

УДК 541.6

РАСЧЕТ СТАБИЛЬНОСТИ КОМПЛЕКСОВ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ С β -ЦИКЛОДЕКСТРИНОМ МЕТОДОМ QSPR

Н.И. Жохова, Е.В. Бобков, И.И. Баскин, В.А. Палиолин, А.Н. Зефирова, Н.С. Зефирова

(кафедра органической химии; e-mail: zhokhova@org.chem.msu.ru)

Методом множественной линейной регрессии с использованием процедуры двойного скользящего контроля построена QSPR-модель для расчета изменения свободной энергии Гиббса при образовании комплексов органических соединений различных классов с β -циклодекстрином. Модель обладает хорошими прогнозирующими свойствами.

Разработка фармакологических форм с желаемыми фармакокинетическими и терапевтическими свойствами на основе включения молекул лекарственного вещества в полость циклодекстринов (ЦД) является перспективным направлением в решении проблемы транспорта лекарств к биологической мишени. Комплексообразование с ЦД препятствует биоразложению органического соединения, уменьшает его летучесть, улучшает растворимость и биодоступность, снижает риск нежелательных побочных эффектов [1–4]. Принципиальными преимуществами природных ЦД как носителей молекул лекарственных препаратов являются хорошо идентифицируемая структура, возможность химической модификации, низкая токсичность и фармакологическая активность [1]. Вследствие относительной доступности и дешевизны наибольшие перспективы для практического использования имеют β -ЦД и его производные.

Актуальность разработки надежного теоретического метода расчета и прогнозирования стабильности комплексов органических соединений с ЦД обусловлена, в частности, сложностями экспериментального определения соответствующих параметров из-за низкой растворимости “молекул-гостей” в водных растворах. Расчетные методы квантовой химии и молекулярной динамики имеют ряд ограничений, связанных с большими размерами и конформационной подвижностью молекул ЦД в водном растворе [5]. Альтернативу этим методам представляет методология QSPR (*Quantitative Structure-Property Relationships*), которая успешно используется для прогнозирования свойств химических соединений [6, 7]. Для расчета термодинамической стабильности комплексов органических соединений различных классов с ЦД ранее был предложен ряд QSPR-моделей [8–11]. Общий недостаток этих моделей связан с завышенной оценкой их прогнозирующих свойств. Традиционные приемы, применяемые с этой целью в большинстве ис-

следований, заключаются либо в использовании для независимого прогноза одного фиксированного набора соединений, либо в проведении после отбора дескрипторов процедуры обычного скользящего контроля, что часто приводит к завышению прогнозирующей способности моделей, поскольку информация из контрольной выборки может быть косвенным образом учтена при построении моделей. Этого недостатка позволяет избежать процедура двойного скользящего контроля. Ранее мы успешно использовали эту процедуру при построении моделей с помощью метода линейной регрессии и искусственных нейронных сетей для прогнозирования физико-химических свойств и биологической активности органических соединений [12]. В данной работе мы использовали метод множественной линейной регрессии и процедуру двойного скользящего контроля для расчета стабильности комплексов “гость-хозяин” органических соединений различных классов с β -циклодекстрином.

В работе использовали экспериментальные данные по изменению свободной энергии (ΔG , кДж/моль) при образовании комплексов 218 органических соединений (ароматические углеводороды, спирты, фенолы, простые и сложные эфиры, альдегиды, кетоны, кислоты, серосодержащие соединения, нитрилы, анилины, гетероциклические соединения, стероиды и барбитураты) с β -ЦД в воде при температуре 25°C [8]. Построение QSPR-моделей и расчет дескрипторов (физико-химических, топологических, зарядовых, стерических и др.) осуществляли с помощью программного комплекса NASAWIN, разработанного на химическом факультете МГУ [13–16]. Для генерации фрагментных дескрипторов использовали алгоритм, реализующий схему иерархической классификации структурных фрагментов и атомов [15]. Схема позволяет генерировать большое число наборов фрагментов химических структур и обеспечивает гибкую классификацию типов атомов. При построении многомерных (т.е.

включающих множество дескрипторов) линейных регрессионных моделей с использованием процедуры двойного скользящего контроля исходная база химических соединений разбивается на 3 части, обучающую, внутреннюю контрольную и внешнюю контрольную выборки в соотношении 3:1:1. Каждое соединение последовательно включается во все три выборки. Информация внутренней контрольной выборки используется для контроля и оптимизации прогнозирующих свойств моделей: процесс пошагового отбора дескрипторов и построение моделей останавливается по достижении наименьшей ошибки прогнозирования на этой выборке. Информация внешней контрольной выборки не используется при построении и отборе моделей и служит для независимой оценки прогнозирующей способности результирующей линейно-регрессионной модели. Предсказанное значение свойства для каждого соединения вычисляется как среднее из предсказанных значений при всех разбиениях, при которых оно попадает во внешнюю контрольную выборку. На выходе на основе усреднения статистических параметров множества (в данной работе 20) частных линейно-регрессионных моделей была получена многомерная модель, включающая: 1) информацию о статистических параметрах 20 линейно-регрессионных моделей и усредненной многомерной модели; 2) данные независимого прогноза свойства для каждого из соединений базы; 3) данные о типе и вкладе каждого дескриптора, включенного в многомерную модель. Такая методология обеспечивает корректную оценку прогнозирующей способности многомерной линейной модели.

Обсуждение результатов

На первом этапе работы для построения QSPR-моделей мы использовали дескрипторы, учитывающие число вхождений структурных фрагментов в химическую структуру "молекул-гостей". Ранее мы успешно использовали фрагментный подход для оценки физико-химических свойств органических соединений, в частности, хроматографических индексов удерживания [17], температуры кипения [17], температуры вспышки [18], магнитной восприимчивости [19], поляризуемости [20], плотности (для жидких веществ), вязкости и давления насыщенного пара органических соединений [16] и др. Достоинством фрагментных дескрипторов является прозрачный смысл каждого дескриптора и возможность их быстрой автоматической генерации на основе одной лишь структурной формулы соединения без учета информации о пространственном строении или электронной структуре моле-

кул [15]. С целью выбора фрагментов, наиболее адекватно описывающих геометрию и топологию исследуемых органических структур, мы рассчитывали число вхождений в структуру следующих фрагментов: цепочек длиной от 1 до 15 неводородных атомов; циклов размером от 3 до 15; разветвленных фрагментов, включающих от 4 до 6 неводородных атомов; бициклов, включающих от 6 до 15 атомов; трициклов, включающих от 12 до 15 атомов.

Наилучшими прогнозирующими характеристиками обладала многомерная модель, основанная на учете фрагментов размером от одного до четырех неводородных атомов:

для внешних контрольных выборок

параметр Q^2 для двойного

скользящего контроля

$$Q^2_{\text{прогн.}} = 0,704,$$

среднеквадратичное отклонение

$$RMSE_{\text{прогн.}} = 2,87 \text{ кДж/моль},$$

средняя абсолютная ошибка

$$MAE_{\text{прогн.}} = 1,98 \text{ кДж/моль},$$

для обучающей выборки

усредненный квадрат коэффициента корреляции

$$R^2_{\text{обуч.}} = 0,876,$$

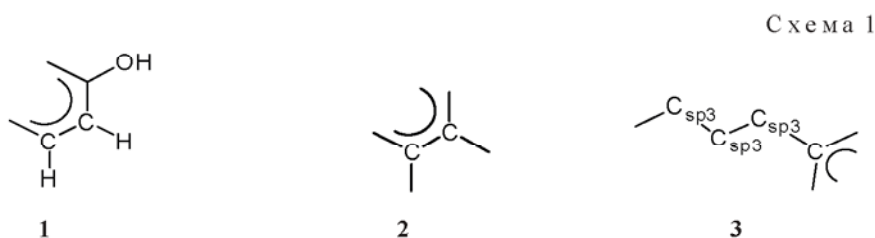
среднеквадратичное отклонение

$$RMSE_{\text{обуч.}} = 1,50 \text{ кДж/моль},$$

средняя абсолютная ошибка

$$MAE_{\text{обуч.}} = 1,11 \text{ кДж/моль}.$$

С целью повышения качества полученных нами моделей, основанных только на фрагментных дескрипторах, представляло интерес включение в эти модели других типов дескрипторов, которые могут охарактеризовать моделируемое свойство. Известно, что по условиям термодинамического равновесия гидрофобные молекулы органических соединений, способные по своим геометрическим размерам проникать во внутреннее пространство кольца ЦД, при добавлении в водный раствор ЦД образуют комплексы включения. Изменение свободной энергии связано с "выталкиванием" гидрофобной молекулы из водной среды и образованием дополнительных водородных связей между молекулами воды, находившимися в контакте с гидрофобной молекулой [1]. К факторам, оказывающим влияние на образование комплексов, относят, в частности, гидрофобные взаимодействия, образование водородных связей между полярными группами "гостя" и гидроксильными группами "хозяина", стерические затруднения, конформационные изменения аддукта ЦД-вода, изменение (высвобожде-



ние) свободной энергии в процессе включения субстрата в полость ЦД и вытеснения воды. Поэтому при дальнейшем построении моделей в комплексе с фрагментными дескрипторами мы использовали различные варианты наборов следующих дескрипторов: 1) липофильность (IgP), 2) дескрипторы водородных связей, характеризующие способность органического соединения выступать в качестве донора либо акцептора водородной связи, 3) дескрипторы, описывающие стерические факторы, характеризующие молекулярный объем и площадь молекулярной поверхности, 4) дескрипторы, характеризующие распределение зарядов внутри молекул, 5) дескрипторы, связанные с распределением свойств атомов внутри фрагментов, (цепочки от 1 до 5 атомов) и характеризующие число электронов, атомный радиус, электроотрицательность, потенциал ионизации и др. Было построено более 500 линейно-регрессионных моделей на основе комбинаций различных типов дескрипторов. Выбор дескрипторов в процессе построения многомерной модели осуществлялся по следующим критериям: улучшение коэффициента корреляции на прогнозе и высокий индекс вхождения в частные линейные модели. Наилучшие результаты были получены при добавлении к фрагментным дескрипторам параметра липофильности IgP и дескрипторов, характеризующих распределение свойств атомов внутри структурных фрагментов. Следует отметить, что параметр липофильности имеет наибольшую статистическую значимость практически для всех моделей, в построении которых он участвует. Это свидетельствует о важности учета гидрофобной составляющей свободной энергии взаимодействия органической молекулы с ЦД, что обусловлено наличием в последнем гидрофобной полости. Наименьший вклад в совокупности с фрагментными дескрипторами вносят дескрипторы, характеризующие стерические особенности “молекул-гостей” и дескрипторы водородных связей. В результате была получена многомерная модель, построенная на основе фрагментных дескрипторов, липофильности и дескрипторов, характеризующих свойства атомов фрагментов. Эта модель имеет следующие статистические характеристики:

для внешних контрольных выборок
параметр Q^2 для двойного скользящего контроля

$$Q^2_{\text{прогн.}} = 0,828,$$

среднеквадратичное отклонение

$$RMSE_{\text{прогн.}} = 2,16 \text{ кДж/моль},$$

средняя абсолютная ошибка

$$MAE_{\text{прогн.}} = 1,56 \text{ кДж/моль};$$

для обучающей выборки

усредненный квадрат коэффициента корреляции

$$R^2_{\text{обуч.}} = 0,936,$$

среднеквадратичное отклонение

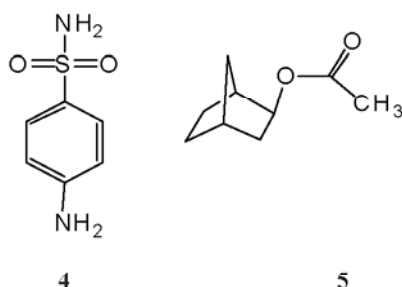
$$RMSE_{\text{обуч.}} = 1,19 \text{ кДж/моль},$$

средняя абсолютная ошибка

$$MAE_{\text{обуч.}} = 0,91 \text{ кДж/моль}.$$

Наибольшую значимость для данной комбинированной модели имеют следующие дескрипторы (по степени убывания вклада): липофильность, среднее значение произведений атомных радиусов в положениях 1–5 по всем 5-атомным цепочкам, число вхождение фрагментов типа **1**, **2** и **3** (схема 1), а также сумма разностей электроотрицательностей для всех связей X–H (где X – любой неводородный атом) в молекуле. Дескрипторы, характеризующие фрагментный состав “молекул-гостей”, необходимы для построения прогнозирующих моделей и позволяют компенсировать ограниченность других типов дескрипторов. Для данной модели наибольшие ошибки при независимом прогнозе исследуемого свойства (ΔG) имеют комплексы β -ЦД со следующими соединениями: стрептоцидом (разница между экспериментальной и предсказанной величинами $\Delta G_{\text{эксп.}} - \Delta G_{\text{расч.}} = -8,22$ кДж/моль) и 2-норборнилацетатом ($-10,21$ кДж/моль), соответственно структуры **4** и **5** (схема 2). Исключение из базы этих соединений не приводит к существенному изменению статистических характеристик модели. Ниже представлена диаграмма разброса данных независимого прогноза ΔG (кДж/моль) образования комплексов исследованных органических соединений с β -ЦД, полученных согласно этой модели, и экспериментальных данных (рисунок).

Схема 2



Таким образом, с использованием метода пошагового построения множественной линейной регрессии и процедуры двойного скользящего контроля нами была получена многомерная QSPR-модель на основе фрагментных дескрипторов, учитывающих число вхождений фрагментов размером до 4 неводородных атомов, а также параметра липофильности $\lg P$ и дескрипторов, характеризующих свойства атомов внутри фрагментов. Модель обладает высокой прогнозирующей способностью в отношении изменения свободной энергии ΔG об-

разования комплексов β -циклодекстрина с органическими соединениями различных классов. Использование процедуры двойного скользящего контроля при построении модели обеспечивает надежную и беспристрастную оценку ее прогнозирующих свойств.

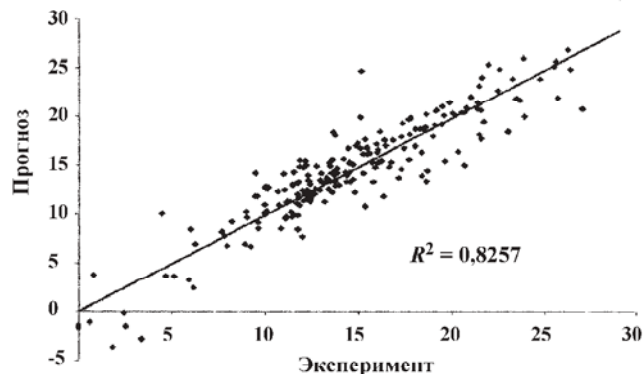


Диаграмма разброса данных независимого прогноза ΔG (кДж/моль) образования комплексов исследованных органических соединений с β -ЦД, полученных согласно модели, и экспериментальных данных

СПИСОК ЛИТЕРАТУРЫ

1. Uekama K., Hirayama F., Irie T. // Chem. Rev. 1998. **98**. P. 2045.
2. Астахова А.В., Демина Н.Б. // Хим.-фарм. журнал. 2004. **38**. № 2. С. 46.
3. Cyclodextrins and their industrial uses / Ed. by D. Duchene. Paris, 1987.
4. Cyclodextrins in pharmacy. Dordrecht, 1994.
5. Lipkowitz K.B. // Chem. Rev. 1998. **98**. P. 1829.
6. Hansch C., Leo A. Exploring QSAR. Fundamentals and applications in chemistry and biology. V.1. ACS Professional Ref. Book, 1995.
7. Zefirov N.S., Palyulin V. // J. Chem. Inf. Comput. Sci. 2002. **42**. P. 1112.
8. Katrizky A.R., Fara D.C., Yang H., Karelson M., Suzuki T., Solov'ev V.P., Varnek A. // J. Chem. Inf. Comput. Sci. 2004. **44**. P. 529.
9. Klein C.Th., Polheim D., Viernstein H., Wolschann P. // J. Incl. Phe. Macr. Chem. 2000. **36**. N 4. P. 409.
10. Suzuki T.A. // J. Chem. Inf. Comput. Sci. 2001. **41**. P. 1266.
11. Suzuki T., Ishida M., Fabian W. M. // J. Comput.-Aided Mol. Design. 2000. **14**. P. 669.
12. Baskin I.I., Zhokhova N.I., Palyulin V.A., Ivanova A.A., Zefirov A.N., Zefirov N.S. // XVI European Symposium on Quantitative Structure-Activity Relationships and Molecular Modeling. Mediterranean Sea, Italy. September 10-17, 2006. P. 206.
13. Baskin I.I., Palyulin V.A., Zefirov N.S. // QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications/Eds. F. Sanz, J. Giraldo, F. Manaut. Barcelona, 1995. P. 30
14. Baskin I.I., Palyulin V.A., Zefirov N.S. // J. Chem. Inf. Comput. Sci. 1997. **37**. P. 715.
15. Артеменко Н.В., Баскин И.И., Палюлин В.А., Зефирова Н.С. // Докл. РАН. 2001. **381**. N 2. С. 203.
16. Артеменко Н.В., Баскин И.И., Палюлин В.А., Зефирова Н.С. // Изв. РАН. Сер. хим. 2003. № 1. С. 19.
17. Zefirov N.S., Palyulin V.A. // J. Chem. Inf. Comput. Sci. 2001. **41**. P. 1022
18. Жохова Н.И., Баскин И.И., Палюлин В.А., Зефирова А.Н., Зефирова Н.С. // Изв. РАН. Сер. хим. 2003. № 9. С. 1787.
19. Жохова Н.И., Баскин И.И., Палюлин В.А., Зефирова А.Н., Зефирова Н.С. // Жур. структ. хим. 2004. **45**. N 4. С. 660.
20. Жохова Н.И., Баскин И.И., Палюлин В.А., Зефирова А.Н., Зефирова Н.С. // Изв. РАН. Сер. хим. 2003. № 5. С. 1005.

Поступила в редакцию 09.04.07

CALCULATION OF STABILITY OF β -CYCLODEXTRIN COMPLEXES OF ORGANIC COMPOUNDS USING QSPR APPROACH

N.I. Zhokhova, E.V. Bobkov, I.I. Baskin, V.A. Palyulin, A.N. Zefirov, N.S. Zefirov

(Division of Organic Chemistry)

The QSPR-model for predicting the free energies of complexation of organic compounds of various classes with β -cyclodextrin was constructed using multiple linear regression and double cross-validation procedure. The model possesses good predictive ability.