

水溶液中有有机化合物 pK_a 值的量子化学计算研究

丁晓琴, 孙 杨, 丁俊杰, 陈冀胜
(北京药物化学研究所, 北京 102205)

摘要 采用密度泛函量子化学计算方法, 在 B3LYP/6-31 + G(d,p) 水平上采用极化连续模型(PCM), 对醇、酚、羧酸、芳香胺和脂肪胺等多类化合物在水溶液中的 pK_a 值进行了计算. 与实验值比较, 多数化合物绝对误差小于 1 个 pK_a 单位, pK_a 计算值与实验值相关系数均大于 0.94. 可见, 采用量子化学计算的方法可准确预测化合物的 pK_a 值, 并可以扩展到药物结构与性能关系、ADMET 虚拟筛选、有机合成路线设计和优化等研究领域.

关键词 pK_a 值; 密度泛函; 连续极化模型

中图分类号 O641

文献标识码 A

文章编号 0251-0790(2009)S0-0103-06

药物在水或非水溶液中的离子化过程是一种常见的化学现象, 在诸多化学或生物化学反应过程中起着相当重要的作用. 化合物的酸度通常以公式(1)的平衡常数 K_a 来定量或者用 pK_a 来简化表示:



$$pK_a = -\lg \frac{[H^+]_{\text{sol}} [A^-]_{\text{sol}}}{[HA]_{\text{sol}}} = \frac{1}{2.303RT} \Delta G_{\text{sol}}^* \quad (2)$$

式(2)中 ΔG_{sol}^* 为溶液相质子解离反应的自由能变化^[1]. 正是由于化合物离子化过程的常见性和重要性, 多年来有机或无机化合物的溶剂化自由能和酸度成为了化学家的一个主要研究对象, 近年来建立了多种微量的实验方法来测定其 pK_a , 并且将实验测定值用于组合化学药物筛选^[2].

但是, 运用具体的实验方法精确测定各种化合物的溶剂化能或 pK_a 是一项繁琐的工作, 尤其对于一些反应中间体、极强酸或极弱酸, 仅仅采用实验的方法往往显得力不从心. 因此, 寻找一种较为精确的理论计算方法来补偿现有实验技术上的不足与缺陷, 或者替代那些冗长乏味的实验操作过程, 以至完全用计算的方法得到化合物的溶剂化能或 pK_a 成为了众多理论化学家的重要研究方向. 同时, 作为化合物的主要理化性质之一, pK_a 的精确计算在药物分子设计、反应机理计算、合成路线设计以及量子化学理论等中具有重要的地位. 经过长期努力, 目前已经总结出多种理论模型和计算方法, 其中部分方法的计算结果与实验值相比, 已经达到相当的精确度.

1 溶液相化合物 pK_a 的计算方法

pK_a 的计算通常采用两种方法. 一种是根据公式(1)的定义直接利用如下热力学循环(图1)通过公式(2)和(3)计算其溶液相质子解离反应自由能得到^[3]. 但是由于量子化学计算不便于处理与质子有关的热力学数据, 因此计算中通常需要利用质子气相自由能和溶液相溶剂化自由能的实验数据. 此外, 由于计算中涉及到不同电荷的化合物, 往往也会带来较大的误差.

$$\Delta G_{\text{sol}}^* = \Delta G_{\text{gas}}^* + \Delta G_{\text{solv}}^*(A^-) + \Delta G_{\text{solv}}^*(H^+) - \Delta G_{\text{solv}}^*(HA) \quad (3)$$

式(3)中, ΔG_{gas}^* 为气相质子解离反应自由能, ΔG_{sol}^* 为溶液相质子解离反应自由能, ΔG_{solv}^* 为溶剂化自由能, 参考态均为 1 mol/L.

为此, 在 pK_a 的计算中也常常采用另外一种方法, 即利用等键反应公式(4)通过图2所示的热力学循环来计算其溶液相中质子转移反应自由能的变化得到. 其中 B^- 可以是参考酸 HB 的共轭碱, 也可

以是参考溶液. 计算中既可以使用 HB 的实验值, 也可以使用计算值.

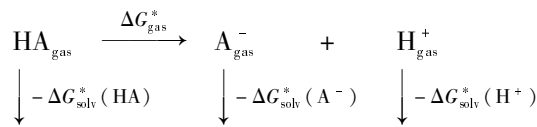


Fig. 1 Thermodynamic cycle 1

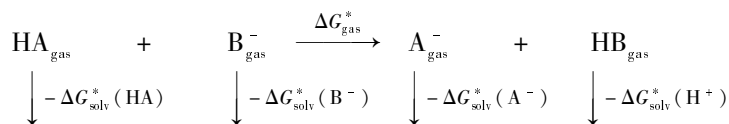


Fig. 2 Thermodynamic cycle 2



$$\Delta G_{\text{sol}}^* = \Delta G_{\text{gas}}^* + \Delta G_{\text{solv}}^*(\text{A}^-) + \Delta G_{\text{solv}}^*(\text{HB}) - \Delta G_{\text{solv}}^*(\text{B}^-) - \Delta G_{\text{solv}}^*(\text{HA}) \quad (5)$$

$$\text{p}K_{\text{a}}(\text{HA}) = \frac{1}{2.303RT} \Delta G_{\text{sol}}^* + \text{p}K_{\text{a}}(\text{HB}) \quad (6)$$

式(6)中, ΔG_{gas}^* 为气相质子解离反应自由能, ΔG_{sol}^* 为溶液相质子解离反应自由能, ΔG_{solv}^* 为溶剂化自由能, 参考态均为 1 mol/L.

1999 年, DaSilva 等^[4]建立了一种新的热力学循环路径, 结果如图 3 所示. 他们在研究中发现, 带电体系在溶剂化能的计算中, 不同相内的结构优化是十分重要的. 对于醇和羧酸类化合物, 除甲醇和甲酸外, 两相中的电子相关影响不大且可以相互补偿, 随着溶质分子的增大, 分子运动所带来的影响也可以忽略. 但对于硫醇化合物, 两相间电子相关的影响可能是其计算值偏离实验值的主要原因. 而对于卤代羧酸, 极化连续介质模型(PCM)^[5~8]的空腔参数不适合可能是其偏离实验值的主要原因.

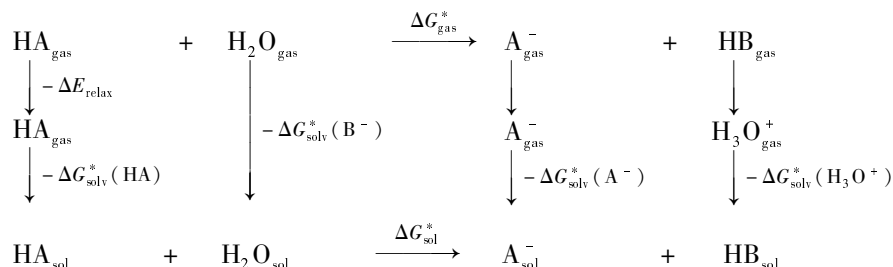


Fig. 3 Thermodynamic cycle 3

除了上述的三种热力学循环路径外, 文献中还曾报道过其他类型的热力学循环路径, 差别主要是路径中的能量采用了不同的方式获得, 如直接引用实验值、间接实验值、理论值、修正值等^[9~11]. 由于这些路径的基本结构与上述三种模型一样, 这里不再详细介绍. 在本文中我们采用了如图 4 所示的热力学循环路径, 并根据公式(8)和(9)计算得到了各化合物的 $\text{p}K_{\text{a}}$ 理论计算值.

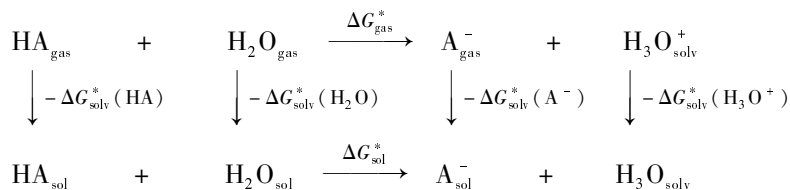


Fig. 4 Thermodynamic cycle used in this calculations



$$\Delta G_{\text{sol}}^* = \Delta G_{\text{gas}}^* + \Delta G_{\text{solv}}^*(\text{A}^-) + \Delta G_{\text{solv}}^*(\text{H}_3\text{O}^+) - \Delta G_{\text{solv}}^*(\text{H}_2\text{O}) - \Delta G_{\text{solv}}^*(\text{HA}) \quad (8)$$

$$\text{p}K_{\text{a}}(\text{HA}) = \frac{1}{2.303RT} \Delta G_{\text{sol}}^* \quad (9)$$

通过大量的计算研究,我们发现采用 DFT 方法的计算精度明显高于 Hartree-Fock 方法,而 MP2 等方法虽然有更好的计算精度,但是耗时更长,计算效率低.在同一方法下,最小的可靠计算基组应为 6-31 + G(d,p),采用更小的基组会导致误差偏大.如果采用更大的基组,计算结果比 6-31 + G(d,p)基组要理想.但是通过对大量数据的分析发现,较大基组与 6-31 + G(d,p)基组之间的误差存在线性关系,也就是说两者之间的误差是可以通过修正来消除的,采用 6-31 + G(d,p)基组计算完全满足需要,并且通过采用修正系数的方法可以轻松的消除计算偏差.最终,我们确定采用 DFT/B3LYP/6-31 + G(d,p)基组作为计算的条件,溶剂化模型采用 PCM(连续极化模型),设置采用软件默认配置.所有计算均在 P4(3.0 GHz, 2 G 内存)计算机上采用 Gaussian 03 软件完成.

2 结果与讨论

2.1 计算结果分析

由于 pK_a 值实验方法的局限性,文献报道的实验值也存在一定的偏差,甚至相同化合物在不同的资料中 pK_a 值相差 1 个单位.同时量子化学计算的方法本身会因为基组选择、构象的选择、计算方法等引入一定的误差,需要对计算所得原始数据进行相关性分析,并建立修正方程.在进行相关性分析时,有个别化合物由于分子中基团的影响,特别是强极性基团、分子内氢键、含有 π 电布云的基团,如硝基、氟、多取代基(特别是三取代)等.由于结构不具有普遍性,这部分计算化合物的数据需要进行单独的修正.图 5 ~ 图 7 为不同类型化合物在水溶液中的 pK_a 值得理论计算值与实验值的线性相关示意图.我们可以看到各种不同类型的化合物理论计算值与实验值的相关系数均大于 0.94,个别类型的相关系数达到了 0.98.同时我们也注意到一个现象,如果增加化合物的数目,会明显提高相关系数.如对胺类化合物进行分析时,开始仅有 24 个数据时相关系数仅达到 0.75 左右,而在最后使用 52 数据时,结果达到了 0.98227.这说明了量子化学从头算的误差是存在的,但不具有确定性,可以通过大量数据的拟合予以消除.这在量子化学从头算的其他应用中也是存在的.

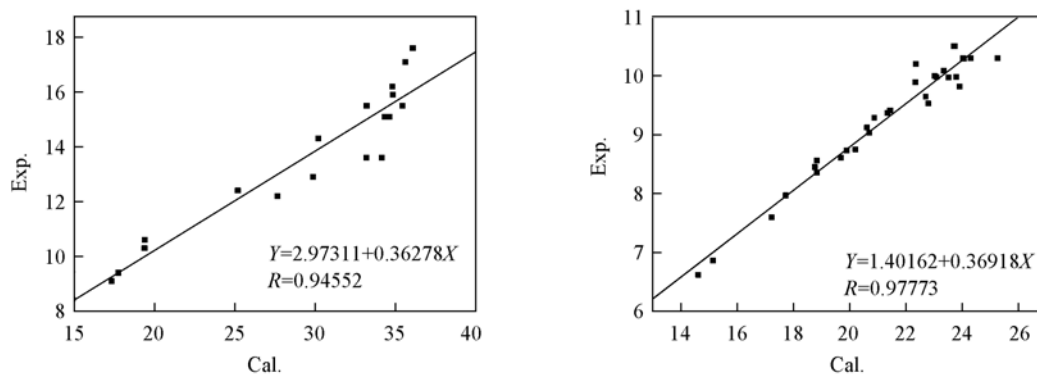


Fig. 5 Correlation between the experimental and theoretical pK_a values for alcohol and hydroxybenzene compounds in H_2O

Table 1 Compared of experimental^[12-18] and theoretical pK_a values, Correlation coefficient(R) and Residue

| Kind of compound | Number | Correlation coefficient | Standard deviation of the fit | Mean absolute error | RMS error |
|------------------|--------|-------------------------|-------------------------------|---------------------|-----------|
| Alcohol | 19 | 0.94 | 0.85 | 0.65 | 0.48 |
| Hydroxybenzene | 33 | 0.98 | 0.21 | 0.16 | 0.14 |
| Carboxyl | 25 | 0.98 | 0.32 | 0.24 | 0.20 |
| Heterocyclic | 36 | 0.96 | 0.58 | 0.40 | 0.40 |
| Aromatic imides | 13 | 0.97 | 0.53 | 0.39 | 0.29 |
| Amines | 52 | 0.98 | 0.64 | 0.45 | 0.43 |

通过对修正后的理论值与实验值的分析发现,醇类中大多数化合物计算误差绝对值在 0.7 个 pK_a 单位左右,分子中含有较强的极性原子或多取代时的误差相对较大,如三氟乙醇误差为 -2.79 个 pK_a 单位,在实际运用时要考虑加入特定基团的修正参数.在醇的计算中我们没有严格区别伯醇、仲醇和叔醇.酚类误差在 0.2 pK_a 单位左右,明显低于醇类的计算误差,这主要是因为芳香环的存在在一定

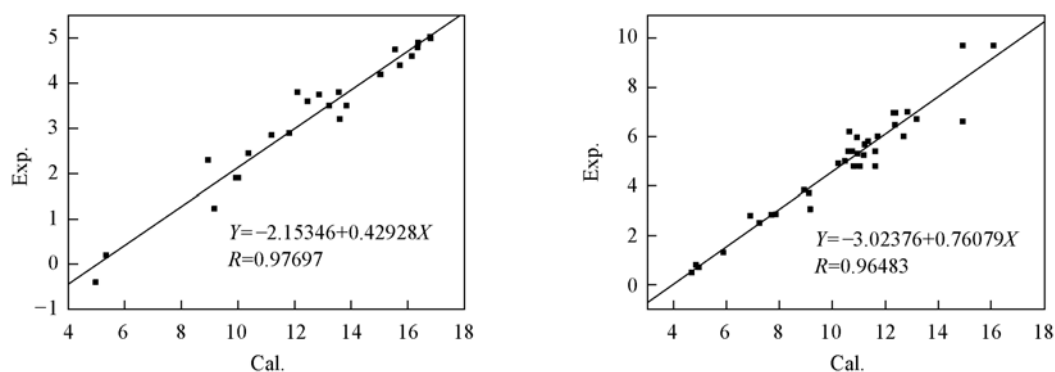


Fig. 6 Correlation between the experimental and theoretical pK_a values for carboxyl and heterocyclic compounds in H_2O

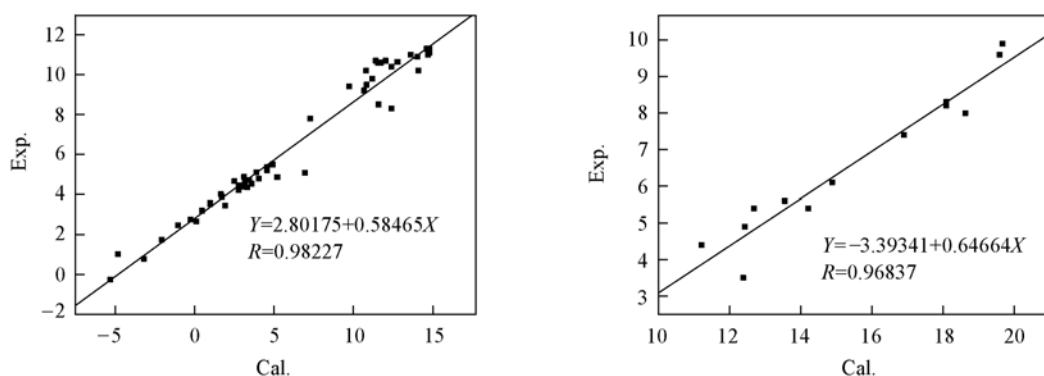


Fig. 7 Correlation between the experimental and theoretical pK_a values for amines and aromatic imides compounds in H_2O

程度上降低了其他基团对羟基解离的影响。羧酸类误差基本小于 0.3 个 pK_a 单位, 个别分子误差达到了 0.6 个 pK_a 单位。芳香氮杂环类平均绝对误差和均方根误差都是 0.4 个 pK_a 单位。该类中个别分子误差达到了 1 个 pK_a 单位, 可能是受分子内氢键形成影响, 如邻二氮杂菲。R— NH_2 分子计算误差在 0.1~0.7 个 pK_a 单位之间。含— NH_2 的脂肪类化合物和芳香类化合物计算误差均小于 1 个 pK_a 单位。通过对大量不同类型的分子在水溶液中 pK_a 值的计算, 可以看到计算结果的精度和准确度都能够令人满意, 能够准确的体现出不同分子间的差异。

2.2 应用实例

开展 pK_a 值的计算研究, 对于有机合成路线设计和优化具有重要的现实意义, 它可以从原子的角度来理解分子反应活性、反应机理、反应动力学等问题, 并帮助我们优化反应路线, 对合成路线设计提供理论指导。

2.2.1 去甲肾上腺素在水溶液中 pK_a 值的计算 结合前面的计算方法, 我们对去甲肾上腺素在水溶液中的 pK_a 值进行了计算。由于去甲肾上腺素中有多个可旋转的单键存在, 我们先对该分子进行了能量扫描, 选择了能量最低的三种构象进行了 pK_a 值计算研究。然后根据玻尔兹曼分布, 计算出了去甲肾上腺素— NH_2 和苯环上— OH 的 pK_a 值, 数据如表 2 所示。同时, 文献[19]报道了针对该化合物的计算结果, 其误差为 0.46 个 pK_a 单位。我们详细研究了其采用的计算方法, 从原理到去甲肾上腺素的构象分析我们基本是一样的, 唯一的不同是他们在计算能量过程中, 对每一步都进行了误差修正。而我们在计算过程中允许了软件计算误差的存在, 在完成所有计算后统一修正。从表 2 中的数据也可以看出, 最低能量构象的 pK_a 值与玻尔兹曼统计值十分接近, 如果能够在计算初期寻找到合适的低能量构象, 其他构象可以不用进行计算, 这样可以在不明显增大误差的情况下最大限度减少计算量。

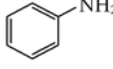
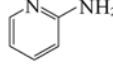
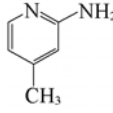
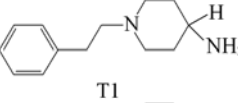
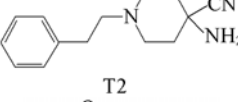
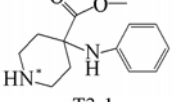
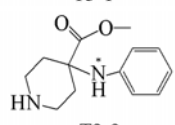
2.2.2 不同分子结构对— NH_2 反应活性的影响 如表 3 中所示的各种化合物, 我们分别计算了它们在水溶液中 pK_a 值, 可以发现 pK_a 值可以很准确的体现不同的 NH —基团在不同的化学环境下反应活性

Table 2 Experimental and theoretical pK_a values for norepinephrine

| Conformation analysis | Predictive values | Energy | Boltzmann average value | Reported value | Experimental value |
|-------------------------------|-------------------|-------------|-------------------------|----------------|--------------------|
| Conformer-1 —NH ₂ | 9.88 | -592.231236 | 9.87 | 10.16 | 9.70 |
| DWConformer-2—NH ₂ | 9.88 | -592.22848 | | | |
| Conformer-3—NH ₂ | 9.40 | -592.227378 | | | |
| Conformer-1—OH | 9.08 | -591.783234 | 9.08 | — | 8.64 |
| Conformer-2—OH | 9.27 | -591.780471 | | | |
| Conformer-3—OH | 8.96 | -591.781184 | | | |

有明显区别. 如苯胺可以与羰基等亲电试剂发生反应, 而吡啶胺或者 4-位甲基取代的吡啶胺则较难发生反应, pK_a 的计算值很准确显示出了这些分子间反应活性的差别. 化合物 T1 和 T2 中的两个氨基受邻位基团的影响, T1 和容易和其他亲电试剂反应, 而 T2 反应则困难很多. 在实验中我们发现, T1 可以很容易的与 2-氯吡啶或 2-氯-4-甲基吡啶发生取代反应, 这反过来也验证了吡啶胺与甲基吡啶胺反应活性要低于苯胺的事实. 同时再如化合物 T3, 采用 EDC 缩合剂, 碱性条件下, 哌啶环上的 NH 容易与羧酸发生缩合反应, 而侧链上的 NH 几乎不与羧酸发生反应.

Table 3 Theoretical prediction of pK_a values and reactivity of different compounds

| No. | Structure of compound | Predictive pK_a value | Predictive reaction activity | Reality of reaction activity |
|-----|---|-------------------------|------------------------------|------------------------------|
| 1 |  | 4.63 | Active | Easy response |
| 2 |  | 1.29 | No-active | Almost no react |
| 3 |  | 1.46 | No-active | Little react |
| 4 |  | 9.82 | Active | Easy response |
| 5 |  | 4.46 | Less active | Reactively slow |
| 6 |  | 10.26 | Active | Easy response |
| 7 |  | 6.50 | No-active | No react |

3 结 论

随着理论研究和计算机技术的发展, 量子化学从头算的方法目前正逐渐成为我们研究药物理化性质的有力工具, 并且随着其计算精度、准确度的日益提高, 也正逐渐成为我们进行药物合成研究、药物生物活性研究、计算机辅助药物分子设计的有力工具. 本文通过建立 pK_a 值计算方法和不同类型化合物的修正方法, 可以对感兴趣的药物分子进行 pK_a 值计算与预测. 同时通过文中的应用实例, 可以将 pK_a 计算值用于解释实验现象, 设计、优化实验路线. 当然, 我们也清醒地认识到了存在的问题, 随着分子结构的复杂化, 计算的难度会成几何级数的增长. 例如去甲肾上腺素这样的结构中存在多个可旋转单键的化合物, 要想获得准确的结果必须首先找到分子的最低能量构象, 这就要求计算前对所有稳定的分子构象进行分析, 并以此计算其 pK_a 值, 这无疑又增加了计算的难度. 通过建立药物分子气

相、液相条件下解离能的高精度计算及计算结果误差的修正方法,为下一步开展有机化学反应、药物与受体相互作用、主客体作用等涉及自由能计算的研究提供技术支持,为虚拟组合化学和虚拟组合筛选提供更精确的结构参数.

参 考 文 献

- [1] Liu Y. H., Wang Y.. J. Progress in Chemistry[J], 2008, **20**(9): 1241—1250
- [2] Hong W., Johan U.. Expert Opin. Drug Metab. Toxicol. [J], 2006, **2**(1): 139—155
- [3] Klicic J. J., Friesner R. A.. J. Phys. Chem. [J], 2002, **106**: 1327—1335
- [4] DaSilva C. O., Dasilva E. C.. J. Phys. Chem. A. [J], 1999, **102**: 11194—11199
- [5] Cossi M., Barone V., Cammi R.. Chem. Phys. Lett. [J], 1996, **255**: 327—335
- [6] Cancès M. T., Mennucci B.. J. Phys. Chem. [J], 1997, **107**: 3032—3041
- [7] Barone V., Cossi M.. J. Phys. Chem. [J], 1997, **107**: 3210—3221
- [8] Cossi M., Barone V., Mennucci B.. Chem. Phys. Lett. [J], 1998, **286**: 253—260
- [9] Toth A. M., Liptak M. D., Phillips D. L.. J. Chem. Phys. [J], 2001, **114**(10): 4595—4606
- [10] Liptak M. D., Shields G. C.. J. Quan. Chem. [J], 2001, **85**: 727—741
- [11] Matlew D. L., George C.. J. Am. Chem. Soc. [J], 2001, **123**: 7314—7319
- [12] Perrin D. D.. Dissociation Constants of Organic Bases in Aqueous Solution[M], London: Butterworths, 1972
- [13] Serjeant E. P., Dempsey B.. Ionization Constants of Organic Acids in Aqueous Solution[M], Oxford: Pergamon, 1979
- [14] Albert A. R.. Physical Methods in Heterocyclic Chemistry[M], New York, Academic Press: 1963
- [15] Sober H. A.. CRC Handbook of Biochemistry[M], Cleveland, Ohio, CRC Press, 1968
- [16] Perrin D. D., Dempsey B.. pK_a Prediction for Organic Acids and Bases[M], London: Chapman & Hall, 1981
- [17] Albert A. R., Serjeant E. P.. The Determination of Ionization Constants[M], London: Chapman & Hall, 1984
- [18] Budavari S.. The Merck Index[M], New York: Merck, 1996
- [19] Lu H. T., Chen X., Zhan C. G.. Journal of Central China Normal University[J], 2004, **38**: 205—206

Quantum Chemical Calculations on pK_a Values of Organic Compounds in Aqueous Solution

DING Xiao-Qin*, SUN Yang, DING Jun-Jie, CHEN Ji-Sheng
(Beijing Institute of Pharmaceutical Chemistry, Beijing 102205, China)

Abstract Quantum chemical calculations of the pK_a values were performed using DFT method with 6-31 + g (d,p) basis set for more than 200 compounds of alcohol, phenol, carboxylic acid, aliphatic amines, aromatic amine and nitrogen heterocyclics etc. The solution-phase calculations use polarized continuum solution(PCM) model with default set in water. The residue of most compounds is less than 1 pK_a unit. Correlation coefficient between experiment results and forecast results is better than 0.94. The results can help us to research reaction activity difference between different molecules or different groups in one molecule. Quantum Chemical calculations of the pK_a values will be instructive to computer aided molecular design and organic synthetic route design and optimization.

Keywords pK_a value; Density functional theory, Polarized continuum solution(PCM)