

有机化合物脂水分配系数 $\log P$ 的计算*

来鲁华 王任小 唐有祺
(北京大学物理化学研究所, 北京 100871)

关键词: 脂水分配系数, 片段加合法, 原子参数, 有机化合物

$\log P$ 作为有机化合物疏水性的量度, 在定量构效关系研究中扮演了重要的角色, 广泛应用于生物化学、药学以及环境科学等各领域. 运用计算的方法获得 $\log P$ 值由于能够弥补实验方法的不足, 在药物分子设计等应用中具有重要意义. 随着计算机辅助药物分子设计的发展, 高精度、简便易行的 $\log P$ 计算方法日益受到重视. $\log P$ 的计算方法见诸报导的已有多种^[1-7]. 这些方法各具特色, 但也都分别存在一些明显的缺点. 目前流行的片段加合法^[1-3]的出发点是将分子划分为基本片段, 每个基本片段对 $\log P$ 具有特定的贡献, 但实际上相同片段对 $\log P$ 的影响不尽相同. 另外, 片段加合法只考虑了分子的二维结构而忽视了分子的三维结构对疏水性的影响.

表 1 XLOGP 中原子的分类
Table 1 The Atom Types Used in XLOGP

No.	Symbol	Meaning	No.	Symbol	Meaning
1	C.3	C <i>sp</i> ³	12	N.pl3	N trigonal
2	C.2	C <i>sp</i> ²	13	N.am	N amide
3	C.ar	C aromatic	14	S.3	S <i>sp</i> ³
4	C.1	C <i>sp</i>	15	S.2	S <i>sp</i> ²
5	H	Hydrogen	16	S.0	S sulfoxide
6	O.3	O <i>sp</i> ³	17	S.02	S sulfone
7	O.2	O <i>sp</i> ²	18	F	Fluorine
8	N.3	N <i>sp</i> ³	19	Cl	Chlorine
9	N.2	N <i>sp</i> ²	20	Br	Bromine
10	N.1	N <i>sp</i>	21	I	Iodine
11	N.ar	N aromatic			

通过分析和研究已有的 $\log P$ 计算方法, 我们在保留片段加合法优点的基础上引入分子的微观参数藉以反映分子内精细的化学环境以及三维结构对 $\log P$ 值的影响, 提出了一种新的 $\log P$ 的计算方法——XLOGP. XLOGP 的基本思想是: 分子的 $\log P$ 值为分子中各个原子的贡献之和; 每个原子的贡献可由它在分子中所占的可达表面积 S 、所占的范

1994-07-12 收到初稿, 1994-08-31 收到修改稿. 联系人: 来鲁华. * “863” 资助项目 863-103-22-2

氏体积 V 以及它的部分电荷 q 这些微观参数来计算；具体的计算表达式与原子的类型有关：

$$\log P = \sum f_i(S_i, V_i, q_i)$$

其中 $f_i(S_i, V_i, q_i)$ 是分子中第 i 个原子对 $\log P$ 的贡献，它是以 S_i, V_i, q_i 为变量的函数。我们把常见有机化合物中的 C、H、O、N、S、F、Cl、Br、I 共 9 种元素的原子按其成键情况划分为 21 种类型，如表 1 所示。

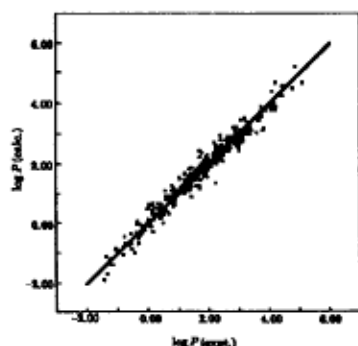


图 1 计算值和实验值的拟合直线

Fig. 1 The line fitting by the calculated and experimental $\log P$ values

The number of compounds $n=400$, correlation coefficient $r=0.982$, standard deviation $s=0.252$, and the F -factor $F=187.86$

为了得出 XLOGP 的具体模型，我们构建了含有 400 个化合物的训练集，尽量包括各种类型的有机化合物，并且尽可能保证每种类型的化合物都有足够的数量。分子的搭建在 INDIGO² 图形工作站上利用 SYBYL 软件进行。每一个分子经 Tripos 力场进行构象优化后，计算其中各原子的可及表面积 S 、范氏体积 V 和部分电荷 q 以备回归分析之用。

我们主要采用了逐步回归分析的方法以获得最后的回归方程。在每次回归分析得出一个模型后，以其对训练集拟合计算的结果以及交叉验证的计算结果为检验标准。经多次回归分析后，得到了每一类原子 f_i 的具体形式。例如对于卤族元素，

$$f(F) = -6.73V/V_i - 29.6S_q/S_i$$

$$f(Cl) = -0.0708S_q$$

$$f(Br) = 0.0120V \quad f(I) = 95.7q/V$$

式中 S_i 和 V_i 分别是指分子的总表面积和总体积， S 为该原子在分子中所占的表面积， V 为其体积， q 是其部分电荷。通过 f_i 的具体表达式将分子的微观参数引入到了 $\log P$ 的计算中。

以最终的 XLOGP 模型对训练集中 400 个化合物进行拟合计算的结果如图 1 所示。拟合直线的斜率为 1.00，截距为 -0.000141，非常接近于理想状态（斜率为 1，截距为 0）。分析计算值与实验值的误差分布可以发现，400 个化合物中约有 87% 以上计算值误差在 ± 0.40 之内，整体数据的标准偏差为 0.252，而 $\log P$ 的实验误差通常认为在 0.40 左右。因而 XLOGP 对训练集拟合计算的结果是相当令人满意的。Klopman 等计算了 1600 个化合物的 $\log P$ 值^[6]，其中有近 100 个化合物的误差大于 0.8；在我们选用的 400 个化合物中误差大于 0.8 的只有 3 个。我们又用 XLOGP 对训练集进行了交叉验证。交叉验证的相

关系数 r 值为 0.971, 标准偏差为 0.317, 约有 83% 化合物的预测值误差在 ± 0.40 之内, 误差大于 0.8 的有 7 个化合物, 说明 XLOGP 的预测能力也很强.

XLOGP 的计算精度与现有的 LOGP 计算方法相类似, 其特点是: ① 利用 SYBYL 软件搭建模型, 操作简便易行; ② XLOGP 用同一个模型就可以计算所有类型的有机化合物, 既不需要对任何基团设立标志变量, 也不需要附加任何修正因子, 模型简单而清晰, 所用的变量数目远较现行的片段加合法为少; ③ XLOGP 在根据化学环境对原子进行分类的基础上, 用原子的 S 、 V 、 q 来计算每个原子对 $\log P$ 值的贡献. 由于即使是同一类型的原子也可以有不同的 S 、 V 、 q , 因而突破了基本片段应有一致片段值的限制, 更能反映实际情况.

参 考 文 献

- 1 Rekker R F. The Hydrophobic Fragmental Constant, New York: Elsevier, 1977
- 2 Leo A. Comprehensive Medicinal Chemistry, Vol.4, Hansch C, Ed., Pergamon Press, 1990
- 3 Hansch C, Leo A. Substituent Constants for Correlation Analysis in Chemistry and Biology, New York: Wiley, 1979
- 4 Viswanadhan V N, Ghose A K, Revankar G R. *J. Chem. Inf. Comput. Sci.*, 1989, 29:163
- 5 Suzuki T. *J. Comput-Aided Mol. Des.*, 1991, 5:149
- 6 Klopman G, Wang S. *J. Comput. Chem.*, 1991, 12:1025
- 7 Bodor N, Gabanyi Z, Wong C K. *J. Am. Chem. Soc.*, 1989, 111:3783

A New Model for Calculating $\log P$ for Organic Compounds

Lai Luhua Wang Renxiao Tang Youqi

(Institute of Physical Chemistry, Peking University, Beijing 100871, China)

Abstract: $\log P$ is an important parameter for evaluating the hydrophobicity of organic compounds and other species. It is widely used in 2D-QSAR and recently in 3D-QSAR studies. On the basis of the current fragment addition method, we proposed a new algorithm called XLOGP by introducing atomic parameters like atom accessible surfaces, van der Waals volumes and partial charges. Step-wise multiple regression on a test set of 400 organic compounds were carried out. The model was then tested by cross-validation. The final correlation coefficient for the whole data set fitting and for cross-validation is 0.982 and 0.971, respectively. This shows that the model obtained is valid both at data fitting and at prediction. Compared to other algorithm, this method can discriminate a common atom or fragment in different environment.

Keywords: Partition coefficient, Fragment addition, Atom parameters, Organic compound