

УДК 547.21

МЕТОД ОБРАТНЫХ ФУНКЦИЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ СТРУКТУРА–СВОЙСТВО

© 2007 г. Е. А. Смоленский, А. Н. Рыжов, В. М. Бавыкин,
И. В. Чуваева, член-корреспондент РАН А. Л. Лапидус

Поступило 07.03.2007 г.

Нахождение зависимостей между физико-химическими свойствами органических соединений и структурой их молекул является одной из важнейших задач теоретической химии, а построенные в результате аппроксимирующие функции весьма актуальны для предсказания свойств неисследованных и еще несинтезированных соединений. Тем не менее для некоторых свойств, например, для поверхностного натяжения и, в особенности, для температур кристаллизации до сих пор не удалось получить сколько-нибудь удовлетворительного решения этой задачи [1, 2], хотя для ряда других свойств, например, для молекулярной рефракции, плотности, энтальпии образования и т.п. получены вполне удовлетворительные по точности приближенные функции [1, 3, 4]. На качественном уровне можно рассматривать эти свойства как более простые и более сложные. В работе одного из авторов [2] введены количественные оценки для простоты и сложности физико-химических свойств, иллюстрирующие эти представления.

Рассматривая физико-химические свойства нормальных алканов C_nH_{2n+2} , можно легко заметить (это отмечалось еще в работе [5]), что для простых свойств, начиная с определенного n_0 (различного для разных свойств), зависимость свойства от числа атомов углерода n становится линейной. Именно такие свойства принято называть аддитивными. Это давно известное положение широко применяется для описания подобных свойств. Однако удивительно то, что до сих пор никто не обращал внимания на такие очевидные факты (и тем более не пытался использовать их в расчетах), состоящие в том, что многие простые свойства фактически можно выразить через другие свойства, которые формально достаточно сложны. Так, молекулярная рефракция R – одно из самых простых свойств [4] – является функци-

ей от коэффициента преломления n_D (формула Лоренца):

$$R = \frac{n_D^2 - 1}{n_D^2 + 2} \cdot \frac{M}{d_{20}^4} = \frac{4}{3} \pi N_A \alpha, \quad (1)$$

где M – молекулярная масса вещества, d_{20}^4 – плотность вещества, N_A – постоянная Авогадро, α – электронная поляризуемость. Еще более важное и известное физико-химическое свойство – плотность d_{20}^4 , связанная гиперболической зависимостью с молекулярным объемом V_M при температуре 20°C:

$$d_{20}^4 = \frac{M}{V_M}. \quad (2)$$

В действительности же плотность всегда рассчитывается именно через молекулярный объем.

Эти и некоторые другие факты позволяют сделать очень серьезные выводы и обобщить их для разработки принципиально нового подхода к расчету произвольных физико-химических свойств. Именно для всех физико-химических свойств (а не только для плотности и коэффициента преломления) можно построить как функции новые формальные свойства, которые, во-первых, связаны с исходными взаимнооднозначно, а во-вторых, линейно зависят для n -алканов от n , что позволяет надеяться на значительное упрощение решения исходной задачи, поскольку они по своему построению уже будут аддитивными. Это упрощение состоит в том, что можно найти аппроксимирующие функции для новых формальных (обобщенных) свойств, а затем получить рассчитанные значения для исходного свойства, поскольку они связаны взаимнооднозначной зависимостью.

В настоящей работе в общем виде предлагается простой математический метод для решения рассмотренной проблемы, а затем иллюстрируются способы применения и очевидные

преимущества нового метода на конкретном примере расчетов октановых чисел алканов и циклоалканов.

Схема зависимости некоего гипотетического экспериментально измеренного свойства P от n для n -алканов C_nH_{2n+2} представлена на рис. 1. Квадраты для $n = 1, 2, \dots$ соответствуют экспериментальным значениям свойства P_n , а линия $f(n)$ соответствует некой функции, для которой вы-

полняется условие $f(n) = P_n$ (обозначены треугольниками).

Таких функций, для которых имеет место последнее соотношение, вообще говоря, может быть очень много, но мы для практического использования выберем такую, которая была бы конечной комбинацией элементарных функций и максимально проста (например, имела бы минимальное число точек перегиба). Например, в качестве такой функции можно всегда построить известный полином Лагранжа:

$$L_n(x) = \sum_{i=0}^n \frac{(x-x_0)\dots(x-x_{i-1})(x-x_{i+1})\dots(x-x_n)}{(x_i-x_0)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_n)} y_i, \quad (3)$$

хотя это весьма сложная и неудобная для практического использования функция. Более удобными в этом смысле являются, например, полиномы Чебышева [6].

Впрочем, даже если такую функцию в классе суперпозиций элементарных функций трудно или невозможно найти, то можно взять такую, чтобы выполнялось условие:

$$\sum_n [P_n - f(n)]^2 \leq \varepsilon, \quad (4)$$

где ε – достаточно малое положительное число. В этом случае экспериментальные значения P_{exp} не обязательно совпадают с рассчитанными по

функции $f(x)$ значениями (показано на рис. 1), которые мы обозначим:

$$P_{\text{calc}} = f(n), \quad (5)$$

при этом неравенство (4) может быть записано в виде:

$$\sum_n [P_{\text{exp}} - P_{\text{calc}}]^2 \leq \varepsilon. \quad (6)$$

Итак, пусть мы нашли функцию $f(x)$, для которой выполняется условие (5), причем рассчитанные значения свойства P отличаются от экспериментальных сколь угодно мало, в соответствии с неравенством (6).

Сделаем небольшое отступление и вспомним элементарные свойства обратных функций: если имеем $y = f(x)$, то для обратной функции $f^{-1}(x)$ получим $f^{-1}(y) = f^{-1}[f(x)] = x$. В качестве примеров обратных функций можно привести такие пары функций, как экспонента и логарифм, парабола n -ой степени и соответствующий корень, тангенс и арктангенс и т.д.

Рассмотрим обратную в математическом смысле функцию

$$x = f^{-1}(P), \quad (7)$$

для которой будет точно выполняться условие:

$$n = f^{-1}(P_{\text{calc}}). \quad (8)$$

Умножив обе части уравнения (7) на число a и прибавив к правой и левой части число b , получим:

$$ax + b = af^{-1}(P) + b. \quad (9)$$

Обозначив правую часть уравнения (9) как новую функцию φ , будем иметь:

$$af^{-1}(P) + b = \varphi(P) = ax + b, \quad (10)$$

что для целых значений $x = n$ означает:

$$\varphi(P_{\text{calc}}) = ab + b. \quad (11)$$

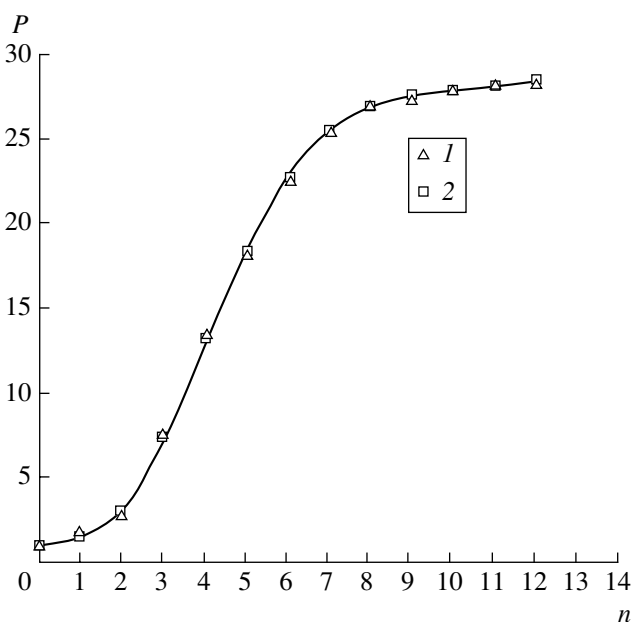


Рис. 1. Зависимость функции P от числа атомов углерода в n -алканах: 1 – экспериментальные данные, 2 – результат моделирования.

Приближенно можно полагать, используя (6), что

$$\varphi(P_{\text{exp}}) = an + b. \quad (12)$$

Теперь определим новое свойство P_1 , в некотором смысле являющееся производным от свойства P :

$$P^1 = \varphi(P) = af^{-1}(P_{\text{exp}}) + b, \quad (13)$$

т.е. каждому значению свойства P для произвольной молекулы, а не только для n -алканов, соответствует значение свойства P^1 , вычисляемое по формуле (10). При этом очень важно подчеркнуть, что введенное новое свойство будет уже строго аддитивным [5], выражающимся для n -алканов линейной функцией от числа атомов углерода согласно формуле (12).

Для нахождения аппроксимирующей функции этого нового свойства P^1 можно использовать способы нахождения оптимальных топологических индексов (ОТИ), описанные в работах [4, 7]. Применение предложенного метода продемонстрируем на примере расчета октановых чисел алканов и циклоалканов. Для упрощения в качестве переменной будем использовать $n_1 = n - 1$ (число связей C–C в алкане) вместо числа n (число атомов C в алкане). На рис. 2 отчетливо видно, что эта зависимость хорошо аппроксимируется одной из ветвей гиперболы, которая однозначно определена параметрами: координатами точки пересечения своих асимптот n_0, P_0 , углом между осью абсцисс и действительной осью гиперболы α (здесь и далее угол отсчитывается от действительной оси гиперболы), длинами действительной осью $2a$ и мнимой $2b$.

Параметры уравнения второго порядка

$$An_1^2 + BnP_{\text{calc}} + CP_{\text{calc}}^2 + Dn + EP_{\text{calc}} + F = 0, \quad (14)$$

$$P_{\text{calc}} = \frac{-Bn_1 - E - \sqrt{(Bn_1 + E)^2 - 4C(An_1^2 + Dn_1 + F)}}{2C}, \quad (21)$$

$$n_1 = \frac{-BP_{\text{calc}} - D - \sqrt{(BP_{\text{calc}} + D)^2 - 4A(CP_{\text{calc}}^2 + EP_{\text{calc}} + F)}}{2A}. \quad (22)$$

Преобразуем (21) таким образом, чтобы левая часть уравнения линейно зависела от n_1 и не зависела от P_{calc} . Получим выражение:

$$2An_1 + D = -BP_{\text{calc}} - \sqrt{(BP_{\text{calc}} + D)^2 - 4A(CP_{\text{calc}}^2 + EP_{\text{calc}} + F)}. \quad (23)$$

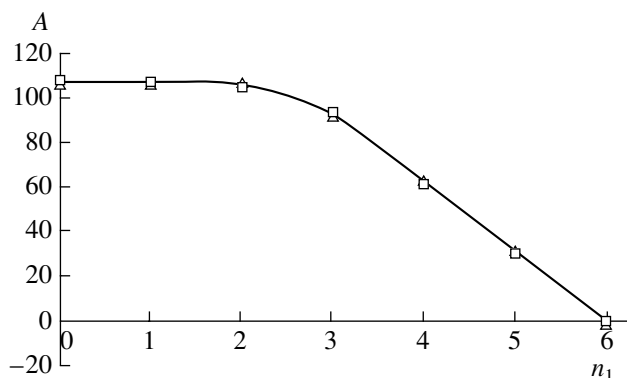


Рис. 2. Зависимость октанового числа от числа атомов углерода в n -алканах (обозначения см. рис. 1).

описывающего искомую гиперболу в общем случае, могут быть вычислены, исходя из a, b, n_0, P_0, α по формулам (15–20) с точностью до произвольного ненулевого множителя M :

$$A = M \left(\frac{\cos^2 \alpha}{a^2} - \frac{\sin^2 \alpha}{b^2} \right), \quad (15)$$

$$B = -2M \left(\frac{1}{a^2} + \frac{1}{b^2} \right) \sin \alpha \cos \alpha, \quad (16)$$

$$C = M \left(\frac{\sin^2 \alpha}{a^2} - \frac{\cos^2 \alpha}{b^2} \right), \quad (17)$$

$$D = -(BP_0 + 2An_0), \quad (18)$$

$$E = -(Bn_0 + 2CP_0), \quad (19)$$

$$F = An_0^2 + Bn_0P_0 + CP_0^2 - M, \quad (20)$$

где P_{calc} – вычисленное значение октанового числа алкана C_nH_{2n+2} .

Искомая ветвь гиперболы описывается формулами (21, 22):

Возвращаясь к n (используя замену $n_1 = n - 1$), получим:

$$\begin{aligned} 2A(n-1) + D &= 2An + D - 2A = an + b, \\ -BP_{\text{calc}} - \sqrt{(BP_{\text{calc}} + D)^2 - 4A(CP_{\text{calc}}^2 + EP_{\text{calc}} + F)} &= \\ &= an + b. \end{aligned} \quad (24)$$

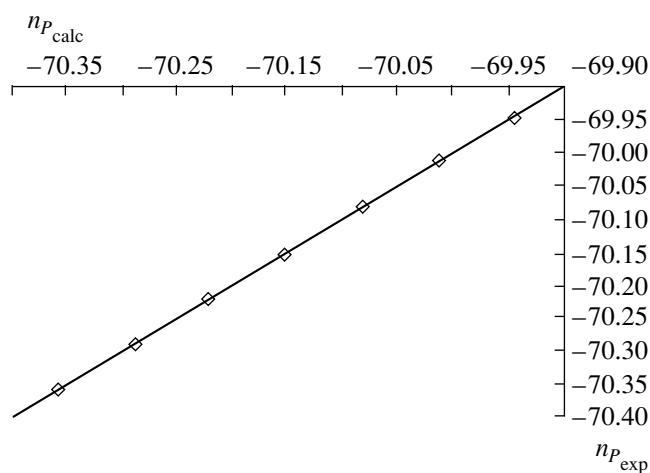


Рис. 3. Зависимость экспериментальных данных новой функции от расчетных значений для n -алканов.

Таким образом, величина

$$n_{P_{\text{exp}}}(g_i) = -BP_{\text{exp}} - \sqrt{(BP_{\text{exp}} + D)^2 - 4A(CP_{\text{exp}}^2 + EP_{\text{exp}} + F)}, \quad (25)$$

полученная из правой части уравнения (24) заменой расчетных значений октановых чисел экспериментальными, и будет новой функцией P^1 . Аналогично можно построить новые функции P^1 любого другого физико-химического свойства, если только соответствие между величиной свойства и n взаимно однозначно, т.е. функция $P(n)$ не имеет экстремумов. Эта величина зависит только от октанового числа, но в отличие от последнего успешно аппроксимируется линейной функцией от n , как следует из (23, 24)

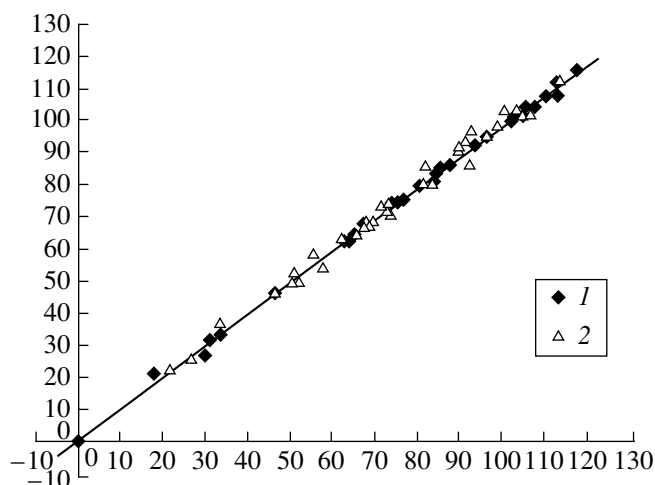


Рис. 4. Результаты моделирования для обучающей (1) и контрольной (2) выборки для алканов и циклоалканов.

(это справедливо и для любых производных P^1 других физико-химических свойств).

На рис. 3 представлена зависимость нового свойства P^1 для октановых чисел n -алканов. Видно, что в отличие от исходного свойства P (рис. 2) зависимость линейна. Подставляя в формулу (25) экспериментальные значения октанового числа, получим значения для нового свойства P^1 (табл. 1).

Далее мы провели расчеты по схеме нахождения ОТИ, предложенной в работах [4, 7]. Все имеющиеся экспериментальные данные по октановым числам мы разделили на обучающую выборку (41 соединение) и контрольную (37 соединений), чтобы получить дополнительное подтверждение достоверности прогностических свойств получае-

Таблица 1. Моделирование октановых чисел (функция P) и новой функции P^1 для n -алканов

n_1	P_{exp}	P_{calc}	$ P_{\text{exp}} - P_{\text{calc}} $	P^1_{exp}	P^1_{calc}	$ P^1_{\text{exp}} - P^1_{\text{calc}} $
0	107.5	107.3	0.2	-70.35693	-70.35686	0.00007
1	107.1	107.2	0.1	-70.28866	-70.28826	0.00040
2	105.7	106.7	1.0	-70.21862	-70.21966	0.00104
3	93.6	93.1	0.5	-70.15361	-70.15104	0.00257
4	61.8	62.4	0.6	-70.08021	-70.08246	0.00225
5	31.0	31.1	0.1	-70.01312	-70.01386	0.00074
6	0.0	-0.3	0.3	-69.94612	-69.94526	0.00086
R^2		0.99984				0.99989
s		0.496				0.00142
$ \Delta_{\text{max}} $		1.0				0.00257
a		0.9702				1.6844
b		1.0016				1.8324
n_0		2.5756				3.6211
P_0		107.5061				107.9458
α		-0.8013				-0.7747

Таблица 2. Экстраполяционная модель октановых чисел углеводородов

Углеводород	Коэффициенты
Метилпропан	-0.13768
3-Метилпентан	0.03664
Этилпентан	-0.13510
Этилгексан	0.05997
<i>n</i> -Бутан	-0.06570
Диметилпропан	0.43487
2,2-Диметилпентан	-0.24809
3,3-Диметилгексан	0.10240
2-Метил-3-этилпентан	0.13360
3-Метил-3-этилпентан	0.00408
Метан	-0.01524
2-Метилбутан	-0.02542
2,2,3-Триметилпентан	-4.30253
2,2-Диметил-3-этилпентан	-6.49473
2,4-Диметил-3-этилпентан	-5.96745
2,2,3,3-Тетраметилпентан	-12.82695
Метан	-35.59253
Этан	40.65145
Циклопентан	-76.38207
<i>n</i> -Пропилциклопентан	-0.14130
Метан	0.45885
2-Метилпропан	-0.33128
2,2-Диметилбутан	0.23695
Циклогексан	-73.85972
Метилциклогексан	0.55415
Этилциклогексан	0.38205
1,1-Диметилциклогексан	0.15148
1,4-Диметилциклогексан	0.10376
Метилпропан	-0.21985
Метилбутан	-0.13432
1,2-Диметилбутан	0.19106

Таблица 3. Обучающая выборка

Углеводород	P_{exp}	P_{calc}	$ P_{\text{exp}} - P_{\text{calc}} $
Метан	107.5	105.4	2.1
Этан	107.1	104.9	2.2
Бутан	93.6	93.7	0.1
Метилпропан	102.1	101.7	0.4
Метилбутан	93.0	93.3	0.3
Диметилпропан	85.5	85.6	0.1
Гексан	31.0	31.7	0.7
3-Метилпентан	74.5	74.3	0.2
Гептан	0.0	-0.7	0.7
Этилпентан	65.0	65.0	0.0
3,3-Диметилпентан	80.8	80.9	0.1
2,2-Диметилпентан	92.8	92.6	0.2
Триметилбутан	112.1	109.1	3.0
Этилгексан	33.5	33.6	0.1
3,3-Диметилгексан	75.5	75.6	0.1
3,4-Диметилгексан	76.3	76.3	0.0
2-Метил-3-этилпентан	87.3	87.2	0.1
3-Метил-3-этилпентан	80.8	80.8	0.0
2,2,3-Триметилпентан	109.6	108.6	1.0
2,3,4-Триметилпентан	102.7	103.0	0.3
Диэтилпентан	84.0	84.0	0.0
2,2-Диметил-3-этилпентан	112.1	112.7	0.6
2,4-Диметил-3-этилпентан	105.3	105.5	0.2
2,2,3,3-Тетраметилпентан	116.8	117.4	0.6
3,3,4-Триметилгептан	86.4	86.4	0.0
Циклопентан	101.6	101.2	0.4
1,3-Диметилциклопентан	79.9	79.9	0.0
<i>n</i> -Пропилциклопентан	31.2	31.6	0.4
Изопропилциклопентан	81.1	80.9	0.2
Изобутилциклопентан	33.4	33.2	0.2
1,1,2,4-Тетраметилциклопентан	96.2	96.3	0.1
Циклогексан	84.0	82.9	1.1
Метилциклогексан	73.8	75.5	2.1
Этилциклогексан	46.5	46.8	2.2
1,1-Диметилциклогексан	87.3	87.5	0.7
1,2-Диметилциклогексан	80.9	81.0	0.4
1,3-Диметилциклогексан	69.3	68.0	0.2
1,4-Диметилциклогексан	67.7	68.0	0.0
Изопропилциклогексан	62.8	62.8	0.1
1-Метил-1-этилциклогексан	68.7	68.6	0.1
1-Изопропил-3-метилциклогексан	67.3	67.2	0.1
	R^2	0.99894	
	s	0.82896	
	$ \Delta_{\text{max}} $	3.0	

Таблица 4. Контрольная выборка

Углеводород	P_{exp}	P_{calc}	$ P_{\text{exp}} - P_{\text{calc}} $
Пропан	105.7	103.8	1.9
2-Метилпентан	73.4	71.0	2.4
2,2-Диметилбутан	91.8	87.1	4.7
2,3-Диметилбутан	104.3	102.7	1.6
Пентан	61.8	63.9	2.1
2-Метилгексан	46.4	46.7	0.3
3-Метилгексан	52	50.1	1.9
2,3-Диметилпентан	91.1	94.5	3.4
2,4-Диметилпентан	83.1	80.8	2.3
2-Метилгептан	21.7	22.1	0.4
3-Метилгептан	26.8	25.6	1.2
4-Метилгептан	26.7	25.6	1.1
2,2-Диметилгексан	72.5	71.9	0.6
2,4-Диметилгексан	65.2	65.1	0.1
2,5-Диметилгексан	55.5	58.7	3.2
2,3-Диметилгексан	71.3	74.3	3.0
2,3,3-Триметилпентан	106.1	103.1	3.0
2,2,4-Триметилпентан	100	104.4	4.4
Тетраметилбутан	103	104.8	1.8
2,2-Диметилгептан	50.3	49.5	0.8
2,2,3,3-Тетраметилгексан	112.8	113.6	0.8
1,1-Диметилциклопентан	92.3	97.9	5.6
1,2,3-Триметилциклопентан	89.2	91.1	1.9
1-Метил-3-тилциклопентан	57.6	54.0	3.6
Метилпентан	89.3	93.1	3.8
Этилпентан	67.2	68.9	1.7
1,1,3-Триметилциклопентан	81.7	86.5	4.8
<i>n</i> -Пропилциклогексан	17.8	20.9	3.1
1,1,2-Триметилциклогексан	95.7	97.1	1.4
Изобутилциклогексан	33.7	36.9	3.2
1,2,4-Триметилциклопентан	72.9	73.6	0.7
1,3,5-Триметилциклопентан	63.8	60.4	3.4
<i>трет</i> -Бутилциклогексан	98.5	99.2	0.7
<i>втор</i> -Бутилциклогексан	51	52.9	1.9
1,1,3-Триметилциклогексан	81.3	80.4	0.9
1,2,3-Триметилциклопентан	84.8	86.2	1.4
1-Метил-2- <i>n</i> -пропилциклогексан	29.9	26.6	3.3
	R^2	0.99012	
	s	2.62038	
	$ \Delta_{\text{max}} $	5.6	

мых моделей. В табл. 2 представлены фрагменты, использованные для построения экстраполяционной модели, и их вклады в общую величину свойства P^1 в соответствии с решением матричного уравнения для связи структура–свойство. Подробно эта процедура в общем виде описана в работах [4, 7]. В табл. 3 приведены результаты ин-

терполяции по полученной модели для обучающей выборки и экстраполяции для контрольной выборки. Общий результат наших расчетов представлен в графическом виде на рис. 4.

Хотя результаты для обучающей выборки $R^2 = 0.999$ (см. табл. 3) превосходят по точности результаты для контрольной выборки $R^2 = 0.990$ (табл. 4), но все же даже последнее заметно лучше, чем у всех известных моделей: $R^2 = 0.937$ [8], $R^2 = 0.987$ [9], $R^2 = 0.971$ [10]. При этом следует заметить, что последние величины R^2 относятся к интерполяции.

Таким образом, можно сделать вывод, что предложенное нами новое направление в решении задач структура–свойство может весьма существенно повысить точность и надежность расчетных методов для предсказания физико-химического свойств органических соединений.

В рассмотренном случае обратная функция (13) была найдена непосредственно при аппроксимации зависимости P , представленной на рис. 2. Из общего соотношения (22) мы в явном виде выделили линейную часть (23) и получили сопряженное свойство P^1 , т.е. уравнение (25). Возможно, это не самая лучшая функция, но тем не менее даже она позволяет получить принципиально более точные результаты.

В общем случае обратную функцию можно построить и другими способами, исходя из конкретного вида зависимости, представленной на рис. 1. Предложенное в настоящей работе направление указывает, что для нахождения количественных соотношений структура–свойство весьма перспективным является использование обратных или максимально близких к ним функций.

Работа выполнена при поддержке гранта Президента РФ (НШ–9071.2006.3).

СПИСОК ЛИТЕРАТУРЫ

1. Needham D.E., Wei I.C., Seybold P.C. // JACS. 1988. V. 110. P. 4186.
2. Смоленский Е.А. // ДАН. 1999. Т. 368. № 2. С. 219–223.
3. Смоленский Е.А. // ДАН. 1976. Т. 30. С. 373.
4. Смоленский Е.А., Власова Г.В., Платунов Д.Ю., Рыжов А.Н. // Изв. АН. Сер. хим. 2006. № 9. С. 1454–1461.
5. Смоленский Е.А. // ЖФХ. 1966. Т. 40. № 12. С. 2982–2985.
6. Демидович Б.П., Марон И.А., Шувалова Э.З. Численные методы анализа М.: Физматиздат, 1963. 400 с.
7. Смоленский Е.А. // Изв. АН. Сер. хим. 2006. № 9. С. 1448–1453.
8. Randic M. // J. Chem. Inf. Comput. Sci. 1997. V. 37. № 4. P. 672–685.
9. Сидорова А.В., Баскин И.И., Петелин Д.Е. и др. // ДАН. 1996. Т. 350. № 5. С. 642–646.
10. Смоленский Е.А., Власова Г.В., Липидус А.Л. // ДАН. 2004. Т. 397. № 2. С. 219–223.