

[通 讯]

计算有机化合物脂水分配系数的新方法*

王任小 傅 瑛 来鲁华

(北京大学物理化学研究所, 北京 100871)

关键词: 脂水分配系数, 原子加和法, 修正项, XLOGP

有机化合物在正辛醇和水两相间的分配系数 $\log P$, 是被普遍接受的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, 广泛地应用于生物化学、药学以及环境科学等各领域. 从分子的结构出发, 运用计算的方法获得 $\log P$ 值, 能够弥补实验方法的不足, 并能提供相当的精度. 因此, 随着计算机辅助药物分子设计的发展, 简便易行的 $\log P$ 的计算方法日益受到重视, 在药物设计中的应用也愈加广泛. $\log P$ 的计算方法目前见诸报导的已有多种^[1,2]. 发展较为成熟的大致有片段加和法^[3-10]和基于分子性质的计算方法^[11,12]两类. 其中片段加和法的出发点是把分子划分为基本片段, 每种特定的基本片段具有特定的贡献值, 整个分子的 $\log P$ 值是其所含的所有片段贡献的加和. 片段加和法概念清楚, 计算简便快速, 计算结果精度较高, 适用范围广, 是目前实际应用中最为流行的方法.

在分析和研究已有方法的基础上, 我们提出了一种新的基于原子加和的 $\log P$ 计算方法 - XLOGP, 这也是我们以前所做工作的发展^[13]. XLOGP 的基本思想是: 把分子中的各种原子划分成特定的原子类型, 每种原子类型具有特定的贡献值, 分子的 $\log P$ 值为分子中各个原子贡献之和. 基于原子的加和法与基于片段的加和法相比, 它可以避免诸如片段划分方法不唯一等问题, 并且使用较少的参数, 从而使得计算方法更加简便.

我们把有机化合物中的 C、H、O、N、S、P、F、Cl、Br、I 共 10 种元素的原子按照其杂化状态、成键状况以及相连原子的状态划分为 76 种基本类型. 另外, 氰基、异硫氰酸酯基、硝基和亚硝基这四种末端基团可以当作一个整体来看待, 定义为四种“假原子”类型. 所以, 在 XLOGP 中我们一共使用了 80 种原子类型, 具体的分类方法如表 1 中所示. 对于多数较为简单的化合物, 采用基本的原子加和就可以获得满意的结果. 但是, 我们以及他人的计算结果都表明, 如果只采用基本的原子加和, 某些类型的化合物 $\log P$ 的估算值将会有明显的偏差. 这些化合物包括烃类化合物、氨基酸、含分子内氢键的化合物等等, 引起偏差的原因各异. 在这种情况下, 引入适当的修正项来说明分子内这些特殊的长程作用的影响就变得非常必要了. 在 XLOGP 中, 我们定义了 4 类修正项, 它们是: (1)“疏水碳原子”, 即烃类化合物中 sp^3 和 sp^2 杂化的碳原子; (2) α -氨基酸的标志变量, 当某化合物为 α -氨基酸时, 此项为 1, 否则为 0; (3) 分子内氢键; (4) 卤素原子 1-3 相互作用. 这些修正项的引入在改善计算结果的过程中起到了非常明显的作用.

为了得到每种原子类型以及修正项的疏水贡献值, 我们构建了含有 1831 个有机化合物的训练集. 为了使回归分析得到可靠的结果, 训练集中包括了各种类型的化合物, 并且尽可能保证每种类型的化合物具有足够的数量. 分子的搭建在 INDIGO2 图形工作站上利用 SYBYL 软件进行.

1996-10-08 收到初稿, 1996-11-04 收到修改稿. 联系人: 来鲁华. * “863” 高技术资助项目.

表 1 XLOGP 中使用的原子类型和修正项

Table 1 Atom types and correction factors used in XLOGP

No.	Description ^a	Contribution	No.	Description ^a	Contribution
	<i>sp</i> ³ carbon in			<i>sp</i> ³ carbon in	
1	CH ₃ R (<i>P</i> ₁ =0) ^b	0.484	2	CH ₃ R (<i>P</i> ₁ ≠ 0) ^b	0.168
3	CH ₃ X	-0.181	4	CH ₂ R ₂ (<i>P</i> ₁ =0)	0.358
5	CH ₂ R ₂ (<i>P</i> ₁ ≠ 0)	0.009	6	CH ₂ RX	-0.344
7	CH ₂ X ₂	-0.439	8	CHR ₃ (<i>P</i> ₁ =0)	0.051
9	CHR ₃ (<i>P</i> ₁ ≠ 0)	-0.138	10	CHR ₂ X	-0.417
11	CHRX ₂ , CHX ₃	-0.454	12	CR ₄ (<i>P</i> ₁ =0)	-0.378
13	CR ₄ (<i>P</i> ₁ ≠ 0)	0.223	14	CR ₃ X	-0.598
15	CR ₂ X ₂	-0.396	16	CRX ₃	-0.699
17	CX ₄	-0.362			
	<i>sp</i> ² carbon in			<i>sp</i> ² carbon in	
18	R=CH ₂	0.395	19	R=CHR	0.236
20	R=CHX	-0.166	21	X=CHR, X=CHX	1.726
22	R=CR ₂	0.098	23	R=CRX, R=CX ₂	-0.108
24	X=CR ₂ , X=CXR	1.637	25	X=CX ₂	1.774
	aromatic carbon in			aromatic carbon in	
26	R...C(H)...R	0.281	27	R...C(H)...X	0.142
28	X...C(H)...X	0.715	29	R...C(R)...R	0.302
30	R...C(X)...R	-0.064	31	R...C(R)...X	0.079
32	R...C(X)...X	0.200	33	X...C(A)...X	0.869
34	A...C...(A)...A	0.316			
	<i>sp</i> carbon in			<i>sp</i> carbon in	
35	R≡CH	0.054	36	R≡CR, R≡CX	0.347
	hydrogen				
37	H	0.046			
	<i>sp</i> ³ oxygen in			<i>sp</i> ³ oxygen in	
38	R-OH (<i>P</i> ₁ =0)	-0.399	39	R-OH (<i>P</i> ₁ ≠ 0)	-0.029
40	X-OH	-0.330	41	R-O-R	0.397
42	R-O-X, X-O-X	0.068	43	<i>P</i> ₁ -O- <i>P</i> ₁ (in ring)	0.327
	<i>sp</i> ² oxygen in			<i>sp</i> ² oxygen in	
44	O=R	-2.057	45	O=X	0.218
	<i>sp</i> ³ nitrogen in			<i>sp</i> ³ nitrogen in	
46	R-NH ₂ (<i>P</i> ₁ =0)	-0.582	47	R-NH ₂ (<i>P</i> ₁ ≠ 0)	-0.449
48	X-NH ₂	-0.774	49	R-NH-R	0.040
50	R-NH-X, X-NH-X	-0.381	51	NR ₃	0.443
52	NR ₂ X, NRX ₂ , NX ₃	-0.117			
	<i>sp</i> ² nitrogen in			<i>sp</i> ² nitrogen in	
53	R=NH, R=NR	-2.052	54	R=NX	-1.716
55	X=NR	0.321	56	X=NX	-0.921
	aromatic nitrogen				
57	A...N...A	-0.704			
	trigonal planar nitrogen ^c in			trigonal planar nitrogen ^c in	
58	R-NH-R	0.119	59	R-NH-X, X-NH-X	1.192
60	A-NH-A (in ring)	0.434	61	NA ₃	0.587
62	NA ₃ (in ring)	0.668			
	amide nitrogen in			amide nitrogen in	
63	-NH ₂	-0.791	64	-NHR, -NHX	-0.212
65	-NR ₂ , -NRX	0.016			
	<i>sp</i> ³ sulfur in			<i>sp</i> ² sulfur in	
66	A-SH	0.752	67	R-S-R, R-S-X	1.071
68	<i>P</i> ₁ -S- <i>P</i> ₁ (in ring)	0.964			

(continued)

No.	Description ^a	Contribution	No.	Description ^a	Contribution
69	sp^2 sulfur S=R	-1.817			
70	sulfoxide sulfur A-SO-A	-1.214			
71	sulfone sulfur A-SO ₂ -A	-0.778			
72	halogen F	0.493	73	halogen Cl	1.010
74	Br	1.187	75	I	1.489
76	phosphur A-P(O(A))-A	-0.802		phosphur	
77	terminal groups -C≡N	-0.256	78	terminal groups -N=C=S	1.626
79	-N=O	0.077	80	-NO ₂	0.264
1	correction factors hydrophobic carbon	0.19	2	correction factors amino acid indicator	-2.27
3	internal H-bond	0.60	4	halogen 1-3 pair ^d	0.08
5	halogen 1-3 pair ^e	-0.26			

a: - represents single bond; = represents double bond; ≡ represents triple bond; ... represents aromatic bond; R represents any group linked through carbon; X represents any heteroatom(O, N, S, P and halogens); A represents any atom except for hydrogen; P_i represents any atom involved in conjugated system, such as sp^3 , sp^2 hybridized atom and aromatic atoms.

b: $P_i=0$ means that there are no P_i atom in the neighboring atoms while $P_i \neq 0$ means this atom is connected to a conjugated system.

c: This is a special atom type defined in Tripos force field.

d: Two halogen atoms attached to the same carbon, with fluorine involved in the pair.

e: Two halogen atoms attached to the same carbon, without fluorine involved in the pair.

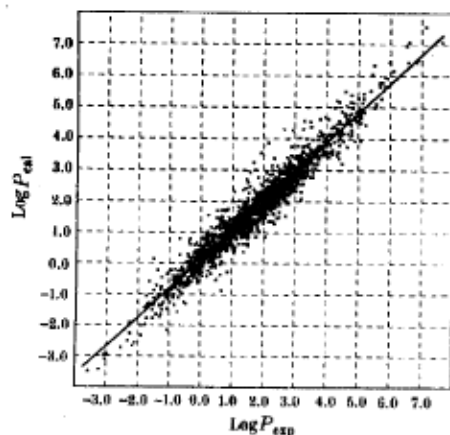


图1 计算值和实验值的拟合直线

Fig.1 The correlation between the measured $\log P_{exp}$ and calculated $\log P_{cal}$ values

每个分子的构象都经过 Tripos 力场的优化. 较复杂的分子的构象均通过系统搜索来确定. 所有分子的最终结构均以 MO1.2 文件格式存储.

多元线性回归的方法被用来关联 XLOGP 的模型和训练集中化合物的 $\log P$ 实验值. 回归分析的结果为: 样本数 $n=1831$, 回归系数 $r=0.968$, 标准偏差 $SD=0.37$. 回归分析所得出的各种原子类型以及修正项的贡献值均列在表 1 中. 1831 个化合物的 $\log P$ 值的计算结果与实验值的比较如图 1 中所示. 由图中可以看出, 计算值与实验值之间得出了较好的相关关系. 整个训练集的标准偏差为 0.37, 小于通常认为的实验值的误差 0.40. 因此, XLOGP 给出的结果是相当令人满意的. 我们也用

XLOGP 对整个训练集进行了交叉验证. 交叉验证的结果为: 回归系数 $r=0.963$, 标准偏差 $SD=0.40$,

其标准偏差也与实验误差相当,从而证明了 XLOGP 出色的预测能力.

基于我们得出的最后模型,我们编制了一个 C 语言的程序,可以对给定的化合物自动进行 XLOGP 的计算,给出 $\log P$ 的计算值.此程序唯一需要的输入是存储成 MOL2 格式的分子结构,所有的 SYBYL 用户可以很方便地使用它进行计算.

与已见诸报导的其它几种同类型的计算方法^[6-10]相比,我们的方法具有以下优点:(1)采用了更大的训练集,从而更能代表有机化合物的全貌;(2)使用了更少的参数,简化了计算方法.这主要是由于我们得出了一套合理的原子类型划分方法和引入了合理的修正项;(3)得出了总体上与实验值更为相近的计算结果,而且对整个训练集进行了交叉验证,证实了其预测能力.与目前国际上最负盛名的基于片段加和法的 CLOGP 相比,我们的方法可以给出精度相近的结果并且大大简化了使用.我们相信, XLOGP 会在定量构效关系研究的实际应用中发挥它应有的作用.

参 考 文 献

- 1 Leo A. *Chem. Rev.*, **1993**, **93**:1281
- 2 朱龙观,俞庆森. *国外医学药学分册*, **1995**, **22**:206
- 3 Hansch C, Leo A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*, New York, Wiley, 1979
- 4 Leo, A. In *Comprehensive Medicinal Chemistry*, Vol.4, Hansch, C. Ed., Oxford, Pergamon, **1990**:295-319
- 5 Rekker R F. *The Hydrophobic Fragment Constant*, New York, Elsevier, 1977
- 6 Suzuki T, Kudo Y. *J. Comput. -Aided Mol. Des.*, **1990**, **4**:155
- 7 Moriguchi I, Hirono S, Liu Q, et al. *Chem. Pharm. Bull.*, **1992**, **40**:127
- 8 Klopman G, Li J Y, Wang S, et al. *J. Chem. Inf. Comput. Sci.*, **1994**, **34**:752
- 9 Broto P, Moreau G, Vandycke C. *Eur. J. Med. Chem.*, **1984**, **19**:71
- 10 Ghose A K, Pritchett A, Crippen G M. *J. Comput. Chem.*, **1988**, **9**:80
- 11 Klopman G, Namboordiri K, Schochet M. *J. Comput. Chem.*, **1985**, **6**:28
- 12 Bodor N, Gabanyi Z, Wong C K. *J. Am. Chem. Soc.*, **1989**, **111**:3783
- 13 来鲁华,王任小,唐有祺. *物理化学学报*, **1994**, **10**:963

A New Method for Calculating Partition Coefficients of Organic Compounds

Wang Renxiao Fu Ying Lai Luhua

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

Abstract A new method is presented for the calculation of octanol/water partition coefficients. On the basis of summation of atomic contributions, our algorithm, namely XLOGP, also incorporate correction factors into the calculation. Multivariate regression analysis was performed on a training database of 1831 organic compounds with diverse structures to give the final model. The correlation coefficients for the whole set fitting is 0.968 and the standard deviation is 0.37. The result shows that our model is accurate enough for $\log P$ estimation in QSAR studies. Compared to other similar approaches, our method gives better results and is more convenient to use.

Keywords: Partition coefficient, Atomic contribution, Correction factor, XLOGP