

# 化合物脂水分配系数计算软件及比较研究

苏笠, 杨劲\*, 王友群, 王广基

(中国药科大学药物代谢与药理学研究中心, 南京 210009)

**摘要** 目的: 介绍几种常用的化合物脂水分配系数计算软件, 并比较它们的准确度和适用范围等。方法: 从分子结构出发, 运用 CLOGP、ACD/LogP、Pallas 和 Molinspiration 4 种计算软件获得  $\log P$  值, 比较其计算准确度。结果: 本文选择的 4 种计算软件均能比较准确的计算出化合物的  $\log P$  值。结论: 运用软件计算的方法得到大批量化合物的  $\log P$  值, 有助于计算机辅助药物设计。

**关键词** 定量构效关系; 计算机辅助药物设计; 脂水分配系数; 分子结构; CLOGP; Pallas; ACD/LogP; Molinspiration

**中图分类号** R911 **文献标识码** A **文章编号** 1000-5048(2008)02-0178-05

## Introduction and comparison of softwares used in calculating lipid/water partition coefficient of compounds

SU Li, YANG Jin\*, WANG You-qun, WANG Guang-ji

Center of Pharmacokinetics and Drug Metabolism, China Pharmaceutical University, Nanjing 210009, China

**Abstract** **Aim:** To introduce the available softwares used to calculate the lipid/water partition coefficients of compounds and to compare the estimation accuracy. **Methods:** Based on the molecular structures selected, the values of  $\log P$  were calculated using softwares CLOGP, Pallas, ACD/LogP and Molinspiration, and the estimation accuracy was assessed one by one. **Results:** The four softwares could accurately calculate the values of  $\log P$ . **Conclusion:** Application of these softwares in the calculation of  $\log P$  values of compounds facilitates computer-aided drug design.

**Key words** QSAR; computer-aided drug design; lipid/water partition coefficient; molecular structure; CLOGP; Pallas; ACD/LogP; Molinspiration

脂水分配系数( $\log P$ )是用来表示化合物的亲脂性、透过生物膜的性能及与受体(包括酶)间疏水性结合力的一种参数。疏水性影响药物吸收、生物活性、疏水性药物受体相互作用、分子代谢变化毒性等,在定量构效关系的研究中很重要<sup>[1,2]</sup>。新药开发过程中,可以所得定量构效和构效相关式为基础,设计生物活性好并具有良好药理学参数的化合物。在分子设计阶段,可以预先计算所设计分子的  $\log P$  值,进行虚拟筛选,从而节省大量的时间和财力,提高工作效率。总之,用计算的方法得到化合物的  $\log P$  值的目的之一就是为药物设计服务<sup>[3]</sup>。

### 1 基础与原理

脂水分配系数( $\log P$ )为化合物在非水相中的平衡浓度与其在水相中的中性形式平衡浓度的比值的对数,公式表达如下:

$$\log P = \log(c_o/c_w)$$

其中,  $c_o$  和  $c_w$  分别表示化合物在非水相中的平衡浓度与其在水相中的中性形式的平衡浓度,  $\log P$  值越大则脂溶性越强。

$\log P$  数值的获得有实验法和计算法两种。实验法可靠,但花费较大,并需要预先合成化合物,不能进行虚拟筛选;计算法简便快捷,但是计算结果不够精确。

Log  $P$  的计算方法有两种:一是通过有机化合物的片断加和性来进行计算,二是通过化合物其他物理化学参数以及其与脂水分分配系数的内在联系来计算。

片断加和法的原理是将分子划分为基本片断,每种特定的基本片断具有特定的贡献值,整个分子的 log  $P$  是所有片断的加和。片断加和法计算简便、快捷,计算结果精度较高、适用范围广,目前常用的脂水分分配系数计算软件都是基于此原理开发的。

## 2 计算软件介绍

随着计算机辅助药物分子设计的发展,从分子结构出发,运用计算的方法获得 log  $P$  值能够弥补实验方法的不足,因此涌现出大量专业软件。

目前常用的 log  $P$  计算软件有 CLOGP 软件、Pallas 软件的 ProLogP 模块、ACD/LogP 的 I-Lab 模块、SYBYL 软件的 ClogP/CMR 模块和 Molinspiration 网络版等。本文主要介绍 CLOGP、Pallas 软件的 ProLogP 模块、ACD/LogP 模块和 Molinspiration 网络版 4 种计算工具。

### 2.1 CLOGP 软件

CLOGP 是基于碎片常数加和法的原理,1979 年实现程序化,中间几经修改,1994 年又有新版本 (Biobyte 公司),商品名为 MalogP<sup>TM</sup>,目前发展为 Biolum。CLOGP 曾经是脂水分分配系数计算领域的黄金标准,随着计算化学的发展,目前流行的计算软件准确度比 CLOGP 有明显的提高。

### 2.2 Pallas 软件

Pallas 软件是 CompuDrug 公司的产品,用于计算 log  $P$  的部件称作 ProLogP。log  $P$  的计算分为 5 个步骤<sup>[4]</sup>:首先对结构式进行分解,然后探查碎片相互作用和邻近效应的测定根据数据库,计算各碎片的贡献值及它们之间的相互影响,最后把各个数据库计算出来的值通过公式计算出综合 log  $P$  值。

Pallas 软件中的 ProLogP 默认计算公式为:
$$\log P_{\text{combined}} = 0.019 \log P_{\text{CDI-Rekker}} + 0.235 \log P_{\text{ANNLogP}} + 0.746 \log P_{\text{ANNLogP2005}}$$

这是 CompuDrug 公司根据约 8 000 个化合物的实验数据,通过神经网络方法进行数据分析得出的经验公式,其中 CDI-Rekker、ANNLogP 和 ANNLogP 2005 为默认数据库。

由 ProLogP 计算的 log  $P$  值是根据碎片数据库直接计算而得,碎片数据库的容量、构成、权重直接

影响最后的计算值。目前,ProLogP 7.0 有 3 个根据不同方法建立的基本数据库:CDI-Rekker、Atomic 和 Atomic 5。

CD-Rekker 数据库是根据 Rekker 对化合物疏水性碎片的研究成果<sup>[5]</sup>建立的,其中包括约 1 000 个化合物的实验值。此外,一些在这些化合物的基础上的修饰体的值也加入到其中,从而使该方法的适用性更强。

Atomic 数据库是使用原子作为片断进行研究的,然而在很多情况下这种计算方法会因为数据库缺少相应的原子类型而造成计算错误。

Atomic 5 数据库也是把原子当作碎片来进行计算的,它和 Rekker 数据库一样,加入修饰步骤,从而使其预测更准确。该数据库有如下特点:加入了新的原子类型;碎片的相互作用中加入了离域作用、氢键作用等;贡献值通过线性回归进行重新计算。

此外还有 Atomic 6 数据库、ANNLogP 和 ANNLogP 2005 数据库。

ProLogP 一个显著特点就是,进行具体工作时,可以根据经验选择数据库并决定他们的权重值,这可以极大的提高预测系列化合物 log  $P$  值的准确性。

### 2.3 ACD/LogP (I-Lab) 软件

ACD/LogP 用于计算 log  $P$  的程序为 ACD/LogP<sup>[5]</sup>。ACD/LogP 计算中性分子的 log  $P$  值大部分误差都小于  $\pm 0.3$  的范围。该算法是建立在 18 400 个 log  $P$  实验值的数据库基础上,拥有多于 1 200 个基本功能基团。

2.3.1 ACD/LogP 计算步骤 首先将化合物分解成多个碎片,然后在内部数据库中搜索相同的碎片,如果有内部数据库中不存在的碎片,系统将运用二级算法估算该碎片与数据库中碎片的差值,最后系统将权衡整体的互变异构体、离子平衡,计算误差的影响,得出最终结果。

2.3.2 ACD/LogP 的特性 同时显示化合物结构和 log  $P$  计算结果;建立计算过程备忘,分子中用于计算 log  $P$  的各个原子碎片的贡献值均有显示,如果有对计算结果有重要影响的结构因素,例如互变异构体,ACD/LogP 会在结果中显著地显示出来;该软件同时还可以计算 Polar Surface Area (PSA)、Freely Rotatable Bonds (FRB),和 Lipinski's "Rule-of-5" 参数;能评价中性分子的生物浓度指数 (bioconcentration factor) 和吸收系数;可以将新的数值加入到数据库中,提高预测的准确性;图片数

数据库评价趋势;能同时计算化合物的各种属性,在历史窗口一起显示,且能同时以 SDfile 输出多个化合物计算结果;可以将新的数值加入到数据库中,提高预测的准确性。

2.3.3 计算限制 能用于计算的化合物,其结构中只能包括下列原子:H、O、S、P、N、F、Cl、Br、I、Se、Si、Ge、Pb、Sn、As 和 B;log *P* 不能计算带电荷的分子(两性氨基酸分子、多肽分子、与氧结合四价氮结构例外);能计算的化合物所含原子不能多于 255 个;程序不能辨别立体异构体<sup>[5]</sup>。

## 2.4 Molinspiration 网络版

Molinspiration 网络版预测 log *P* 的模块版本号是:mib batch molecule processing V2007.03,基本原理也是基于基团贡献的计算方式<sup>[6]</sup>。它通过超过 12 000 个类药性好的化合物的实验数值作为训练系统,来预测恰当的 log *P* 值。在该系统中,通过 35 个小的基本碎片,185 个大的碎片来表征分子内氢的键合对 log *P* 的贡献和电荷相互影响。Molinspiration 计算 log *P* 的方法非常强大,它可以预测几乎全部的有机分子和大多数有机金属分子。

通过对 12 202 个类药性比较好的化合物进行回归分析,其复测定系数为 0.944,标准误差为 0.428,平均绝对误差为 0.328,模型比较稳定;对于 12 202 个分子,50.5% 的 log *P* 值预测误差小于

0.25,80.2% 的 log *P* 值预测误差小于 0.5,96.5% 的 log *P* 值预测误差小于 1.0,只有 3.5% 的误差大于 1.0,准确性也较好。可用于新化合物的预测。

Molinspiration 运用很广泛,其计算出来的 log *P* 值被证明与不同药物体内转运性质有很好的相关性,它计算的稳定性和良好的预测能力使它在 ZINC 数据库 (<http://zinc.docking.org/>) 虚拟筛选中有重要的运用。

Molinspiration 有在线网络免费版本和付费商业版本。其网络版目前是免费的,预测化合物数目每月限制在 100 以内。

## 3 计算软件准确性比较

log *P* 的原始值参见参考文献[7],该网络数据库提供了约 4 300 种上市药物和类药性好的化合物的具体参数。从中随机选取了 100 个药物,用 CLOGP、ACD/LogP、Molinspiration 和 Pallas 这 4 种软件计算 log *P*,结果见表 1。

由于篇幅关系,表 1 中仅附上 30 个化合物的计算数据;表 1 中加下划线的数据是与实验值最接近的计算结果;ACD/LogP 输出结果含有置信区间,这里仅取其中值,例如:Amphetamine(实验值 = 1.799),计算结果为 1.81 + / - 0.20,本文选取平均值 1.81。

Table 1 Calculation of log *P* values by 4 softwares

Compd.	Experimental value	ACD/LogP	Molinspiration	Pallas	CLOGP
Acetazolamide	-0.22	<u>-0.260</u>	-1.146	-0.04	-1.130 9
Acyclovir	-1.31	-1.760	-1.614	<u>-1.48</u>	-2.422 4
Amphetamine	1.80	<u>1.806</u>	1.323	1.57	1.742 0
Aspirin	1.43	1.190	<u>1.434</u>	1.56	1.023 5
Atropine	2.90	1.528	1.767	<u>1.82</u>	1.299 2
Benazepril	3.22	<u>3.860</u>	2.524	2.29	1.823 7
Benzocaine	1.54	1.945	<u>1.560</u>	1.66	1.923 0
Bepidil	5.15	5.800	4.965	<u>5.23</u>	6.200 2
Carbenicillin	1.22	1.009	1.327	<u>1.32</u>	1.637 1
Carbenoxolone	6.96	7.083	5.597	4.33	<u>7.001 2</u>
Chlorhexidine	4.07	4.576	<u>3.989</u>	1.05	4.810 0
Chlorpheniramine	3.56	<u>3.389</u>	3.310	3.81	3.148 0
Dexamethasone	1.77	1.870	2.056	2.10	<u>1.785 2</u>
Diltiazem	3.14	3.634	3.336	<u>3.24</u>	3.647 1
Dicoumarol	3.45	<u>3.548</u>	3.325	2.26	3.657 6
Fluoxymesterone	2.54	2.170	2.761	<u>2.55</u>	2.654 0
Fenoprofen	3.64	3.844	3.891	3.86	<u>3.821 0</u>
.....					
Indapamide	2.09	<u>2.095</u>	2.455	2.53	2.957 0
Mephénytoin	1.52	1.400	<u>1.706</u>	1.23	1.741 0
Metoclopramide	2.46	2.216	<u>2.535</u>	2.54	2.229 4

(Continued)

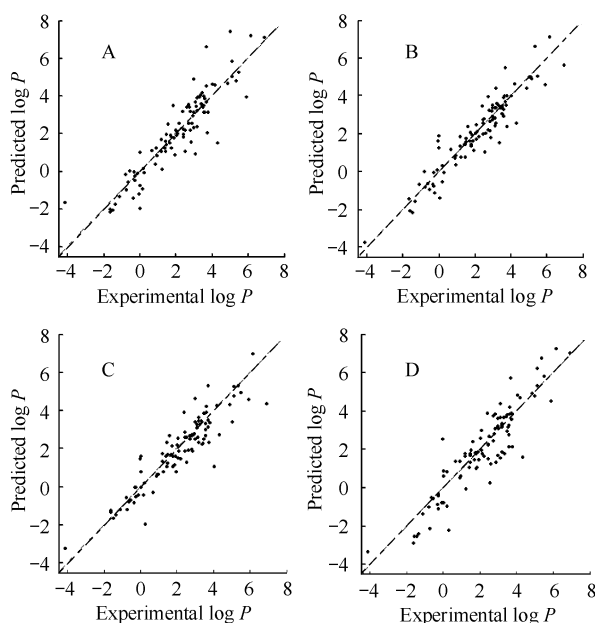
Compd.	Experimental value	ACD/LogP	Molinspiration	Pallas	CLOGP
Mexiletine	2.04	2.160	<u>2.002</u>	2.16	2.569 0
Pentazocine	4.21	4.535	4.594	<u>4.24</u>	4.667 0
Quinacrine	5.39	4.787	6.563	<u>5.29</u>	6.722 8
Ramipril	3.15	<u>3.409</u>	2.398	2.54	1.539 9
R-mephobarbital	1.92	1.771	1.592	1.55	<u>1.851 0</u>
Succinic acid	-0.60	<u>-0.59</u>	-0.655	-0.80	-0.525 6
Testosterone	3.60	<u>3.48</u>	3.245	3.30	3.219 0
Trifluridine	-0.52	0.01	-0.974	-0.71	<u>-0.565 6</u>
Trimeprazine	5.16	4.98	4.854	4.75	<u>5.298 0</u>
Warfarin	3.51	<u>3.417</u>	3.031	2.46	2.901 3

线性回归的方法被用来关联计算数值和原始值,回归分析所得的结果见表2。其中,表2中相关系数是指4种软件计算数值分别与实验数据的相关系数<sup>[8,9]</sup>。由图1可以看出,计算值和原始值之间得出了较好的相关关系。4种计算软件的标准偏差在0.47到0.86之间,稍大于通常认为的实验值的误差0.40,不过由于选取的比较样本只有100个,4种软件给出的结果是相当令人满意的。

**Table 2** Data comparison of the different softwares

Name	MAE <sup>a</sup>	STD <sup>b</sup>	r
CLOGP	0.691	0.857	0.951
ACD/LogP	0.622	0.78	0.951
Molinspiration	0.512	0.469	0.972
Pallas	0.589	0.665	0.961

<sup>a</sup> Mean absolute error; <sup>b</sup> Standard deviation



**Figure 1** Simulation of experimental log *P* and predicted log *P* by different softwares

A: ACD/LogP; B: Molinspiration; C: Pallas; D: CLOGP

CLOGP、ACD/LogP、Molinspiration 和 Pallas 的计算值中最接近原始数据的个数分别为 19、23、31 和 27; 误差  $\leq \pm 0.8$  的数据数分别为 65%、81%、84% 和 79%; 误差  $\leq \pm 0.5$  的数据数分别为 45%、53%、63% 和 59%。综合表 2 的结果可以看出: ACD/LogP、Molinspiration 和 Pallas 这 3 个软件的计算准确度比 CLOGP 有明显的提高; 就这 100 个化合物而言, Molinspiration 的计算准确度是最好的。

另外我们考虑一些特殊化合物作为测试样本, 来比较这几种软件的计算准确度及适用范围, 这里选取了金属化合物进行计算, 结果见表 3。

通过回归分析比较, Molinspiration 的标准偏差和相关系数分别为 1.35, 0.959; CLOGP 的标准偏差和相关系数分别为 12.28, 0.851。Molinspiration 的计算准确度优于 CLOGP。而 ACD/LogP 和 Pallas 不能计算含金属的化合物。

**Table 3** Log *P* values of metallic compounds

Compd.	Experimental values	ACD/LogP	Molinspiration	Pallas	CLOGP
<b>1</b>	1.96	—	1.84	—	1.91
<b>2</b>	-0.55	—	-0.37	—	-1.72
<b>3</b>	-2.08	—	-1.72	—	-2.89
<b>4</b>	-2.70	—	-2.93	—	-3.22
<b>5</b>	-2.19	—	-4.89	—	-1.68
<b>6</b>	6.97	—	8.55	—	15.98
<b>7</b>	1.56	—	0.55	—	6.09
<b>8</b>	1.95	—	0.53	—	2.41
<b>9</b>	-4.22	—	-4.85	—	-1.36
<b>10</b>	1.87	—	1.88	—	-1.32
STD			1.35		12.28
r			0.959		0.85

**1**: 2-(Ethylmercuri-thio) benzoic acid; **2**: 3-Mercuri-4-aminobenzenesulfonamide; **3**: Adenosine mononucleotide; **4**: Adenosine-5'-diphosphate; **5**: Cisplatin; **6**: Ferrocene iron; **7**: Delta-bis(2,2'-bipyridine)-(5-methyl-2,2'-bipyridine)-C2-adamantane ruthenium (II); **8**: Delta-bis(2,2'-bipyridine) imidazole osmium (II); **9**: Desferal; **10**: Mercury diiodide

#### 4 讨论与结论

本文随机选取的 100 个化合物其相对分子质量基本都在 400 以内,  $\log P$  值介于  $-4.102 \sim 6.96$ 。对这 100 个化合物而言, Molinspiration 计算结果更准确一些。

CLOGP 曾经是脂水分配系数计算领域的黄金标准, 目前常用的其他计算软件都是在其算法基础上发展而来, 计算准确度有明显的提高。

Pallas 软件拥有多个数据库, 在实际工作中可根据经验选择数据库并决定其权值, 例如对同系列化合物计算时可根据经验, 选择合适的数据库和权值, 提高计算的准确度。另外该软件还有 LogD 计算功能, 实际应用意义很大。

ACD/Log  $P$  计算结果带有置信区间, 在实际工作中, 对虚拟合成的有机分子进行筛选时可利用 ACD/Log  $P$  的置信区间, 降低误筛选率。我们在工作中还可以将新的实验数据加入数据库, 提高计算准确性。

Molinspiration 网络版几乎可以计算全部的有机分子、大多数有机金属分子和带有电荷的分子。实际工作中碰到特殊化合物时可使用该软件。

总之,  $\log P$  计算值有误差, 这种误差是不可避免的, 计算值不能替代实验值。在应用该参数进行

定量构效关系的研究工作时, 要有清醒的认识。

#### 参考文献

- [1] 卡米尔·乔治·维尔穆特. 创新药物化学[M]. 第2版. 广州: 广东世界图书出版公司, 2005: 447-479.
- [2] 王连生(Wang LS), 韩朔滕(Han SK), 孔令仁(Kong LR), 等. 分子结构、形式与活性[M]. 北京: 化学工业出版社, 1997: 26-40.
- [3] 陈凯先(Chen KX), 蒋华良(Jiang HL), 嵇汝运(Ji RY). 计算机辅助药物设计——原理、方法及应用[M]. 上海: 上海科技出版社, 2000: 15.
- [4] PALLAS 软件参考文件(版本号 3.5.1.2, CompuDrug 公司).
- [5] Rekker RF, DeKort HM. The hydrophobic fragmental constant [J]. *Eur J Med Chem*, 1979, **14**: 479-488.
- [6] Robin Martin, Sanjivanjit K. Bhal. ACD/Labs freeware short start tutorial [EB/OL]. (2007-3-25) [2007-10-20]. [http://www.acdlabs.com/download/docs/logp\\_start\\_tutorial.pdf](http://www.acdlabs.com/download/docs/logp_start_tutorial.pdf).
- [7] Anonymous. Molinspiration application tutorial [EB/OL]. (2007-10-1) [2007-10-22]. [www.molinspiration.com](http://www.molinspiration.com).
- [8] Wishart DS, Knox C, Guo AC, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration [J]. *Nucleic Acids Res*, 2006, **D668-D672**.
- [9] 孙振球(Sun ZQ). 医学统计学[M]. 北京: 人民卫生出版社, 2002: 342-346.
- [10] 王静龙(Wang JL). 现代外国统计学优秀著作译丛——应用线性回归[M]. 第二版, 北京: 中国统计出版社, 1998: 20-65.

#### · 会 讯 ·

### 第九届全国青年药学工作者最新科研成果交流会

为广泛交流我国青年药学工作者的最新科研成果, 促进青年药学工作者健康成长, 加速培养药学科技人才, 为青年学者提供学术交流的机会和平台, 由中国药学会主办的第九届全国青年药学工作者最新科研成果交流会将于 2008 年 5 月底在北京市召开。

#### 联系方式:

中国药学会学术部: 孙文虹、鲁毅、王爱国

地址: 北京市朝阳区建外大街 4 号建外 SOHO 九号楼 18 层 1802(100022)

电话: 010-58699275-819/820

传真: 010-58699125

E-mail: sunwenhong2002@163.com 或 yxhwag@sohu.com