G, dont on ne sait pas à l'heure actuelle obtenir des monocristaux. Ils permettent en outre de prédire les effets des modifications d'énergie libre engendrés lors de mutagénèses dirigées sur la stabilité des protéines. Enfin ils sont particulièrement utiles pour analyser les interactions non-covalentes lors des processus de fixation de ligands sur leurs sites actifs.

Références

1. Collander, R. *Ann. Rev. Plant. Physiol.* **1957**, 8, 335-348.
2. Hansch, C.; Quinlan, J. E.; Lawrence G. L. *J. Org. Chem.* **1968**, 33, 347-350.
3. Leo, A. J.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, 71, 525-616.
4. OECD Guidelines for Testing of Chemicals No. 107, OECD, Paris, 1992.
5. OECD Guidelines for Testing of Chemicals No. 117, OECD, Paris, 1992.
6. Danielsson, L.G.; Zhang, Y.H. *Trends Anal. Chem.* **1996**, 15, 188-196.
7. Van de Waterbeemd, H.; Mannhold, R. In: *Lipophilicity in Drug Action*, (Pliska, V.; Testa, B.; Van de Waterbeemd, H., Eds.), VCH, Weinheim, 1996; pp 401-418.
8. Mannhold, R.; Dross K. *Quant. Struct.-Act. Relat.* **1996**, 15, 403-409.
9. Wang, R. X.; Fu, Y.; Lai, L. H. *J. Chem. Inf. Computer. Sci.* **1997**, 37, 615-621.
10. Breindl, A.; Beck, B.; Clark, T.; Glen, R. C. *J. Mol. Model.* **1997**, 3, 142-155.
11. Haerberlein, M.; Brinck, T. *J. Chem. Soc. Perkin Trans.* **1997**, 289-294.
12. Schaper, K. J.; Rosado Samitier, M. L. *Quant. Struct.-Act. Relat.* **1997**, 16, 224-230.
13. Bodor, N.; Buchwald, P. *J. Phys. Chem. B.* **1997**, 101, 3404-3412.
14. Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, 55, 379-384.

15. Pearlman, R. S. In: Partition Coefficient Determination and Estimation, Dunn, W. J.; Block, J. H.; Pearlman, R. S. Eds., Pergamon, New York, 1986; pp 3-20.
16. Silla, E.; Tunon, I.; Pascual-Ahuir, J. L. *J. Comput. Chem.* **1991**, *12*, 1077-1088.
17. Petitjean, M. *J. Comput. Chem.* **1994**, *15*, 507-523.
18. Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1995**, *16*, 422-440.
19. Audry, E.; Dubost, J. P.; Colleter, J. C.; Dallet, P. *Eur. J. Med. Chem.* **1986**, *21*, 71-72.
20. Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199-203.
21. Furet, P.; Sele A.; Cohen, N. C. *J. Mol. Graphics* **1988**, *6*, 182-189.
22. Fauchère, J. L.; Quarendon, P.; Kaetterer, L. *J. Mol. Graphics* **1988**, *6*, 203-206.
23. Gaillard, P.; Carrupt, P.; Testa, B.; Boudon, A. *J. Comput. Aided. Mol. Des.* **1994**, *8*, 83-96.
24. Heiden, W.; Moeckel, G.; Brickmann, J. *J. Comput. Aided. Mol. Des.* **1993**, *7*, 503-514.
25. Hansch, C.; Leo, A.J. In: Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, 1979; pp 352-384.
26. Rekker, R. F.; de Kort, H. M. *Eur. J. Med. Chem.* **1979**, *14*, 479-488.
27. Broto, P.; Moreau, G.; Vanduycke, C. *Eur. J. Med. Chem.* **1984**, *19*, 71-78.
28. Abraham, D. J.; Leo, A. J. *Proteins. Struct. Funct. Genetics* **1987**, *2*, 130-152.
29. Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163-172.
30. Suzuki, T.; Yoshihiro, K. *J. Comput. Aided. Mol. Des.* **1990**, *4*, 155-198.
31. Klopman, G.; Li, J. Y.; Wang, S.; Dimayuga, M. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 752-781.
32. Meylan, W.; Howard, P. *J. Pharm. Sci.* **1995**, *84*, 83-92.
33. Kellogg, G. E.; Semus, S. F.; Abraham, D. J. *J. Comput. Aided. Mol. Des.* **1991**, *5*, 545-552.
34. Brasseur, R. *J. Biol. Chem.* **1991**, *266*, 16120-16127.
35. Cramer, R. D. III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5965.
36. Laguerre, M.; Dubost, J. P.; Kummer, E.; Carpy, A. *Drug. Des. Discov.* **1994**, *11*, 205-222.
37. Kauzmann, W. *Adv. Prot. Chem.* **1959**, *14*, 1-64.
38. Laguerre, M.; Saux, M.; Dubost, J. P.; Carpy, A. *Pharm. Sci.* **1997**, *3*, 217-222.