

МОДЕЛИРОВАНИЕ РАСТВОРИМОСТИ ОРГАНИЧЕСКИХ ВЕЩЕСТВ В СВЕРХКРИТИЧЕСКОМ ДИОКСИДЕ УГЛЕРОДА

Д.С. Комарова — Российский химико-технологический университет им. Д.И. Менделеева, кафедра химического и фармацевтического инжиниринга. г. Москва, Россия. ORCID: 0009-0001-1313-0141. Эл. почта: darya.komarowa2000@yandex.ru (*для переписки*)

И.В. Лебедев — Российский химико-технологический университет им. Д.И. Менделеева, кафедра химического и фармацевтического инжиниринга. г. Москва, Россия. ORCID: 0000-0001-6682-9813. Эл. почта: lebedev.i.v@muctr.ru

А.Р. Фунтикова — Российский химико-технологический университет им. Д.И. Менделеева. г. Москва, Россия. Эл. почта: ann.funtikova06@mail.ru

Н.В. Меньшутина — Российский химико-технологический университет им. Д.И. Менделеева, кафедра химического и фармацевтического инжиниринга. г. Москва, Россия. ORCID: 0000-0001-7806-1426. Эл. почта: chemcom@muctr.ru

© 2025 г. Поступила в редакцию 08.04.2025 г. Прошла рецензирование 15.05.2025 г.
Принята к публикации 02.06.2025 г.

Предложены математические модели растворимости различных групп органических веществ в сверхкритическом диоксиде углерода. Разработанные модели основаны на методе количественного соотношения «структура—свойство». Ключевая особенность метода — построение зависимости между прогнозируемым свойством и параметрами, численно отражающими особенности молекулярной структуры (молекулярными дескрипторами). В рамках работы был осуществлен отбор молекулярных дескрипторов, влияющих на растворимость и построены зависимости растворимости органических веществ в сверхкритическом диоксиде углерода (СК-СО₂) с использованием метода множественной линейной регрессии. Разработку моделей осуществляли с применением разработанного программно-аналитического комплекса и оригинальной базы данных, содержащей сведения о растворимости веществ в СК-СО₂ и информацию о молекулярных дескрипторах растворяемых веществ. Была проведена оценка прогнозирующей способности разработанных моделей. На основе полученных результатов сделан вывод о перспективности применения разработанных моделей для прогнозирования растворимости веществ в сверхкритическом диоксиде углерода.

Ключевые слова: молекулярный дескриптор, СК-СО₂, растворимость, метод QSPR, программно-аналитический комплекс, база данных.

ВВЕДЕНИЕ

Сверхкритический флюид (СКФ) — состояние, при котором вещество проявляет свойства, характерные как для газа, так и для жидкости. Вязкость и диффузия СКФ сопоставимы со значениями, характерными для газов, плотность и сольватирующие свойства сопоставимы со значениями, характерными для жидкостей [1, 2]. В настоящий момент наиболее широко применяемым СКФ на лабораторном и промышленном уровнях является сверхкритический диоксид углерода (СК-СО₂). Диоксид углерода характеризуется относительно низкими параметрами перехода в СК-состояние (7,4 МПа, 31,1 °С),

что является значительным преимуществом при работе с термолабильными веществами [1, 3]. Диоксид углерода не токсичен, не взрывоопасен, экологически безопасен [1–3]. Кроме того, он легко удаляется из конечных продуктов, что исключает необходимость в дополнительных этапах очистки [4].

В настоящее время в промышленности наблюдается растущая потребность в разработке эффективных и экологически безопасных методов извлечения, разделения и обработки веществ, в связи с чем процессы на основе СКФ находят все большее применение в химической, фармацевтической, пищевой отраслях [3–6]. СКФ, в частности, СК-СО₂, характеризуются высокой проникающей способностью, вследствие чего возможна глубокая экстракция целевых компонентов из растительного сырья [2]. СК-СО₂ обеспечивает высокую селективность извлечения веществ, что способствует получению продуктов высокой степени чистоты с заданными характеристиками при минимизации количества побочных продуктов [4].

С использованием СК-СО₂ возможно получение композиционных материалов посредством адсорбции растворенных в нем веществ на твердую матрицу-носитель [7]. Уникальные свойства СК-СО₂, в частности, высокая степень диффузии и высокая плотность, позволяют эффективно осуществлять процесс адсорбции. Технологии на основе СК-СО₂ могут быть использованы для микронизации посредством метода быстрого расширения СК-СО₂ с растворенным веществом, что приводит к образованию микро- или наночастиц [3]. Данная технология имеет высокий потенциал применения в области фармацевтики, поскольку с использованием метода быстрого расширения возможно получение частиц требуемого размера с увеличенной биодоступностью [8].

Для проведения всех перечисленных выше процессов необходимы данные о растворимости веществ в СК-СО₂, на основе которых можно оценить целесообразность и перспективность их проведения. Однако, зачастую исследователи сталкиваются с недостаточным количеством экспериментальных данных о растворимости веществ в СК-СО₂. Для определения растворимости вещества в СКФ (в том числе, в СК-СО₂) требуется проведение трудоемких и дорогостоящих экспериментальных исследований. Для прогнозирования растворимости веществ в СКФ активно применяют математическое моделирование.

Для моделирования процессов растворения в СКФ применяют модели, которые можно условно разделить на три типа. Первый тип основан на использовании параметров системы, включая критические параметры; в основу таких моделей положены уравнения состояния, такие как уравнения Ван-дер-Ваальса, Пенга—Робинсона, Соавва—Редлиха—Квонга и другие. Среди сложностей при применении данного типа моделей можно выделить недостаточную точность при прогнозировании растворимости, поскольку модели часто не учитывают сложные взаимодействия между молекулами в условиях СКФ, где условия значительно отличаются от тех, что наблюдаются в обычных жидкостях и газах [9].

Второй тип моделей включает в себя эмпирические модели, применяющиеся для расчета растворимости с использованием плотности СКФ. К данному типу моделей относятся: уравнение Джойбана, уравнение Кумара—Джонстона, уравнение Кешмири и др. [8]. Коэффициенты уравнения модели, определенные на основе экспериментальных данных, носят индивидуальный характер. Данные модели могут иметь недостаточную точность, поскольку, как и в

случае моделей первого типа, при моделировании не учитывается взаимодействие растворяемого вещества с СКФ и полученные результаты могут значительно отклоняться от данных, полученных экспериментально.

Модели третьего типа полуэмпирические; представляют собой комбинацию двух предыдущих типов [9]. Данные модели описывают растворимость веществ, используя параметры растворителя, растворенных веществ и системы. Среди полуэмпирических моделей наиболее широко применяются следующие [8].

Модель Крастила (1) [10]:

$$\ln(S) = k \ln(\rho) + \frac{a}{T} + b, \quad (1)$$

где S — растворимость, кг/м³; ρ — плотность СКФ при параметрах процесса, кг/м³; T — температура, К; k — коэффициент ассоциации; a , b — индивидуальные константы, не зависящие от температуры и определяемые по методу линейной регрессии.

Модель де Валле и Агилеры (2) [11]:

$$\ln(S) = k \ln(\rho) + \frac{a}{T} + \frac{b}{T^2 + c}, \quad (2)$$

где S — растворимость, кг/м³; ρ — плотность сверхкритического флюида при параметрах процесса, кг/м³; T — температура, К; k , a , b , c — индивидуальные константы, определяемые по методу линейной регрессии.

Модель Бартла (3) [10]:

$$\ln\left(\frac{S}{P_{\text{реф}}}\right) = k(\rho - \rho_{\text{реф}}) + \frac{a}{T} + b, \quad (3)$$

где S — растворимость, кг/м³; P — давление, МПа; $P_{\text{реф}}$ — референсное давление (0,1 МПа); ρ — плотность сверхкритического флюида при исследуемых температуре и давлении, кг/м³; $\rho_{\text{реф}}$ — референсная плотность СКФ (700 кг/м³); T — температура, К; k , a , b — индивидуальные константы, определяемые по методу линейной регрессии.

Однако, несмотря на все преимущества полуэмпирических моделей, они обладают существенными ограничениями при прогнозировании растворимости. Это обусловлено тем, что растворение вещества в СКФ представляет собой сложный физико-химический процесс, на который оказывают влияние структурные особенности молекулы растворяемого вещества. Полуэмпирические модели не способны учитывать параметры молекулярной структуры, поэтому прогнозирование растворимости ограничено одним конкретным веществом, для которого и находят подходящее полуэмпирическое уравнение [12]. Кроме того, при разработке математических моделей растворимости веществ в СКФ остро стоит проблема уменьшения количества необходимых экспериментальных данных.

Для решения указанных проблем в данной работе для прогнозирования растворимости предлагается метод количественного соотношения «структура—свойство» (quantitative structure—property relationship, QSPR). Метод QSPR

представляет собой инструмент моделирования, применяемый в хемоинформатике [13]. На основе данных о молекулярных структурах веществ, выраженных в числовой форме, и данных о целевом свойстве веществ разрабатывается модель, которая позволяет спрогнозировать значение целевого свойства вещества с использованием данных о его молекулярной структуре [14].

Параметры, численно отражающие особенности молекулярной структуры, называются молекулярными дескрипторами. К настоящему времени существует несколько тысяч молекулярных дескрипторов, применяющихся для описания уникальных особенностей структуры широкого спектра веществ (дескрипторы, отражающие молекулярную массу, количество атомов, количество связей, количество ароматических связей и др.) [13].

Метод QSPR позволяет представить зависимость между целевым свойством и параметрами молекулярной структуры вещества в форме математических уравнений различного вида, но наиболее универсальным и простым для расчетов является вид линейной зависимости [15]. В данном случае общий вид QSPR-модели может быть записан следующим образом (4):

$$f(x) = a_0 + a_1x_1 + a_2x_2 + \dots + a_ix_i, \quad (4)$$

где a_0 — свободный коэффициент; a_i — коэффициенты при независимых параметрах x_i , которые представляют дескрипторы. В ходе разработки QSPR-модели осуществляется определение коэффициентов a_0 , a_i , для чего могут применяться различные методы, включая множественную линейную регрессию.

Создание QSPR-модели включает несколько этапов. На начальном этапе осуществляется сбор экспериментальных данных о свойствах соединений и их структурах [16]. На следующем этапе все имеющиеся экспериментальные данные разделяют на выборки — обучающую и тестовую. Далее осуществляют разработку математической модели на основе обучающей выборки, осуществляют подбор коэффициентов уравнения. На следующем этапе разработанную модель проверяют на основе данных тестовой выборки для оценки ее предсказательной способности [17].

Перспективность применения метода QSPR при разработке моделей растворимости органических веществ в СК-СО₂ подкрепляется многочисленными работами, в которых подтверждается высокая прогностическая способность QSPR-моделей для прогнозирования физико-химических свойств, токсичности, биологической активности и других свойств веществ [15, 17, 18].

Цель настоящей работы — разработка QSPR-моделей растворимости органических веществ в СК-СО₂ и оценка прогнозирующей способности полученных моделей.

МЕТОДИЧЕСКАЯ ЧАСТЬ

В данной работе для осуществления математического моделирования растворимости органических веществ в СК-СО₂ использовали базу данных, разработанную Гуриковым П.А. [19]. Используемая база данных содержит в себе информацию о растворимости веществ в СК-СО₂, которая представлена в виде натурального логарифма мольной доли растворяемого вещества. В дополнение к этим данным в базе данных также представлены дескрипторы для каждого вещества.

При разработке QSPR-моделей были использованы экспериментальные данные по растворимости (S , г/кг) в СК-СО₂ следующих веществ:

— $S < 0,01$ г/кг: сульфаниламид, сульфатиазол, гидрокортизон, бетаметазон, кофейная кислота, протокатехиновая кислота;

— $0,01$ г/кг $< S < 0,5$ г/кг: метилгаллат, сульфаметазин, сульфадиметоксин, пиазин-2-карбоновая кислота, сульфамеразин, резерпин, ацетаминофен, ацетат кортизона, сирингиновая кислота, теofilлин, хлорамфеникол, пироксикам, пенициллин G, метимазол;

— $0,5$ г/кг $< S < 1$ г/кг: карбамазепин, *n*-ацетоксиацетанилид, тестостерон, тестостерон-17-пропионат, новокаин, феназопиридин;

— 1 г/кг $< S < 10$ г/кг: аспирин, миконазол, 1,3-дигидроксibenзол, ципротерона ацетат, капсаицин, 1-гидрокси-2-метилантрахинон, метилпарабен, бисакодил, прогестерон, 2,4-дихлор-1-нафтол, диазепам;

— 10 г/кг $< S < 50$ г/кг: миндальная кислота, линолевая кислота, индол, артемизинин, салициловая кислота, ретинол, бензойная кислота, тимол.

Использование экспериментальных данных веществ с различной растворимостью в СК-СО₂ в широком диапазоне при разработке QSPR-моделей необходимо для обучения этих моделей. Разнообразие веществ в каждой группе позволяет снизить риск переобученности QSPR-моделей. Модели, разработанные на основе экспериментальных данных веществ с различным строением молекул, являются более универсальными и могут быть применимы для прогнозирования растворимости веществ с различной молекулярной структурой. Однако, стоит учитывать, что применение большого разнообразия структур молекул при разработке QSPR-моделей может приводить к низкой прогностической способности модели в связи с появлением в модели большого количества «выбросов» и сложных взаимосвязей, что затрудняет обнаружение четких закономерностей и может ухудшить прогностические способности модели. Все выбранные вещества проявляют биологическую активность и имеют потенциал применения в фармацевтической и химической промышленности.

Общее количество экспериментальных данных, использованных при разработке QSPR-моделей растворимости веществ в СК-СО₂, составило 1500. Все использованные экспериментальные данные были разделены на две выборки: обучающую (80% данных) и тестовую (20% данных) [17]. Разделение проводили в случайном порядке.

Для разработки QSPR-моделей были отобраны дескрипторы, отражающие структурные характеристики молекул, которые могут оказывать влияние на растворимость веществ в СК-СО₂ [9, 20, 21]. Количество рассматриваемых молекулярных дескрипторов в представленной работе равно 20 (табл. 1). Значения дескрипторов для всех веществ были получены из базы данных MOLE db [22].

Для обозначения дескрипторов в табл. 1 и далее используют общепринятые для них аббревиатуры [23].

В связи с наличием дескрипторов, описывающих схожие молекулярные характеристики, существует значительный риск возникновения мультиколлинеарности — линейной зависимости между двумя и более независимыми дескрипторами. Мультиколлинеарность может негативно сказаться на качестве итоговых QSPR-моделей, приводя к избыточности коэффициентов в уравнении (4) и затрудняя интерпретацию результатов. Необходимо отметить, что

Таблица 1

Список дескрипторов, используемых при разработке QSPR-моделей

| Обозначение | Название | Физико-химическое значение | Обоснование |
|-------------|--|---|--|
| MW | Молекулярная масса | Молекулярная масса | Высокая молекулярная масса может быть косвенным показателем большого размера молекулы и сложности молекулярной структуры, данные факторы могут оказывать влияние на взаимодействие молекулы с СК-СО ₂ |
| Sv | Сумма атомных объемов Ван-дер-Ваальса | Объем молекулы, учитывающий размер электронной оболочки и взаимодействия между атомами | Связан с возможностью взаимодействия молекулы с СК-СО ₂ |
| Se | Сумма атомных электроотрицательностей по Сандерсону | Электроотрицательность атомов — молекулярная мера способности атома к притяжению электронов | Атомная электроотрицательность связана с полярностью молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |
| Ss | Сумма электроположительных состояний Кира—Холла | Электроположительное состояние атомов — оценка активности атомов молекулы с учетом их положения и связей внутри структуры | Электроположительное состояние связано с полярностью молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |
| Ms | Усредненное электроположительное состояние | | |
| MAXDP | Максимальная положительная электроположительная вариация | Электроположительная вариация — изменение состояния атома под влиянием других атомов молекулы, их положения и связей внутри структуры | Электроположительная вариация связана с полярностью молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |
| W | Индекс Винера | Сумма топологических расстояний между всеми парами атомов | Отражает структурную сложность молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |
| Nar | Индекс Харари | Сумма обратных топологических расстояний между всеми парами атомов в молекуле | Отражает структурную сложность молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |

Продолжение таблицы 1

| Обозначение | Название | Физико-химическое значение | Обоснование |
|-------------|---|--|---|
| War | Индекс Винера (для всех путей) | Версия индекса Винера, которая учитывает топологические расстояния между всеми парами атомов, включая все возможные пути | Отражает структурную сложность молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |
| Pol | Число полярности | Представляет собой количество пар вершин графа, разделенных тремя ребрами | Отражает структурную сложность и гибкость молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |
| SOK | Индекс симметрии Кира | Учитывает топологическое распределение атомов и связей в молекуле | Симметричные молекулы могут иметь более однородное распределение заряда, что влияет на растворимость в СК-СО ₂ |
| UNIP | Униполярность | Характеризует распределение заряда в молекуле, указывая на преобладание одного типа полярности | Униполярность связана с полярностью молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |
| XMOD | Модифицированный индекс связности Рандича | Оценивает уровень структурной связности внутри молекулы | Отражает структурную сложность молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |
| ISIZ | Молекулярный размер | Молекулярный размер | Размер молекулы может оказывать влияние на взаимодействие молекулы с СК-СО ₂ |
| QXXe | Взвешенное значение электроотрицательности (по оси X) | Электронная плотность вдоль оси X | Электроотрицательность связана с полярностью молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |
| QYYe | Взвешенное значение электроотрицательности (по оси Y) | Электронная плотность вдоль оси Y | |
| QZZe | Взвешенное значение электроотрицательности (по оси Z) | Электронная плотность вдоль оси Z | |

Окончание таблицы 1

| Обозначение | Название | Физико-химическое значение | Обоснование |
|-------------|--|---|---|
| ASP | Асферичность | Мера отклонения формы молекулы от сферичности | Высокая асферичность может оказывать на более сложную молекулярную структуру, что оказывает влияние на растворимость в СК-СО ₂ |
| TPSA(Tot) | Полярная площадь поверхности | Полярная площадь поверхности, рассчитанная на основе полярных коэффициентов атомов N, O, S, P | Полярная площадь поверхности связана с полярностью молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |
| MLOGP2 | Коэффициент распределения «октанол—вода» | Оценка гидрофобности молекул | Гидрофобность связана с полярностью молекулы, которая оказывает влияние на растворимость в СК-СО ₂ |

Таблица 2

Элемент корреляционной матрицы

| | MW | Sv | Se | Ss | W | Har | QZZe | TPSA (Tot) | Ms | MAXDP | S0K |
|------------|-------|-------|-------|-------|-------|-------|-------|------------|--------|-------|-------|
| MW | 1,000 | 0,971 | 0,956 | 0,927 | 0,987 | 0,984 | 0,933 | 0,422 | -0,437 | 0,684 | 0,609 |
| Sv | 0,971 | 1,000 | 0,994 | 0,845 | 0,976 | 0,968 | 0,953 | 0,254 | -0,594 | 0,663 | 0,581 |
| Se | 0,956 | 0,994 | 1,000 | 0,831 | 0,965 | 0,950 | 0,952 | 0,257 | -0,587 | 0,668 | 0,553 |
| Ss | 0,927 | 0,845 | 0,831 | 1,000 | 0,924 | 0,915 | 0,841 | 0,631 | -0,085 | 0,737 | 0,602 |
| W | 0,987 | 0,976 | 0,965 | 0,924 | 1,000 | 0,969 | 0,967 | 0,407 | -0,440 | 0,671 | 0,621 |
| Har | 0,984 | 0,968 | 0,950 | 0,915 | 0,969 | 1,000 | 0,904 | 0,382 | -0,460 | 0,714 | 0,603 |
| Wap | 0,934 | 0,934 | 0,912 | 0,842 | 0,905 | 0,975 | 0,832 | 0,299 | -0,516 | 0,676 | 0,532 |
| XMOD | 0,997 | 0,962 | 0,944 | 0,931 | 0,987 | 0,982 | 0,934 | 0,446 | -0,423 | 0,679 | 0,622 |
| QXXe | 0,940 | 0,957 | 0,956 | 0,827 | 0,942 | 0,929 | 0,889 | 0,266 | -0,537 | 0,652 | 0,591 |
| QYYe | 0,922 | 0,950 | 0,951 | 0,823 | 0,956 | 0,893 | 0,993 | 0,317 | -0,522 | 0,604 | 0,573 |
| QZZe | 0,933 | 0,953 | 0,952 | 0,841 | 0,967 | 0,904 | 1,000 | 0,331 | -0,508 | 0,600 | 0,589 |
| TPSA (Tot) | 0,422 | 0,254 | 0,257 | 0,631 | 0,407 | 0,382 | 0,331 | 1,000 | 0,440 | 0,283 | 0,273 |

расчет моделей с большим количеством дескрипторов приводит к значительному увеличению требуемых вычислительных мощностей.

Для выявления и исключения мультиколлинеарности была построена корреляционная матрица, учитывающая значения дескрипторов для всех рассматриваемых веществ, которая позволила оценить степень корреляции между парами дескрипторов (табл. 2).

В качестве критерия мультиколлинеарности использовали порог коэффициента корреляции $|r|=0,8$ [24]. Пары дескрипторов с $|r| < 0,8$ считались некоррелирующими и могли совместно использоваться в модели. Кроме применения некоррелирующих дескрипторов (TPSA(Tot), MLOGP2, MAXDP, Ms, S0K, ASP, ISIZ), также было принято решение о включении в QSPR-модель дескриптора, проявляющего сильную взаимную корреляцию с всеми остальными дескрипторами. Для отбора данного дескриптора были вычислены коэффициенты корреляции между неотобранными дескрипторами и значениями растворимости. Расчеты проводились с применением экспериментальных данных обучающей выборки. На основании полученных результатов дескрипторы были отсортированы по степени корреляции с растворимостью, был отобран дескриптор, проявляющий наибольшую степень корреляции с растворимостью (Ss), ввиду его высокой значимости в описании целевого свойства. Таким образом, для дальнейшего моделирования были отобраны следующие дескрипторы: TPSA(Tot), MLOGP2, MAXDP, Ms, S0K, ASP, ISIZ, Ss.

Конечный вид QSPR-модели для каждой группы веществ основан на полуэмпирическом уравнении Крастила, которое было модернизировано путем добавления к нему молекулярных дескрипторов (5):

$$\ln(S) = k \ln(\rho) + \frac{a}{T} + \sum_{i=1}^n a_i x_i + b, \quad (5)$$

где S — растворимость, кг/м³; ρ — плотность СК-СО₂, кг/м³; T — температура, К; k, a, a_i, b — искомые коэффициенты; x_i — значения молекулярных дескрипторов.

В качестве математического инструмента для определения коэффициентов k, a, a_i, b уравнения (5) была выбрана множественная линейная регрессия, позволяющая аппроксимировать зависимость растворимости от совокупности молекулярных дескрипторов. Подбор коэффициентов при разработке QSPR-модели для каждой из групп органических веществ осуществляли методом наименьших квадратов на обучающей выборке экспериментальных данных. Оценку точности моделирования производили на основании коэффициента детерминации (R^2) и средней абсолютной ошибки (mean absolute error, MAE). Коэффициент R^2 характеризует долю объясненной дисперсии экспериментальных данных и указывает на степень согласия между предсказанными и фактическими значениями; принимает значения от 0 до 1. MAE отражает среднее значение абсолютных ошибок между предсказанными и фактическими значениями, выражается в тех же единицах, что и целевая переменная. Полученные QSPR-модели для каждой группы веществ, представлены в следующем разделе.

С целью автоматизации обработки большого объема данных расчет проводили с применением разработанного программно-аналитического комплекса на языке Python. Все этапы предварительной обработки данных и расчетов были выполнены с его использованием. При разработке программного комплекса были задействованы следующие библиотеки: Pandas, Numpy, matplotlib, scipy.stats и sklearn.metrics.

РЕЗУЛЬТАТЫ

На основе выбранных дескрипторов были разработаны QSPR-модели растворимости для каждой группы рассматриваемых веществ (табл. 3).

Таблица 3

Разработанные QSPR-модели растворимости веществ в СК-СО₂

| Группа веществ | QSPR-модель |
|------------------------------|---|
| $S < 0,01$ г/кг | $\ln y = 1,554 + 5,375 \ln \rho - \frac{7540,2}{T} - 19,879 \ln(\text{TPSA}(\text{Tot})) + 7,072 \ln(Ss) -$ $-2,433 \ln(\text{MLOGP2}) - 3,951 \ln(\text{MAXDP}) + 12,014 \ln(\text{Ms}) +$ $+0,161 \ln(\text{S0K}) - 11,676 \ln(\text{ASP}) + 5,606 \ln(\text{ISIZ}) \quad (6)$ |
| $0,01$ г/кг $< S < 0,5$ г/кг | $\ln y = 3,594 + 2,823 \ln \rho - 3129 \cdot \frac{3}{T} - 2,473 \ln(\text{TPSA}(\text{Tot})) + 9,836 \ln(Ss) -$ $-0,049 \ln(\text{MLOGP2}) - 3,527 \ln(\text{MAXDP}) - 10,270 \ln(\text{Ms}) +$ $+0,183 \ln(\text{S0K}) + 0,216 \ln(\text{ASP}) - 4,646 \ln(\text{ISIZ}). \quad (7)$ |
| $0,5$ г/кг $< S < 1$ г/кг | $\ln y = 3,459 + 3,720 \ln \rho - \frac{5026,4}{T} + 4,338 \ln(\text{TPSA}(\text{Tot})) - 0,474 \ln(Ss) +$ $+3,055 \ln(\text{MLOGP2}) + 5,065 \ln(\text{MAXDP}) - 2,751 \ln(\text{Ms}) -$ $-2,527 \ln(\text{S0K}) + 4,006 \ln(\text{ASP}) - 2,717 \ln(\text{ISIZ}). \quad (8)$ |

| Группа веществ | QSPR-модель |
|----------------------------|--|
| 1 г/кг < S < < 10 г/кг | $\ln y = -4,867 + 3,411 \ln \rho - \frac{3583,6}{T} - 2,817 \ln(\text{TPSA}(\text{Tot})) - 2,305 \ln(\text{Ss}) -$ $-0,330 \ln(\text{MLOGP2}) + 2,073 \ln(\text{MAXDP}) + 7,137 \ln(\text{Ms}) + 0,125 \ln(\text{S0K}) -$ $-0,326 \ln(\text{ASP}) + 1,635 \ln(\text{ISIZ}). \quad (9)$ |
| 10 г/кг < S < < 50 г/кг | $\ln y = -0,0251 + 3,221 \ln \rho - \frac{4702,3}{T} - 0,375 \ln(\text{TPSA}(\text{Tot})) - 0,327 \ln(\text{Ss}) -$ $-0,507 \ln(\text{MLOGP2}) - 1,331 \ln(\text{MAXDP}) - 0,830 \ln(\text{Ms}) + 0,388 \ln(\text{S0K}) -$ $-0,868 \ln(\text{ASP}) + 0,985 \ln(\text{ISIZ}). \quad (10)$ |

Для каждой группы веществ были получены графики, отображающие отклонение спрогнозированных значений растворимости веществ с использованием, разработанных QSPR-моделей, от значений растворимости, полученных экспериментально (рис. 1).

Следующим этапом была валидация полученных моделей. Разработанные QSPR-модели были использованы для определения растворимости для каждой группы веществ на основе тестовой выборки экспериментальных данных (рис. 2).

Разработанная QSPR-модель для группы веществ 1 г/кг < S < 10 г/кг была использована для определения растворимости эритромицина; данное вещество не входило в обучающую и тестовую выборки (рис. 3).

Для каждой группы веществ были рассчитаны коэффициенты R² и MAE на основании полученных результатов по тестовым выборкам экспериментальных данных. Анализ точности разработанных QSPR-моделей представлен в следующем разделе.

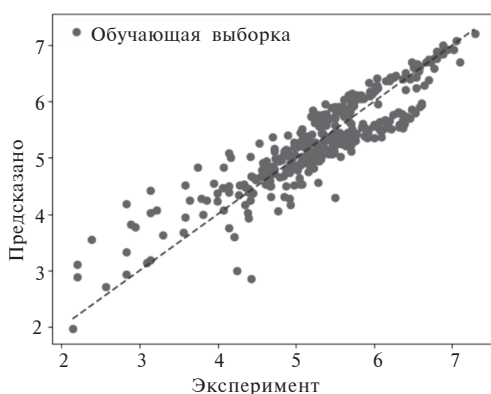


Рис. 1. График «измерено—предсказано» по обучающей выборке экспериментальных данных для группы веществ 1 г/кг < S < 10 г/кг

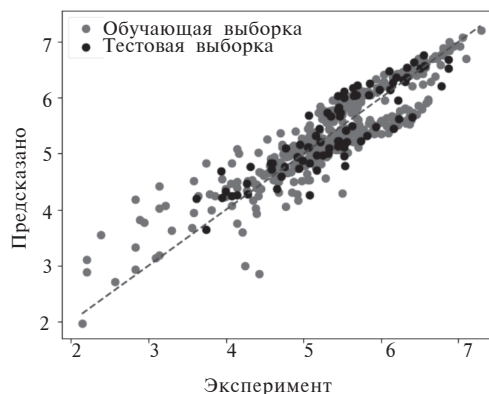


Рис. 2. График «измерено—предсказано» по обучающей и тестовой выборкам экспериментальных данных для группы веществ 1 г/кг < S < 10 г/кг

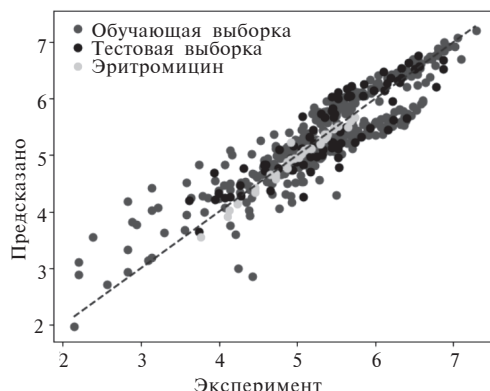


Рис. 3. График «измерено—предсказано» по обучающей и тестовой выборкам экспериментальных данных для группы веществ $1 \text{ г/кг} < S < 10 \text{ г/кг}$ и экспериментальных данных эритромицина

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Разработанные QSPR-модели продемонстрировали высокую точность прогнозирования растворимости:

- $S < 0,01 \text{ г/кг}$, $R^2 = 0,83$, $\text{MAE} = 0,65$ (диапазон экспериментальных значений $\ln(S)$ от $-7,13$ до $0,95$);
- $0,01 \text{ г/кг} < S < 0,5 \text{ г/кг}$, $R^2 = 0,75$, $\text{MAE} = 0,91$ (диапазон экспериментальных значений $\ln(S)$ от $-6,91$ до $5,25$);
- $0,5 \text{ г/кг} < S < 1 \text{ г/кг}$, $R^2 = 0,83$, $\text{MAE} = 0,35$ (диапазон экспериментальных значений $\ln(S)$ от $-0,94$ до $5,31$);
- $1 \text{ г/кг} < S < 10 \text{ г/кг}$, $R^2 = 0,85$, $\text{MAE} = 0,28$ (диапазон экспериментальных значений $\ln(S)$ от $2,14$ до $7,29$);
- $10 \text{ г/кг} < S < 50 \text{ г/кг}$, $R^2 = 0,92$, $\text{MAE} = 0,26$ (диапазон экспериментальных значений $\ln(S)$ от $1,95$ до $9,46$).

Значение $R^2 > 0,8$ было получено для диапазонов самых низких и средних концентраций (уравнения (6), (8—10)). Значения $R^2 > 0,8$ для большинства случаев указывают на высокую прогностическую способность полученных QSPR-моделей, следовательно, разработанные модели могут быть использованы для прогнозирования растворимости веществ в СК- CO_2 . Минимальное значение $R^2 = 0,75$ наблюдалось для группы веществ $0,01 \text{ г/кг} < S < 0,5 \text{ г/кг}$ (уравнение (7)), где, вероятно, сказывается наибольшее структурное разнообразие веществ. Наиболее перспективными для дальнейшего применения являются QSPR-модели, разработанные для групп $0,5 \text{ г/кг} < S < 1 \text{ г/кг}$, $1 \text{ г/кг} < S < 10 \text{ г/кг}$, $10 \text{ г/кг} < S < 50 \text{ г/кг}$ поскольку данные модели демонстрируют $R^2 < 0,8$ и низкие значения MAE, относительно диапазона экспериментальных значений $\ln(S)$ для данных групп. Наибольшее значение MAE наблюдается для QSPR-модели группы $0,01 \text{ г/кг} < S < 0,5 \text{ г/кг}$, что коррелирует с полученным наименьшим коэффициентом R^2 для данной группы веществ. QSPR-модель для группы $S < 0,01 \text{ г/кг}$ также демонстрирует высокое значение MAE, но в то же время R^2 соответствует $0,83$, что может говорить о высокой прогностической способности модели, несмотря на наличие выбросов в экспериментальных данных.

Полученную QSPR-модель для группы $1 \text{ г/кг} < S < 10 \text{ г/кг}$ использовали для прогнозирования растворимости вещества эритромицина, не включенного в первоначальную выборку для разработки моделей. С использованием данной QSPR-модели был достигнут коэффициент R^2 , равный 0,95, значение MAE составило 0,13, что также демонстрирует высокую прогностическую способность полученной модели.

При анализе полученных QSPR-моделей можно сделать следующие выводы. В большинстве моделей коэффициент при дескрипторе ASP (асферичность) отрицательный, следовательно, можно выдвинуть предположение, что сильно несферичные, продолговатые молекулы хуже растворяются в СК-СО₂. Аналогично, в большинстве разработанных QSPR-моделей наблюдаются отрицательные коэффициенты при дескрипторе Ss (сумма электропотологических состояний Кира—Холла), который связан с электротопологическим состоянием атомов в молекуле. На основании данного факта, можно сделать предположение, что более сложная структура молекулы, с большим взаимным электротопологическим влиянием атомов друг на друга затрудняет растворение в СК-СО₂.

Коэффициент при дескрипторе MLOGP2 (коэффициент распределения «октанол — вода», модифицированный log P для большинства моделей отрицательный. Этот дескриптор отражает липофильность соединения с учетом полярности и гибкости молекулы. СК-СО₂, хотя и считается неполярным, имеет уникальные свойства, которые позволяют ему взаимодействовать с некоторыми полярными молекулами, вещества с высокой гидрофобностью (высоким MLOGP2) могут иметь низкую растворимость в СК-СО₂.

Для более детальных выводов необходимо проведение дальнейших исследований в области разработки QSPR-моделей растворимости органических веществ в СК-СО₂, в том числе, требуется значительное увеличение количества рассматриваемых органических веществ и экспериментальных точек, для увеличения прогнозирующей способности математических моделей.

ВЫВОДЫ

В рамках работы были разработаны QSPR-модели для прогнозирования растворимости некоторых групп органических веществ в СК-СО₂. Была осуществлена предварительная кластеризация исследуемых веществ, в результате которой они были распределены по группам по растворимости. Была осуществлена предварительная выборка молекулярных дескрипторов, которые оказывают влияние на растворимость вещества в СК-СО₂ и не коррелируют друг с другом. Всего исследуемая выборка включала в себя данные по 20-ти молекулярным дескрипторам и 1500-м экспериментальным точкам для 45-ти рассматриваемых веществ, распределенных по пяти группам.

Для каждой группы была разработана QSPR-модель зависимости растворимости веществ, входящих в группу, от их молекулярной структуры, температуры, плотности сверхкритического диоксида углерода. Была проведена оценка точности прогнозирования разработанных моделей с использованием тестовой выборки экспериментальных данных. Полученные значения R^2 (от 0,75 до 0,92) и MAE (от 0,26 до 0,91) позволяют сделать вывод о высокой прогностической способности моделей и подтверждают перспективность применения разработанных моделей в химической, фармацевтической, пищевой и других

промышленностях. Был разработан программно-аналитический комплекс, который позволил автоматизировать ключевые этапы разработки QSPR-моделей, включая предобработку данных, отбор дескрипторов, поиск коэффициентов QSPR-модели и расчет коэффициента детерминации. Такой подход обеспечил высокую воспроизводимость вычислений и возможность масштабирования методологии для обработки больших массивов данных.

Предложенный алгоритм отбора дескрипторов и разработки математических моделей на основе программно-аналитического комплекса может быть адаптирован к другим сверхкритическим флюидам, что делает данный подход универсальным и применимым для широкого круга задач в области хемоинформатики и материаловедения. Предложенный подход в дальнейшем позволит прогнозировать растворимость веществ в сверхкритическом диоксиде углерода по их молекулярным структурам и существенно уменьшить количество необходимых экспериментальных исследований по растворимости этих веществ. Полученные QSPR-модели могут быть использованы для снижения затрат времени и ресурсов при разработке технологических процессов с использованием сверхкритических флюидов.

СПИСОК ЛИТЕРАТУРЫ

1. Tsygankov P.Yu., Menshutina N.V. // Theor Found Chem Eng. 2024. Vol. 58. No 5. P. 1648.
2. Artemiev A.I., Demkin K.M., Kazeev I.V., Menshutina N.V. // Biochem. Moscow Suppl. Ser. B. 2024. Vol. 18. No. 4. P. 323.
3. Tsygankov P.Y., Kislinskaya A.Y., Pashkin E.A., Menshutina N.V. // Theor Found Chem Eng. 2024. Vol. 58. No 5. P. 1584.
4. Wang W., Rao L., Wu X., Wang Y., Zhao L., Liao X. // Food Eng Rev., 2021. Vol. 13. No 3. P. 570.
5. Arumugham T., Rambabu K., Hasan S.W., Show P.L., Rinklebe J., Banat F. // Chemosphere. 2021. Vol. 271. P. 129525.
6. Liu H., Chen B.Q., Pan Y.J., Fu C.P., Kankala R.K., Wang S.B., Chen A.Z. // Science and Technology of Advanced Materials. Taylor & Francis. 2021. Vol. 22. No 1. P. 695.
7. Lovskaya D., Bezchasnyuk A., Mochalova M., Tsygankov P., Lebedev A., Zorkina Y., Zubkov E., Ochneva A., Gurina O., Silantyev A., Majouga A., Menshutina N. // Gels. Multidisciplinary Digital Publishing Institute. 2022. Vol. 8. No 12. P. 765.
8. Sodejftian G., Usefi M.M.B. // ChemBioEng Reviews. 2023. Vol. 10. No 2. P. 133.
9. Balbinot Filho C.A., Dias J.L., Rebelatto E.A., Lanza M. // Food Eng Rev. 2023. Vol. 15. No 3. P. 466.
10. Hojjati M., Yamini Y., Khajeh M., Vatanara A. // J. Supercritical Fluids. 2007. Vol. 41. No 2. P. 187.
11. Garcia-Samino C., del Valle E.M., Taberner A. // J. Molecular Liquids. 2024. Vol. 403. P. 124890.
12. Škerget M., Knez Ž., Knez-Hrnčič M. // J. Chem. Eng. Data. Am. Chem. Soc. 2011. Vol. 56. No 4. P. 694.
13. Niazi S.K., Mariam Z. // Intern.J.Molecular Scienc es.Multidisciplinary Digital Publishing Institute. 2023. Vol. 24. No 14. P. 11488.
14. De P., Kar S., Ambure P., Roy K. // Arch Toxicol. 2022. Vol. 96. No 5. P. 1279.
15. Mahboob A., Rasheed M.W., Dhiaa A.M., Hanif I., Amin L. // Heliyon. Elsevier. 2024. Vol. 10. No 4. P. e25908.
16. Wu Z., Zhu M., Kang Y., Leung E.L.H., Lei T., Shen C., Jiang D., Wang Z., Cao D., Hou T. // Briefings in Bioinformatics. 2021. Vol. 22. No. 4. P. bbaa321.
17. Rácz A., Bajusz D., Héberger K. // Molecules. Multidisciplinary Digital Publishing Institute. 2021. Vol. 26. No 4. P. 1111.
18. Zhang X., Saif M.J., Idrees N., Kanwal S., Parveen S., Saeed F. // ACS Omega. Am. Chem. Soc. 2023. Vol. 8. No 44. P. 41417.

19. Gurikov P., Lebedev I., Kolnoochenko A., Menshutina N. // Computer Aided Chem. Engineering / Elsevier. 2016. Vol. 38. P. 1587.
 20. Rangel Pinto J.D., Guerrero J.L., Rivera L., Parada-Pinilla M.P., Cala M.P., López G., González Barrios A.F. // Front. Chem. Frontiers. 2024. Vol. 12. P. 1480887.
 21. Makarov D.M., Kalikin N.N., Budkov Y.A. // Ind. Eng. Chem. Res. Am. Chem. Soc. 2024. Vol. 63. No 3. P. 1589.
 22. http://michem.disat.unimib.it/mole_db/
 23. Handbook of molecular descriptors / Todeschini R., Consonni V. John Wiley & Sons, 2008.
 23. He J., Li J., Liu C., Wang C., Zhang Y., Wen C., Xue D., Cao J., Su Y., Qiao L., Bai Y. // Acta Materialia. 2021. Vol. 209. P. 116815.
-
-

MODELING OF SOLUBILITY OF ORGANIC SUBSTANCES IN SUPERCRITICAL CARBON DIOXIDE

D.S. Komarova — D. Mendeleev University of Chemical Technology of Russia, department of Chemical and Pharmaceutical Engineering, Moscow, Russia. ORCID: 0009-0001-1313-0141. E-mail: darya.komarowa2000@yandex.ru (*for correspondence*)

I.V. Lebedev — D. Mendeleev University of Chemical Technology of Russia, department of Chemical and Pharmaceutical Engineering, Moscow, Russia. ORCID: 0000-0001-6682-9813. E-mail: lebedev.i.v@muctr.ru

A.R. Funtikova — D. Mendeleev University of Chemical Technology of Russia, Moscow, Russia. E-mail: ann.funtikova06@mail.ru

N.V. Menshutina — D. Mendeleev University of Chemical Technology of Russia, department of Chemical and Pharmaceutical Engineering, Moscow, Russia. ORCID: 0000-0001-7806-1426. E-mail: chemcom@muctr.ru

Mathematical models of the solubility of various groups of organic substances in supercritical carbon dioxide are proposed. The developed models are based on the quantitative structure-property ratio method. The key feature of the method is to obtain the relationship between the predicted property and the structure of the substance molecule (molecular descriptors). As part of the work, molecular descriptors affecting solubility were selected and dependences of the solubility of organic substances in supercritical carbon dioxide were obtained using the multiple linear regression method. The models were developed using a developed software and analytical complex and an original database containing information on the solubility of substances in supercritical carbon dioxide and information on the molecular descriptors of soluble substances. The predictive ability of the developed models was evaluated using training and test samples of experimental data on the solubility of substances in supercritical carbon dioxide at various pressure and temperature parameters. Based on the results obtained, it was concluded that the developed models are promising for predicting the solubility of substances in supercritical carbon dioxide.

Key words: molecular descriptor, supercritical carbon dioxide, solubility, method QSPR, software and analytical complex, database.
