

УДК 615.015.11

## ТЕОРЕТИЧЕСКИЕ МЕТОДЫ ОЦЕНКИ ЛИПОФИЛЬНОСТИ ОРГАНИЧЕСКИХ МОЛЕКУЛ

© 2002 В.В.Иванов, Л.П.Логинова, А.С.Мирная, О.С.Чернышева

В работе рассматривается методология теоретического расчета липофильности на примере некоторых модельных молекулярных систем. Рассматриваются эмпирический и квантовохимический подходы. Эмпирический подход основывается на расчете некоторых топологических инвариантов графа молекулы. Квантовохимический метод базируется на соответствующем расчете энергии сольватации в рамках поляризационно-континуальной модели. Тестовые результаты для π-сопряженных систем (метод ППП) и малых молекул (*ab initio*) демонстрируют хорошее согласие с экспериментом.

Логарифм коэффициента распределения вещества в системе октанол-вода (липофильность,  $\text{ЛФ}$ ,  $\lg P$ ) является фундаментальным параметром, определяющим транспортные возможности и биологическую активность молекулярных систем [1,2]. Современная ориентация на "автоматизированный" органический синтез с целью выявления перспективных биоактивных веществ приводит к необходимости просеивания миллионов новых соединений по различным критериям, в том числе и  $\text{ЛФ}$ . Важной разновидностью характеристик  $\text{ЛФ}$  являются также константы распределения вещества в системе вода-мицеллярная псевдофаза [3]. Дифильные мицеллы поверхностно-активных веществ, образующиеся путем самоассоциации, считаются лучшей моделью биологических микроагрегатов, например, липидов.

В связи с этим существует проблема теоретических оценок  $\text{ЛФ}$  и соответствующей экспериментальной проверки таких оценок. В настоящей работе рассматривается методология расчетов  $\text{ЛФ}$  на примере углеводородов и некоторых алкалоидов.

### Эмпирический подход к оценке липофильности

Эмпирический подход к вычислению  $\text{ЛФ}$  заключается в разработке неких калибровочных параметров, которые были выверены по известным экспериментальным данным. Среди таких моделей особую популярность приобрели различные *аддитивные схемы* типа модели Риккера [2,4]. При всей ограниченности аддитивного метода он, по-видимому, является наиболее удобным для грубой оценки  $\text{ЛФ}$  новых соединений. В этом методе предполагается, что структура молекулы каким-то образом разбита на определенные (атомные или связевые) фрагменты, количество которых и определяет  $\text{ЛФ}$  молекулы. Иными словами, если молекула разбита на  $N$  фрагментов и каждый  $i$ -ый фрагмент входит в молекулу под номером  $j$   $n_{ij}$  раз, то суммарная липофильность молекулы ( $\lg P_j$ ) записывается так:

$$\lg P_j = n_{1j}\chi_1 + n_{2j}\chi_2 + \dots + n_{Nj}\chi_N, \quad (1)$$

Коэффициенты  $\chi_i$  могут быть найдены методом наименьших квадратов. В принципе возможны множество способов разбиения молекулы на структурные составляющие, что определяет многообразие популярных в настоящее время аддитивных схем. Некоторые теоретические вопросы связанные с выбором оптимального разбиения молекул на структурные фрагменты описаны в [5].

Очевидным недостатком аддитивного метода является неучет взаимодействия между структурными фрагментами, что иногда приводит к довольно большим ошибкам, особенно при расчетах сопряженных систем. Недостаток этот, впрочем, частично устраняется в более изощренных аддитивных подходах (см., например [4]).

Метод молекулярных индексов в некоторой степени учитывает эффекты неаддитивности, поскольку оперирует величинами, характеризующими структуру молекулы в целом. В качестве таких величин можно использовать самые различные индексы (дескрипторы) [6]. Среди них особой притягательностью, как нам кажется, обладают топологические инва-

рианты – индексы, которые описывают особенности структурной формулы молекулы – разветвленность, порядок связывания атомов и т.д. К таким индексам относятся широко известные индексы Рандича и теоретико-информационный индекс. Индекс Рандича ( $R$ ) записывается следующим образом:

$$R = \sum_{(i,j)} (v_i v_j)^{\frac{1}{2}}, \quad (2)$$

где  $v_i$  и  $v_j$  – степени связанных вершин молекулярного графа, суммирование идет по всем парам связанных атомов. Теоретико-информационный индекс ( $I_D$ ) базируется на известной формуле Шеннона:

$$I_D = -\sum_i \frac{n_i}{n} \log_2 \frac{n_i}{n}, \quad n = \sum_i n_i. \quad (3)$$

Известно множество трактовок этой формулы [7]. В этой работе мы предполагаем, что  $n_i$  – количество путей в молекулярном графе, имеющих длину “ $i$ ”, а  $n$  – их сумма. Таким образом  $I_D$  характеризует неоднородность в распределении маршрутов графа. С использованием этих величин возможно построение корреляционных функций описывающих зависимость ЛФ от структурных индексов. В настоящей работе в качестве примера мы вычислили функцию ЛФ от индексов  $R$ ,  $I_D$  и числа углеродных атомов ( $N_c$ ) в молекуле насыщенного углеводорода:

$$\lg P_{(top)} = 1.035 + 0.477 N_c - 0.375 R + 0.757 I_D,$$

$\sigma=0.074$ , к.к.=0.986.

Результаты расчета представлены в таблице 1. Как видим указанные параметры достаточно чувствительны для описания разветвленной структуры углеводородов. Значения  $\lg P_{(top)}$ , в среднем, достаточно близко к экспериментальным величинам.

#### Квантовохимический подход к оценке липофильностей

Прежде всего отметим, что, как следует из равновесной термодинамики, константа распределения молекул вещества между двумя несмешивающимися фазами может быть выражена через изменение стандартной энергии Гиббса  $\Delta G$ , соответствующее переходу от одного растворителя к другому. Эта величина в свою очередь выражается через энергии сольватации при растворении вещества в нормальном октаноле ( $\Delta G_o$ ) и воде ( $\Delta G_e$ ). Таким образом, мы имеем дело со следующим соотношением:

$$\lg P = -\frac{\Delta G_o - \Delta G_e}{2.3RT}, \quad (4)$$

где  $T$  – температура, а  $R$  – универсальная газовая постоянная.

**Таблица 1.** Топологические индексы и ЛФ насыщенных углеводородов. ( $\lg P_{(top)}$  – теоретические значения ЛФ, вычисленные с использованием топологических индексов;  $\lg P_{(exp)}$  – экспериментальная величина)

Углеводород (брutto формула)	$I_D$	$R$	$\lg P_{(top)}$	$\lg P_{(exp)}$
$C_2H_6$	0	1	1.61	1.81
$C_5H_8$	0.637	1.414	2.42	2.36
$C_4H_{10}$	1.011	1.914	2.99	2.89
$CH(CH_3)_3$	0.693	1.732	2.81	2.76
$C_5H_{12}$	1.280	2.414	3.48	3.39
$(CH_3)_2CH-C_2H_5$	1.055	1.693	3.58	3.77
$C(CH_3)_4$	0.673	2.000	3.18	3.11
$C_6H_{14}$	1.490	2.914	3.93	3.9
$(CH_3)_3C-C_2H_5$	1.044	2.354	3.80	3.82
$(CH_3)_2CH-CH(CH_3)_2$	1.085	2.643	3.73	3.42
$C_7H_{16}$	1.662	3.414	4.55	4.66
$C_8H_{18}$	1.809	3.914	4.75	5.18
$C_{10}H_{22}$	2.05	4.914	5.52	5.01
$C_{12}H_{26}$	2.243	5.914	6.24	6.1
$C_{14}H_{30}$	2.405	6.614	6.94	7.2
$(CH_3)_3C-C(CH_3)_3$	1.075	3.25	4.45	3.96
$(CH_3)_3C-CH_2-C(CH_3)_3$	1.347	3.71	4.96	4.55

Следовательно, задача вычисления липофильности сводится к теоретическому определению свободной энергии сольватации в октаноле и в воде. В настоящее время уровень развития вычислительной техники и квантовой химии позволяет провести довольно точные расчеты молекул в вакууме. Что же касается расчетов сольватационных эффектов и межмолекулярного взаимодействия вообще, то это все еще достаточно сложная проблема. Большая трудоемкость прямого вычисления взаимодействия растворителя и растворенного вещества привела к интенсивной разработке так называемого континуального метода (см. например [8]). В этом методе вся сложность статистического поведения молекул растворителя заменяется диэлектрическими параметрами непрерывной среды. В целом эта модель, получившая название поляризационно-континуальной модели (в английской аббревиатуре PCM), предполагает, что молекула растворенного вещества погружена в полость соответствующей формы и размеров, которая как бы “вырезана” в поляризуемой диэлектрической среде с определенной проницаемостью ( $\epsilon$ ). Форма полости может варьироваться от простой сферической (или эллипсоидальной) до тонко учитывающей геометрические особенности молекулы. На рисунке 1, схематически показана полость в растворителе, в которую погружен углеводород.

Взаимодействие растворенного вещества с растворителем (рассматривается только неспецифическая сольватация) описывается в терминах реактивного поля. Это поле возникает в результате поляризации растворителя молекулой растворенного вещества. Реактивное поле, в свою очередь, взаимодействуя с молекулой растворенного вещества, приводит к смещению ее энергетических уровней, которое и определяет величину энергии сольватации. При квантовохимическом расчете полная энергия системы ( $E$ ) и волновая функция  $|\Psi\rangle$  соответствующие данной электронной модели, являются решением уравнения Шредингера:

$$(\hat{H}_0 + \hat{R})|\Psi\rangle = E|\Psi\rangle, \quad (5)$$

где  $\hat{H}_0$  – многоэлектронный гамильтониан системы в вакууме, а  $\hat{R}$  – поправка, связанная с реактивным полем. В настоящее время разработано множество прикладных квантохимических пакетов, реализующих всевозможные варианты PCM. В нашей работе мы использовали программы на базе комплексов GAMESS и DALTON<sup>1</sup>.

Конкретные расчеты липофильности мы начинаем с использования  $\pi$ -электронного приближения (гамильтониан ППП) в рамках метода самосогласованного поля Хартри-Фока. Этот метод хорошо зарекомендовал себя при расчетах сопряженных систем. Для характеристики растворителя использовались следующие значения диэлектрических проницаемостей:

$$\text{вода: } \epsilon_{(0)} = 80, \epsilon_{(\infty)} = 5.2;$$

$$\text{октанол-1: } \epsilon_{(0)} = 9.8, \epsilon_{(\infty)} = 3.0.$$

Несмотря на то, что ППП – схема явно очень грубая для расчета столь тонких эффектов, как взаимодействие с растворителем, общие результаты оказались вполне удовлетворительными. Расчет некоторых сопряженных систем и сравнение с экспериментом представлены в таблице 2. Как видим, теоретические значения ЛФ для бензола и нафталина оказались достаточно точными и в полуэмпирической модели. Точность расчета для  $\pi$ -систем с гетероатомами (анилин, пиррол) также удивительно стабильна – ошибка не превышает 0.11. Далее в таблице 2 приведены данные для алкалоида теофиллина (1,3-диметилксантин) [9]. Эта система известна как эффективное бронхорасширяющее средство. Рассчитанное значение  $\lg P = 0.2$  здесь уже заметно отличается от экспериментальной величины  $\lg P = -0.02$  (вплоть до знака!), что, по-видимому, связано с необходимостью учитывать сигма-электронную ( $\text{CH}_3$  группы) подсистему, что совершенно игнорируется в

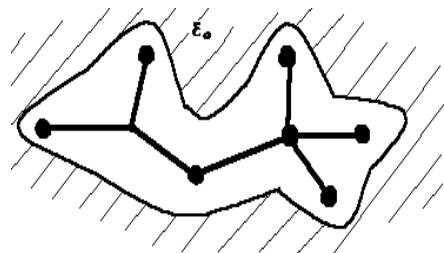


Рис.1. Молекула углеводорода, погруженная в растворитель

<sup>1</sup> Подробности нашей расчетной схемы будут описаны в дальнейшем.

$\pi$ -электронном приближении. Такая же ситуация для антипирина (1-фенил-2,3-диметилпиразолон-5, анальгезирующее, противовоспалительное средство) ошибка – 0.42. Молекулы кофеин и урацил неплохо описываются нашей моделью. В целом данные расчета говорят о том, что в системах не содержащих значительные алкильные радикалы  $\pi$ -электронный квантовохимический расчет ЛФ как правило оказывается вполне удовлетворительным.

Неэмпирические (*ab-initio*) оценки  $\lg P$  мы проводили в валентно-расщепленном двухэкспоненциальном базисе Даннинга (DZV). Расчет показал хорошее согласие с экспериментальными данными (табл.3). Даже в достаточно сложных системах (ацетилен, этилен, этан) точность все еще остается высокой. Наихудший результат среди углеводородов мы получили для молекулы этана. Замечательно также, что для молекулы метанола теоретический метод верно воспроизводит величину и ее отрицательный знак, характеризующий гидрофильность спирта.

На данном этапе мы ограничились лишь методом Хартри-Фока. Открытым остался вопрос о роли эффектов электронной корреляции в проблеме ЛФ, а также влияние выбора базисного набора. Тем не менее полученные данные говорят о том, что методы квантовой химии в настоящее время позволяют оценивать даже такие достаточно тонкие характеристики, взаимодействия со средой, как ЛФ.

**Таблица 2.** Расчетные и экспериментальные<sup>2</sup> значения липофильностей, вычисленные в поляризационно-континуальной модели ППП для некоторых сопряженных систем ( $T = 298\text{K}$ )

Молекула	$\lg P_{(\text{theor})}$	$\lg P_{(\text{exp})}$	Ошибка
Бензол	2.27	2.15	0.12
Нафталин	3.48	3.37	0.11
Анилин	1.01	0.9	0.11
Пиррол	0.86	0.75	0.11
Теофиллин 	0.2	-0.02	0.22
Антипирин 	0.65	0.23	0.42
Кофеин 	-0.03	-0.06	0.03
Урацил 	-1.8	~-2.0	0.2

**Таблица 3.** Липофильности малых молекул (*ab-initio*, метод Хартри-Фока, базис DZV,  $T = 298\text{K}$ )

Молекула	$\lg P_{(\text{theor})}$	$\lg P_{(\text{exp})}$	Ошибка
метан $\text{CH}_4$	1.12	1.09	0.03
ацетилен $\text{C}_2\text{H}_2$	0.47	0.37	0.1
этилен $\text{C}_2\text{H}_4$	1.23	1.13	0.1
этан $\text{C}_2\text{H}_6$	1.56	1.81	-0.25
метиламин $\text{CH}_3\text{NH}_2$	-0.56	-0.57	-0.01
метанол $\text{CH}_3\text{OH}$	-0.79	-0.77	0.02

<sup>2</sup> Экспериментальные данные для кофеина и урацила получены в настоящей работе, остальные величины из работы [9].

### **Выводы**

Рассмотрены различные подходы к оценке ЛФ органических молекул. На примере насыщенных углеводородов построена регрессионная модель, связывающая липофильность с некоторыми топологическими инвариантами.

Предложено использование методов квантовой химии для описания ЛФ в рамках РСМ модели. Показано, что даже в  $\pi$ -электронном приближении ППП можно получить полу-количественное описание  $\lg P$ . Соответствующие неэмпирические расчеты некоторых малых молекул показали хорошее согласие с экспериментом.

### **Литература**

1. Hansch C., Leo. A. Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley Interscience, NY. 1979. 339 р.
2. Leo A.J. Chemical Reviews. 1993. V.93. No.4. P.1281-1306.
3. Логинова Л.П., М?к. симпоз?ум, “Сучасн? проблеми ф?зично? х?м?? Матеріали сим-поз?ума”. 2002. С.98.
4. Rekker R.F. Eur. J. Med. Chem. – Chimica Therapeutica. 1979. V.14. №6. P.479-488.
5. Степанов Н.Ф., Ерлыкина М.Е., Филиппов Г.Г., Методы линейной алгебры в физической химии, из-во Московского университета. 1976. 360 с.
6. Раевский О.А. Успехи химии. Т.68. №6. 1999. С.555-575.
7. Химические приложения топологии и теории графов. М.: Мир, 1987. 387 с.
8. Constanciel R., Tapia O. Theor. Chim. Acta. 1978. №48. P.75-85.
9. Эпштейн Н.А., Нижний С.В. Хим. Фарм. Журн. 1978. С.57-61.

*Поступила в редакцию 2 декабря 2002 г.*

Kharkov University Bulletin. 2002. №573. Chemical Series. Issue 9(32). V.V.Ivanov, L.P.Loginova, A.S.Mirnaya, O.S.Chernisheva. Theoretical methods for estimating the lipophilicity of organic molecules.

The lipophilicity of several model molecular systems were estimated by using both empirical and quantum chemistry methods. The empirical method has been based on the calculation of some topological invariants, which represent the graphs of the molecules. The quantum chemistry method is based on the calculation of solvation energy within polarizable continuum model. Numerical results for  $\pi$ -conjugated systems (PPP approach) and for small molecules (*ab initio*) demonstrate good agreement with experimental data.