

УДК 541.61

© 1990 г.

РАЕВСКИЙ О. А., САПЕГИН А. М., ЧИСТЯКОВ В. В.,  
СОЛОВЬЕВ В. П., ЗЕФИРОВ Н. С.

## ФОРМИРОВАНИЕ МОДЕЛЕЙ ВЗАИМОСВЯЗИ СТРУКТУРА — КОМПЛЕКСООБРАЗУЮЩАЯ СПОСОБНОСТЬ

Статья посвящена выявлению взаимосвязи между структурой лигандов и их способностью связывать кальций в водной среде с использованием усовершенствованной методики развиваемого авторами системного физико-химического подхода к моделированию связи структура—свойство. Для построения моделей использовался собственный и литературный массив данных по энтальпии и константам устойчивости комплексов 56 разнообразных лигандов: органических кислот, фосфорильных соединений, макроциклических полиэфиров, криптанов, подандов, природных соединений. В работе доказана эффективность дискретно-непрерывных моделей связи структура—свойство. Обнаружены различные кластеры соединений по отношению к связыванию кальция в воде. Для каждого кластера установлены количественные зависимости между дескрипторами структуры, с одной стороны, и значениями энтальпии или констант устойчивости кальциевых комплексов, с другой стороны.

Ускорение научно-технического прогресса в такой важной для народного хозяйства области, как разработка эффективных комплексонов, в настоящее время уже не может быть обеспечено за счет дорогостоящих и малоэффективных эмпирических методов поиска соединений с нужными свойствами, основанных на скрининге. Малодоступными для широкого распространения из-за своей сложности остаются и достаточно строгие расчеты прочности и устойчивости комплексов, базирующиеся на таких методах, как, например, молекулярная динамика [1]. Поэтому в последние годы ведущую роль в целенаправленном поиске и конструировании веществ начинают играть методы выявления взаимосвязи структура—свойство, которые нашли практическое воплощение еще в моделях Хэнча и Фри-Вильсона [2] в начале 60-х годов. В настоящее время это научное направление вобрало в себя довольно широкий спектр методов и моделей, включая целые области знаний из смежных наук (такие, например, как развиваемые в кибернетике методы распознавания образов [3] и искусственного интеллекта [4]).

В данной работе предпринята попытка выявления взаимосвязи между структурой лигандов и их способностью связывать кальций в водной среде с использованием усовершенствованной методики развиваемого нами системного физико-химического подхода к моделированию связи структура—свойство [5].

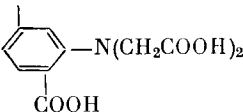
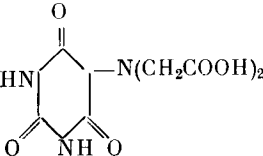
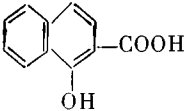
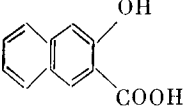
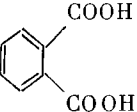
Выбор кальция в качестве акцептора был продиктован его важной ролью в процессах жизнедеятельности [6]. Массив данных по энтальпии ( $\Delta H$ ) и константам устойчивости ( $\lg K$ ) сформирован на основе литературных данных [7] и собственных экспериментальных исследований (табл. 1).

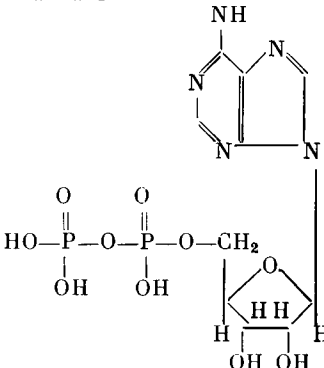


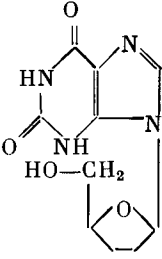
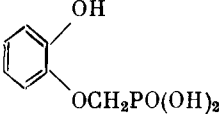
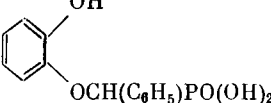
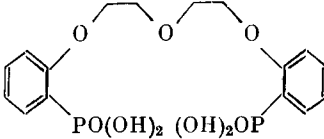
### ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

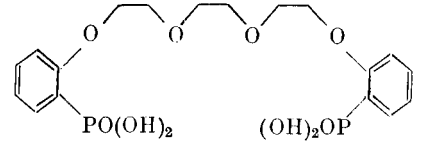
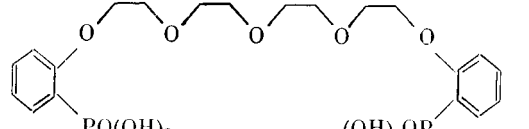
Оценку констант комплексообразования (для соединений 50—56) производили на основе данных по потенциометрическому титрованию (использовали автотитратор RADIOMETR PHM-82) при 298 К, постоянной ионной силе  $\mu = 0,1$  в атмосфере аргона, очищенного от примесей.

Измерения теплот реакций для соединений 48, 49 проводили на калориметре ЛКВ-2107/112 при 298 К. Использовали метод растворной калориметрии.

Формулы соединений, входящих в массив данных по энтальпии ( $\Delta H$ )  
и константам устойчивости ( $\lg K$ )

| № п/п | Формула соединения  |
|-------|---|
| 1     | $\text{CH}_3\text{COOH}$  |
| 2     | $\text{HOCH}_2\text{COOH}$  |
| 3     | $\text{C}_6\text{H}_5\text{CH}(\text{OH})\text{COOH}$   |
| 4     | $\text{HN}(\text{CH}_2\text{COOH})_2$   |
| 5     | $\text{HOOC}(\text{CH}_2)_2\text{N}(\text{CH}_2\text{COOH})_2$  |
| 6     | $(o\text{-HOOC}_6\text{H}_4)_2\text{N}(\text{CH}_2\text{COOH})_2$<br>$\text{COOH}$  |
| 7     |    |
| 8     | $(\text{CH}_2\text{COOH})_2\text{N}(\text{CH}_2\text{COOH})$  |
| 9     | $\text{CH}_3\text{N}(\text{CH}_2\text{COOH})_2$   |
| 10    | $(\text{CH}_2\text{CH}_2\text{NH}_2)\text{N}(\text{CH}_2\text{CH}_2\text{NH}_2)_2$  |
| 11    |    |
| 12    | $(\text{HOOCCH}_2)_2\text{N}(\text{CH}_2)_4\text{N}(\text{CH}_2\text{COOH})_2$  |
| 13    | $(\text{HOOCCH}_2)(\text{COOH})\text{C}(\text{OH})(\text{CH}_2\text{COOH})$   |
| 14    | $\text{C}_6\text{H}_{10}(\text{N}(\text{CH}_2\text{COOH})_2)_2$   |
| 15    | $\text{C}_5\text{H}_8(\text{N}(\text{CH}_2\text{COOH})_2)_2$  |
| 16    | $(\text{HOOCCH}_2)_2\text{N}(\text{CH}_2)_2\text{N}(\text{CH}_2\text{CH}_2\text{OH})(\text{CH}_2\text{COOH})$                     |
| 17    | $(\text{HOOCCH}_2)_2\text{N}(\text{CH}_2)_2\text{N}(\text{CH}_2\text{COOH})_2$  |
| 18    | $(\text{HOOCCH}_2)_2\text{NCH}_2\text{O}(\text{CH}_2)_2\text{N}(\text{CH}_2\text{COOH})_2$  |
| 19    | $\text{HCOOH}$  |
| 20    | $\text{HOOCCH}_2\text{CH}_2\text{SCH}_2\text{CH}_2\text{COOH}$  |
| 21    | $(\text{HOOCCH}_2)_2\text{N}(\text{CH}_2)_6\text{N}(\text{CH}_2\text{COOH})_2$  |
| 22    |    |
| 23    |    |
| 24    | $(\text{HOOCCH}_2)_2\text{N}(\text{CH}_2\text{CH}_2\text{O})_2(\text{CH}_2)_2\text{N}(\text{CH}_2\text{COOH})_2$                  |
| 25    | $\text{HOOC}-\text{COOH}$   |
| 26    | $(\text{HOOCCH}_2)_2\text{N}(\text{CH}_2)_2\text{N}(\text{CH}_2)_2\text{N}(\text{CH}_2\text{COOH})_2$<br>$\text{CH}_2\text{COOH}$ |
| 27    | $(\text{HOOCCH}_2)_2\text{N}(\text{CH}_2)_2\text{O}(\text{CH}_2)_2\text{N}(\text{CH}_2\text{COOH})_2$                             |
| 28    |    |
| 29    | $(\text{HOOCCH}_2)_2\text{N}(\text{CH}_2)_3\text{N}(\text{CH}_2\text{COOH})_2$  |
| 30    | $\text{CH}_3(\text{OH})\text{CH}(\text{COOH})$  |
| 31    | $\text{HOOCCH}_2(\text{OH})\text{CH}(\text{COOH})$  |
| 32    | $(\text{HOOCCH}_2)_2\text{N}(\text{CH}_2)_2\text{S}(\text{CH}_2)_2\text{N}(\text{CH}_2\text{COOH})_2$                             |
| 33    | $\text{HOOCCH}(\text{OH})\text{CH}(\text{OH})(\text{COOH})$   |

| № п/п | Формула соединения  |
|-------|---|
| 34    | Аденозин-5'-дифосфорная кислота<br>                                |
| 35    | Аденозин-2'-монофосфорная кислота   |
| 36    | Аденозин-5'-трифосфорная кислота  |
| 37    | $\text{CH}_3\text{CH}(\text{PO}(\text{OH})_2)_2$  |
| 38    | $\text{CH}_3(\text{OH})\text{C}(\text{PO}(\text{OH})_2)_2$  |
| 39    | $\text{CH}_2-\text{CH}(\text{CHOH})_3\text{CHOPO}_3\text{H}_2$<br> |
| 40    | $\text{CH}_2(\text{PO}(\text{OH})_2)_2$   |
| 41    | $\text{CH}=\text{CHSC}(\text{COOH})=\text{CH}$<br>                 |
| 42    | $(\text{CH}_3)_2\text{C}(\text{PO}(\text{OH})_2)_2$   |
| 43    |   |
| 44    | 3,2,2-крипанд   |
| 45    | 2,2,2-крипанд   |
| 46    | 2,1,2-крипанд   |
| 47    | 1,1,2-крипанд   |
| 48    | 18-краун-6  |
| 49    | 15-краун-5  |
| 50    | $\text{C}_6\text{H}_5\text{OCH}_2\text{PO}(\text{OH})_2$  |
| 51    |    |
| 52    |    |
| 53    | $\text{H}_4\text{C}_6(\text{OCH}_2\text{PO}(\text{OH})_2)_2$  |
| 54    |    |

| № п/п | Формула соединения  |
|-------|---|
| 55    |  |
| 56    |  |

Константы комплексообразования рассчитывали по программе, алгоритм которой описан в [8].

В работе использовали известные данные по термодинамическим характеристикам, полученным в указанных условиях эксперимента [7].

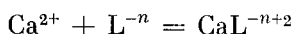
### МЕТОДИКА МОДЕЛИРОВАНИЯ И АНАЛИЗА ДАННЫХ

Для обеспечения полноты и корректности описания структуры лигандов на первом этапе моделирования проводили расчет дескрипторов структуры, охватывавших значительную часть многообразных информационно-топологических, электронно-топографических и физико-химических молекулярных свойств [9]. Структурные дескрипторы (в форме скринов [10]) использовали только для проверки гипотезы об аддитивном вкладе фрагментов структуры в комплексообразующую способность лигандов. Из остального набора дескрипторов (общее число — 41) с помощью пошагового дискриминантного анализа [11] отбирали наиболее информативные, на базе которых строили классификационное правило и проводили кластерный анализ соединений в этом дискриминантном пространстве. Характер кластеризации (форма и относительное расположение кластеров) соединений и описывающих их структуру дескрипторов изучали методами факторного анализа [12]. Для наглядного отображения кластеризации соединений по отношению к дискриминантной гиперплоскости, разделяющей соединения на два класса (в нашем случае, к первому классу были отнесены соединения с  $\lg K > 4$  и  $\Delta H < -5$  кДж/моль, а ко второму — все остальные), было использовано SLD (Steam-and-Leaf-Displays)-представление «разведочного» анализа данных [13].

Для каждого выявленного кластера соединений, опираясь на информативный набор дескрипторов, методом пошаговой регрессии строили количественную модель связи структура—комплексообразующая способность, которую затем проверяли на устойчивость с помощью сингулярного анализа [14] и на адекватность — методами дисперсионного анализа.

### РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Для проверки гипотезы об аддитивном вкладе фрагментов структуры в комплексообразующую способность лигандов из 56 соединений, для которых равновесные константы взаимодействия с  $\text{Ca}^{2+}$  определены при 298 К и ионной силе 0,1 для равновесия



в воде, были отобраны 26 лигандов, в которых каждый фрагмент структуры присутствовал не менее трех раз (исключение составили лишь фрагменты  $(\text{H})_2\text{C}(\text{C})(\text{CO})$ ,  $\text{C}_b(\text{H})$ ,  $\text{O}(\text{C}_b)(\text{H})$ ,  $\text{C}(\text{C})(\text{H})_3$  и  $\text{C}(\text{C})_2(\text{PO})_2$ . Под фрагментом структуры ( $G$ ) подразумевается атом вместе с соседними атомами его ближайшего окружения с учетом их гибридизации. Для автоматического

Фрагменты лигандов и вклады ( $X$ ) этих фрагментов в логарифм равновесных констант комплексообразования с Са в водной среде (вклады рассчитаны по формуле (1))

| Фрагмент *                         | $X$   | Фрагмент *                             | $X$   |
|------------------------------------|-------|--|-------|
| C=O(C) (OH)                        | 0,42  | O(C)                                   | -0,08 |
| O(H) (C=O)                         | 0,85  | C(H) (C) (O) (CO)                      | -0,07 |
| O(H) (C)                           | -0,07 | C(H) <sub>2</sub> (C)                  | 0,42  |
| C <sub>b</sub> (H)                 | -0,04 | O(H) (PO)                              | 0,90  |
| C(H) <sub>2</sub> (CO) (N)         | 0,71  | PO(O) <sub>2</sub> (C)                 | 0,65  |
| N(C)                               | 0,40  | C(C) <sub>2</sub> (PO)                 | 0,58  |
| C(H) <sub>2</sub> (C) (CO)         | -0,23 | O(C) (C <sub>b</sub> )                 | -0,72 |
| C(H) <sub>2</sub> (N) (C)          | 0,47  | C(H) <sub>2</sub> (O) (PO)             | -0,52 |
| C(H) <sub>2</sub> (C) <sub>2</sub> | -1,63 | PO(OH <sub>2</sub> ) (C <sub>b</sub> ) | -0,19 |
| C(H) <sub>2</sub> (C) (O)          | -0,03 |  |       |

\* Обозначения фрагментов и гибридизации соответствуют [10].

определения типов фрагментов и их числа в лигандах была разработана программа TRAIL (мини-ЭВМ NORD-10, FORTRAN-77).

Вклады фрагментов в значение  $\lg K$  — согласно аддитивной схеме — были определены из системы линейных уравнений

$$\sum_{i=1}^{19} X_i n_{ij} (G_i) = \lg K, \quad j = 1, \dots, 26 \quad (1)$$

где  $X_i$  и  $n_{ij} (G_i)$  — вклад и количество  $i$ -х фрагментов ( $G_i$ ) в  $j$ -м лиганде, соответственно. Для обеспечения корректности определения вкладов проводили сингулярный анализ [14] матрицы  $n_{ij} (G_i)$ . Результаты расчета значений  $\lg K$  по определенным из (1) вкладам фрагментов (табл. 2) представлены в табл. 3. Сравнение расчетных и экспериментальных значений  $\lg K$  показывает, что их различие не превышает 10% для большинства лигандов. Исключение составили (как, впрочем, и следовало ожидать, учитывая слишком упрощенный характер описания циклических структур при их фрагментации) криптанды и краун-эфир, относящиеся к группе циклических лигандов.

В целом при рассмотрении результатов расчета термодинамических параметров методом скринов можно отметить существенный вклад в устойчивость комплексов таких фрагментов, как O(H)(C=O), C(H)<sub>2</sub>(CO)(N), O(H)PO, PO(O)<sub>2</sub>(C), что согласуется с общепринятыми представлениями о комплексообразующей способности входящих в них функциональных групп, а также дестабилизирующее влияние на образование комплексов наличия фрагментов типа C(H)<sub>2</sub>(C)<sub>2</sub>, O(C)(C<sub>b</sub>). Полученные результаты показывают, что, несмотря на такое довольно формальное описание структуры, аддитивная схема может быть весьма полезной при моделировании различных свойств соединений с общими структурными фрагментами.

С целью дальнейшего уточнения описания структуры были изучены информационно-топологические, электронно-топографические и физико-химические свойства лигандов, выраженные через набор соответствующих дескрипторов:

1. Молекулярная масса ( $MW$ )
2. Молекулярный объем ( $MV$ )
3. Молекулярная рефракция ( $MR$ )
4.  $\lg P$
5. Число атомов в молекуле ( $N$ )
6. Число донорных центров ( $N_{DC}$ )
7. Число вершин в граф-представлении молекулы ( $N^0$ )
8. Число атомов водорода в молекуле ( $N_H$ )
9. Простой индекс ( $SI$ )
10. Смешанный индекс ( $MI$ )
11. Индекс Винера ( $W$ )
12. Индекс Рандича  $\chi = \Sigma(\nabla_i \nabla_j)^{-1/2}$

## Термодинамические характеристики комплексообразования кальция в воде (298 К)

| № соедине-<br>ния | ΔH, кДж/моль |          |        | lg K    |          |       |       |
|-------------------|--------------|----------|--------|---------|----------|-------|-------|
|                   | экспер.      | расчет * |        | экспер. | расчет * |       |       |
|                   |              | 2        | 3      |         | 1        | 2     | 3     |
| 1                 | 3,81         | 6,26     | —      | 1,24    | —        | 0,94  | 1,66  |
| 2                 | 35,50        | 32,10    | 3,82   | 1,65    | —        | 0,68  | 1,70  |
| 3                 | 32,05        | 3,35     | 72,40  | 1,46    | —        | 3,23  | 1,68  |
| 4                 | -1,26        | 23,48    | 2,90   | 2,59    | —        | 4,07  | 3,17  |
| 5                 | -1,05        | -9,80    | -6,10  | 5,04    | 5,88     | 6,67  | 8,38  |
| 6                 | 5,61         | 16,96    | —      | 5,08    | —        | 6,54  | 5,05  |
| 7                 | 6,70         | 6,20     | 62,80  | 5,05    | —        | 8,02  | 6,99  |
| 8                 | -3,40        | 18,59    | -8,24  | 6,57    | 6,34     | 6,19  | 4,03  |
| 9                 | -5,00        | 14,79    | 6,65   | 3,85    | —        | 4,55  | 3,14  |
| 10                | -23,18       | -35,30   | 1,31   | 8,43    | —        | 3,16  | 8,64  |
| 11                | -28,00       | -28,60   | —      | 8,15    | —        | 5,82  | 3,78  |
| 12                | -3,80        | -16,90   | -8,73  | 5,66    | 6,40     | 3,40  | 8,10  |
| 13                | -10,40       | -2,50    | -15,83 | 3,10    | —        | 3,86  | 4,15  |
| 14                | -25,90       | -28,40   | —      | 12,20   | —        | 11,78 | 9,30  |
| 15                | -20,38       | -24,05   | 3,55   | 11,37   | —        | 11,59 | 9,30  |
| 16                | -27,20       | -40,14   | —      | 8,06    | 8,06     | 8,60  | 6,02  |
| 17                | -25,19       | -6,52    | —      | 10,97   | 9,66     | 11,03 | 9,30  |
| 18                | -28,66       | -36,37   | —      | 10,00   | —        | 10,23 | 9,64  |
| 19                | 4,40         | -5,16    | —      | 1,42    | —        | 0,56  | 1,57  |
| 20                | -8,49        | -6,48    | -9,47  | 2,99    | —        | 1,70  | 2,77  |
| 21                | -5,40        | 7,60     | —      | 3,69    | 3,13     | 5,49  | 9,05  |
| 22                | 191,00       | 156,68   | 134,90 | 4,75    | —        | 4,40  | 0,81  |
| 23                | 137,90       | 136,30   | 134,40 | 4,42    | —        | 2,79  | 0,81  |
| 24                | -33,50       | -25,70   | —      | 10,80   | 9,72     | 10,39 | >15   |
| 25                | -21,30       | -21,90   | -2,42  | 8,64    | —        | 6,50  | 6,20  |
| 26                | -25,50       | -55,00   | —      | 10,70   | 13,00    | 11,16 | >15   |
| 27                | -26,80       | -23,20   | —      | 9,90    | 9,69     | 6,21  | 8,81  |
| 28                | 29,75        | 64,85    | —      | 2,41    | —        | 3,10  | 2,66  |
| 29                | -7,28        | -14,10   | —      | 7,28    | 8,03     | 6,12  | 7,64  |
| 30                | 45,65        | 29,11    | 7,83   | 1,55    | 1,55     | 1,34  | 1,68  |
| 31                | 36,00        | 21,60    | —      | 3,01    | 2,17     | 3,01  | 2,99  |
| 32                | -10,50       | -11,40   | —      | 6,21    | —        | 5,67  | 8,20  |
| 33                | -14,68       | 6,07     | —      | 1,84    | 2,26     | 2,52  | 2,55  |
| 34                | -2,50        | -4,10    | —      | 1,82    | —        | 3,88  | 3,48  |
| 35                | -5,00        | -5,00    | —      | 2,86    | —        | 3,34  | 3,78  |
| 36                | -3,80        | -5,50    | —      | 3,97    | —        | 3,43  | 5,08  |
| 37                | 9,20         | 10,94    | -10,10 | 5,22    | —        | 4,31  | 3,13  |
| 38                | -2,18        | -7,16    | —      | 5,75    | —        | 4,57  | 10,50 |
| 39                | 10,05        | 2,78     | —      | 2,50    | —        | 3,05  | 2,10  |
| 40                | 0,00         | 20,50    | —      | 4,77    | —        | 4,38  | 5,14  |
| 41                | 18,20        | 15,52    | 22,78  | 2,01    | —        | 2,60  | 1,67  |
| 42                | -4,27        | -0,81    | —      | 6,34    | 6,34     | 4,84  | 1,05  |
| 43                | -6,70        | -0,08    | —      | 2,37    | —        | 1,55  | 1,84  |
| 44                | 0,67         | -2,61    | -9,40  | 1,98    | 3,84     | 2,75  | 1,77  |
| 45                | -0,84        | 1,78     | -3,40  | 4,40    | 3,81     | 3,68  | 11,20 |
| 46                | -7,10        | 3,28     | 2,60   | 6,95    | 3,78     | 4,18  | 8,14  |
| 47                | -0,40        | 7,70     | 8,57   | 2,49    | 3,75     | 4,16  | 1,67  |
| 48 **             | -9,44        | -12,74   | -4,24  | 0,50    | 0,18     | 2,86  | 0,65  |
| 49 **             | —            | —        | —      | 0,80    | 0,15     | —     | —     |
| 50 **             | —            | —        | —      | 1,81    | 1,81     | 1,77  | —     |
| 51 **             | —            | —        | —      | 4,19    | 4,19     | 3,67  | —     |
| 52 **             | —            | —        | —      | 4,11    | —        | 5,30  | —     |
| 53 **             | —            | —        | —      | 3,88    | 3,88     | 6,30  | —     |
| 54 **             | —            | —        | —      | 3,42    | 3,05     | 4,60  | —     |
| 55 **             | —            | —        | —      | 3,10    | 3,08     | 2,48  | —     |
| 56 **             | —            | —        | —      | 2,71    | 3,11     | -0,50 | —     |

\* Расчет 1 — по скринам; расчеты 2 и 3 — по регрессионным уравнениям, полученным для кластеров из факторного анализа и SLD-представления соответственно.

\*\* Собственные результаты. Для остальных соединений использованы данные [7].

13. Сумма однокоординированных вершин в графе ( $N_{\sigma=1}$ )
14. Сумма двухкоординированных вершин в графе ( $N_{\sigma=2}$ )
15. Сумма трехкоординированных вершин в графе ( $N_{\sigma=3}$ )
16. Сумма четырехкоординированных вершин в графе ( $N_{\sigma=4}$ )
17. Число простых связей в молекуле ( $N_{sb}$ )
18. Число двойных связей в молекуле ( $N_{db}$ )
19. Число тройных связей в молекуле ( $N_{tb}$ )
20. Число ароматических связей в молекуле ( $N_{ar}$ )
- 21–23. Координационные индексы
24.  $E_{тл} = \Sigma(E_i E_j) / \Delta$ , где  $\Delta$  – число связей между  $i$ -м и  $j$ -м донорными центрами
25.  $E = \Sigma(E_i E_j) / L$ , где  $L$  – сумма длин связей между  $i$ -м и  $j$ -м донорными центрами
26.  $D = \Sigma(E_i E_j) / MW$ , где  $MW$  – молекулярная масса
- 27–38. Информационно-топологические индексы на базе формулы Шеннона
- 39–41. Квадраты некоторых вышеперечисленных дескрипторов.

На основе пошаговой процедуры дискриминантного анализа (результаты которого сведены в табл. 4) были выделены наиболее информативные дескрипторы, пять из которых (молекулярный вес, число двухкоординированных вершин в молекулярном графе, смешанный индекс, донорно-акцепторные индексы  $E_{тл}$ ,  $D$ ) оказались одинаковыми как для модели описания взаимосвязи «структура —  $\Delta H$ », так и для модели «структура —  $\lg K$ ».

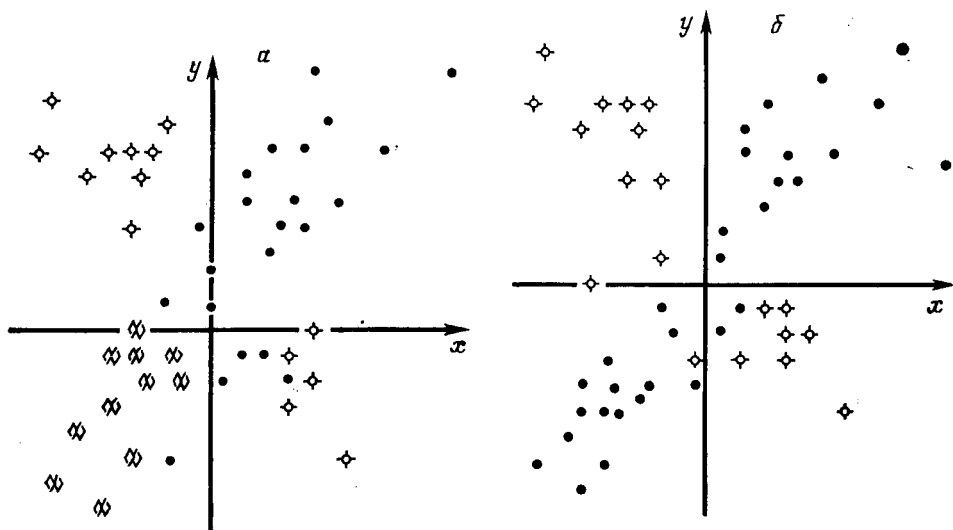
Таблица 4

Результаты дискриминантного анализа данных по комплексообразованию с кальцием в воде

| Класс                                  | Количество соединений | Достоверность классификации обучающей выборки, число (%) | Скользкий контроль, число (%) |             |
|--|-----------------------|--|-------------------------------|-------------|
|  |                       |  | правильно                     | неправильно |
| <b>Данные по <math>\Delta H</math></b> |                       |  |                               |             |
| 1                                      | 21                    | 21 (100)   | 15 (71)                       | 6 (29)      |
| 2                                      | 27                    | 27 (100)   | 22 (82)                       | 5 (18)      |
| Всего                                  | 48                    | 48 (100)   | 37 (77)                       | 11 (23)     |
| Прочн. оценка                          | 240                   | —  | 214 (89)                      | 26 (11)     |
| <b>Данные по <math>\lg K</math></b>    |                       |  |                               |             |
| 1                                      | 26                    | 26 (100)   | 23 (89)                       | 3 (11)      |
| 2                                      | 22                    | 22 (100)   | 19 (86)                       | 3 (14)      |
| Всего                                  | 48                    | 48 (100)   | 42 (88)                       | 6 (12)      |
| Прочн. оценка                          | 240                   | —  | 223 (93)                      | 17 (7)      |

Кроме перечисленных, в модель описания структуры (МОС) для  $\Delta H$  вошли следующие дескрипторы: молекулярная рефракция, молекулярный объем, число донорных центров и однокоординированных вершин, индексы координации и индексы Шеннона, а для  $\lg K$  — гидрофобность, число атомов в молекуле, число вершин в молекулярном графе, число трех- и четырехкоординированных вершин, простой индекс, число простых связей. Полученные в процессе этого моделирования результаты показывают, что уже на стадии дискриминантной модели существует заметное различие между достоверностью классификации при обучении и скользком контроле (табл. 4). Это очевидное свидетельство несовершенства построенной модели в значительной степени объясняется недостаточной представительной выборкой лигандов для выявления всех влияющих на процесс комплексообразования факторов. На основе проведенных расчетов можно предполагать, что увеличение обучающей выборки лигандов, например, до 240 соединений, могло бы снизить вероятность ошибочной классификации до 11% для модели прогноза  $\Delta H$  и до 7% для модели прогноза  $\lg K$ .

На основе входящих в МОС дескрипторов был проведен также  $Q$ -факторный анализ (рисунок), который показал, что набор соединений обучающей выборки не является однородным. Об этом же свидетельствуют и неудачные результаты построения единой регрессионной модели для всей выборки соединений. В связи с этим для модели, описывающей  $\Delta H$ , были выделены три кластера, для каждого из которых построена методом по-



Кластеризация по результатам  $Q$ -факторного анализа для  $\Delta H$  (а) и  $\lg K$  (б). На осях  $x$  и  $y$  отложены первый и второй главные характеристические векторы SVD-разложения [14] соответственно

шаговой регрессии зависимость имела вид:

1. Соединения 7, 12, 13, 18, 21, 23–25, 27, 29, 31–33, 39

$$\Delta H = 224,5 + 3,66 (\lg P) - 11,7 (N_{DC}) - 1592 (D) \quad (2)$$

( $R = 0,97$   $sd = 12,3$   $n = 14$   $F = 59$ );

2. Соединения 1–4, 8, 9, 19, 30, 37, 40, 41

$$\Delta H = 39,70 - 11,86 (\lg P) - 0,32 (SI) - 3198 (D^2) \quad (3)$$

( $R = 0,89$   $sd = 9,7$   $n = 11$   $F = 8,5$ );

3. Соединения 5, 6, 10, 11, 14–17, 20, 22, 26, 28, 32, 34–36, 42–48

$$\Delta H = 96,48 + 3,14 (MR) + 0,44 (MI) + 32,83(\chi) - 4864(D) + 43745 (D^2) \quad (4)$$

( $R = 0,95$   $sd = 15,9$   $n = 23$   $F = 29,7$ )

В случае построения модели для  $\lg K$  (рисунок, б) были выделены два кластера соединений с номерами 1–6, 9–11, 14–17, 19, 22, 26, 30, 34–38, 40–46, 48 (кластер 1) и 7, 12, 13, 18, 20, 21, 23–25, 27–29, 31–33, 39, 47 (кластер 2), описываемые зависимостями:

кластер 1

$$\lg K = -0,58 + 0,12 (MR) - 0,42 (N_{DC}) - 0,03 (W) + 0,26 (E) \quad (5)$$

( $R = 0,93$   $sd = 1,3$   $n = 31$   $F = 40,6$ )

кластер 2

$$\lg K = 1,32 + 0,0035 (W) - 0,74 (E_{\text{ТЛ}}) + 1,095 (E) \quad (6)$$

( $R = 0,85$   $sd = 1,6$   $n = 17$   $F = 12,3$ )

Для выявления характера кластеризации соединений обучающей выборки по отношению к моделируемым термодинамическим параметрам  $\Delta H$  и  $\lg K$  был использован и подход, основанный на многомерном шкалировании [15] в форме SLD-представления [13] с кластеризацией соединений при проецировании описывающих их структуру векторов в дискриминантном пространстве дескрипторов на направление вектора «обобщенного портрета» разделяющей гиперплоскости [11] для МОС  $\lg K$ :



|       |                        |
|-------|------------------------|
| 1,4   |                        |
| 1,3   | 1, 3, 3, 3             |
| 1,2   |                        |
| 1,1   | 1, 1                   |
| 1,0   | 8, 9                   |
| 0,9   | 1, 3, 4                |
| 0,8   |                        |
| 0,7   |                        |
| 0,6   | 2, 4, 6                |
| 0,5   | 1, 1, 1, 1, 2, 3, 3, 3 |
| 0,4   | 6, 8, 9, 9             |
| 0,3   |                        |
| 0,2   |                        |
| 0,1   |                        |
| <hr/> |                        |
| -0,1  |                        |
| -0,2  |                        |
| -0,3  |                        |
| -0,4  | 6, 6, 7, 8             |
| -0,5  | 2, 2, 2, 2, 4, 6, 7, 9 |
| -0,6  |                        |
| -0,7  |                        |
| -0,8  | 0, 2, 2, 2, 4, 5, 6    |
| -0,9  | 5, 8                   |
| -1,0  | 0                      |
| -1,1  |                        |

Для четырех выявленных кластеров соединений с номерами:

кластер 1 — 7, 8, 12, 14—18, 27, 29, 32,

кластер 2 — 5, 6, 10, 22—26, 37, 38, 40, 42, 45, 46,

кластер 3 — 1, 3, 13, 21, 30, 33, 34, 39, 41, 43, 44, 47,

кластер 4 — 2, 4, 9, 11, 19, 20, 28, 31, 35, 36, 48,

были построены регрессионные зависимости:

кластер 1:

$$\lg K = 1,62 + 0,0026(E_{\text{ТЛ}}^2) \quad (7)$$

$$(R = 0,86 \quad sd = 1,9 \quad n = 11 \quad F = 59)$$

кластер 2:

$$\lg K = 3,03 - 0,28(N_{\text{H}}) + 1,01(\lg P^2) \quad (8)$$

$$(R = 0,87 \quad sd = 1,5 \quad n = 14 \quad F = 20)$$

кластер 3:

$$\lg K = -0,1 + 0,33(N(sb)) - 0,0045(IW) + 0,003(E_{\text{ТЛ}}^2) \quad (9)$$

$$(R = 0,90 \quad sd = 1,8 \quad n = 12 \quad F = 27)$$

кластер 4:

$$\lg K = 1,06 - 0,001(W) + 0,17(E) \quad (10)$$

$$(R = 0,95 \quad sd = 0,4 \quad n = 11 \quad F = 31)$$

В отличие от модели для  $\lg K$ , где применение шкалирования привело к большей по сравнению с использованием факторного анализа детализации описания взаимосвязи структура — комплексобразующая способность, модель для  $\Delta H$  значительно упростилась. Был обнаружен только один кластер соединений (с номерами 2—5, 7—10, 12, 13, 15, 20, 22, 23, 30, 34, 37, 41, 44—48), для которого построенная регрессионная зависимость имела вид:

$$\Delta H = 31,65 - 12,39(N_{DC}) + 0,26(MI) \quad (11)$$

$$(R = 0,87 \quad sd = 2,5 \quad n = 25 \quad F = 34)$$

Такое различие между моделями для  $\lg K$  и  $\Delta H$ , особенно при сравнении характера кластеризации в факторном и дискриминантном пространствах МОС, по-видимому, связано с тем, что на формирование модели по  $\Delta H$  влияет широкий спектр взаимосвязанных структурных свойств сое-

динений. Общие тенденции изменения их хорошо отражаются главными компонентами в факторном анализе, но не могут быть выделены при выборе только одного направления в пространстве структурных дескрипторов. Наоборот, модели по  $\lg K$  характеризуются более специфичными структурными свойствами, которые могут быть выделены только по отношению к определенному, задающему максимальное изменение моделируемого свойства, направлению (как в случае многомерного шкалирования), а не из описывающего общие закономерности  $Q$ -факторного анализа.

Возможность использования полученных МОС для предсказания устойчивости ( $\lg K$ ) комплексов полидентатных лигандов с кальцием в водной среде может быть продемонстрирована сравнением рассчитанных и полученных нами экспериментальных значений  $\lg K$  для соединений 50—56. Расчет величин  $\lg K$  осуществляли в два этапа: на первом проводили отнесение новых соединений к выявленным для этого свойства кластерам из факторного анализа, а на втором — использование соответствующих этим кластерам регрессионных уравнений. Результаты, представленные в табл. 3, свидетельствуют, что для двух из семи соединений имеется существенное различие между экспериментальными и рассчитанными по указанной схеме значениями  $\lg K$ . Для оставшихся пяти соединений согласие эксперимента и расчета удовлетворительное.

В целом проведенное исследование показало, что при моделировании комплексообразующей способности лигандов с разнородной структурой необходимо учитывать неоднородность обучающей выборки соединений. С этой точки зрения методика построения дискретно-регрессионных моделей связи структура — свойство, где на первом этапе проводится кластеризация соединений, а затем для каждого полученного кластера строится количественная модель, выглядит довольно эффективной. Дальнейшее улучшение согласия между экспериментальными и расчетными значениями для  $\Delta H$  и  $\lg K$  может быть достигнуто как за счет уточнения описания структуры (в том числе специфических особенностей строения «выпадающих» из модели лигандов), так и за счет расширения экспериментальных данных, используемых при формировании обучающей выборки.

#### СПИСОК ЛИТЕРАТУРЫ

1. *Naray-Szabo G., Surjan P.* // Theoretical Chemistry of Biological System / Ed. *Naray-Szabo G.*, Amsterdam: Elsevier, 1986. P. 1.
2. *Franke R.* Theoretical Drug Design Methods. Berlin: Akademie-Verlag, 1984. 412 S.
3. *Голендер В. Е., Розенблат А. Б.* Логико-комбинаторные методы в конструировании лекарств. Рига: Зинатне, 1983. 352 с.
4. Построение экспертных систем / Под ред. *Хейес-Рот Ф.* М.: Мир, 1987. 434 С.
5. *Раевский О. А., Сапегин А. М.* // Успехи химии. 1988. Т. 57. № 9. С. 1525.
6. *Calcium in Biology*/Ed. *Spiro T. G.* N. Y.: J. Wiley, 1983. 248 p.
7. *Christensen J. J., Izatt R. M.* Handbook of metal ligand heats. N. Y.: M. Dekker. 1983. 783 p.
8. *Новиков В. П., Раевский О. А.* Изв. АН СССР. Сер. хим. 1983. № 6. С. 1336.
9. *Стьюпер Э., Брюггер У., Джурс П.* Машинный анализ связи химической структуры и биологической активности. М.: Мир, 1982. 235 с.
10. *Рид Р., Праусниц Дж.* Свойства газов и жидкостей. Л. Химия, 1982. 382 С.
11. Алгоритмы и программы восстановления зависимостей / Под. ред. *Валника В. Н.* М.: Наука. 1984. 816 с.
12. *Иберла К.* Факторный анализ. М.: Статистика, 1980.
13. *Velleman P. F., Hoaglin D. C.* Applications, Basics and Computing of Exploratory Data Analysis. Boston, Mass.: Duxbury Press, 1981.
14. *Форсайт Дж., Малькольм М., Маулер К.* Машинные методы математических вычислений. М.: Мир, 1980. 210 с.
15. *Дэйвисон М.* Многомерное шкалирование. Методы наглядного представления данных. М.: Финансы и статистика, 1988.