Оценка и моделирование риска химической опасности

УДК: 544.165:615.917 DOI: 10.25514/CHS.2020.2.18001

QSPR анализ способности индивидуальных авермектинов к биоконцентрированию

О. В. Тиньков[1,2], В. Ю. Григорьев[3], Л. Д. Григорьева[4]

[1]Государственное образовательное учреждение высшего профессионального образования Военный институт Министерства обороны, г. Тирасполь, Приднестровье, Молдова, e-mail: oleg.tinkov.chem@mail.ru
[2]Государственное образовательное учреждение Приднестровский государственный университет им. Т.Г. Шевченко, г. Тирасполь, Приднестровье, Молдова
[3]Федеральное государственное бюджетное учреждение науки Институт физиологически активных веществ Российской академии наук, Черноголовка, Московская область, Россия
[4]Федеральное государственное бюджетное образовательное учреждение высшего образования Московский государственный университет имени М.В. Ломоносова, Москва, Россия

Поступила в редакцию: 01.10.2020 г., после доработки: 27.10.2020 г., принята в печать: 10.11.2020

Аннотация – В настоящий момент распространение коронавируса представляет собой глобальную проблему для человечества. Одним из перспективных соединений для борьбы с коронавирусом SARS-CoV-2 является ивермектин – комплекс полусинтетических производных природных авермектинов, которые уже много лет эффективно используются в медицине, ветеринарии и сельском хозяйстве как противопаразитарные препараты. Также ведутся успешные исследования по использованию авермектинов в качестве противоопухолевых препаратов. Однако, несмотря на распространенность во всем мире, высокую физиологическую значимость и перспективность, до сих пор отсутствуют достоверные данные для большинства индивидуальных авермектинов по основным экотоксикологическим характеристикам, что в соответствии с действующим законодательством Европейского союза является обязательным условием для веществ, производимых или импортируемых в количестве свыше одной тонны в течение года. С использованием интернет-платформы «OCHEM» (https://ochem.eu) построены адекватные модели Количественной Связи Структура – Свойство (Quantitative Structure – Property Relationship, QSPR), позволившие оценить влияние структуры органических соединений на величину биоконцентрационного фактора (BCF). QSPR модели разработаны с использованием различных фрагментарных, топологических, физико-химических молекулярных дескрипторов, а также таких методов машинного обучения как Случайный Лес (Random Forest, RF) и Ассоциативные Нейронные Сети (Associative Neural Networks, ASNN). Для количественной оценки способности индивидуальных авермектинов к биоконцентрированию была разработана консенсусная QSPR модель, находящаяся в свободном доступе в сети Интернет по адресу: https://ochem.eu/model/20673575. При сравнительном QSPR моделировании были применены молекулярные дескрипторы, в том числе фрактальные, рассчитанные с помощью программы HYBOT. Разработанные модели обладают сопоставимой предсказательной способностью и могут быть полезны при определении стратегии синтеза, испытаний новых лекарственных средств на основе индивидуальных авермектинов, в том числе при селекции соединений – лидеров,

ингибирующих репликацию SARS-CoV-2. При проведении структурной интерпретации методом молекулярных пар определены наиболее часто встречающиеся молекулярные трансформации, повышающие и понижающие биоконцентрационный фактор, что может быть учтено при рациональном молекулярном дизайне новых физиологически активных соединений.

Ключевые слова: макролиды, авермектины, биоконцентрационный фактор, молекулярные дескрипторы, машинное обучение.

Chemical hazard assessment and risk modeling

UDC: 544.165:615.917 DOI: 10.25514/CHS.2020.2.18001

QSPR analysis of bioconcentrating ability of individual avermectins

Oleg V. Tinkov[1,2], Veniamin Yu. Grigoriev[3], and Lyudmila D. Grigorieva[4]

[1]Military Institute of the Ministry of Defense, Tiraspol, Transdniestria, Moldova,
e-mail: oleg.tinkov.chem@mail.ru
[2]Transnistrian State University, Tiraspol, Transdniestria, Moldova
[3]Institute of Physiologically Active Compounds of the Russian Academy of Science,
Chernogolovka, Moscow Region, Russia
[4]Lomonosov Moscow State University, Moscow, Russia

Received: October 1, 2020, Revised: October 27, 2020, Accepted: November 10, 2020

Abstract – At the moment, the spread of coronavirus infection is a global problem for humanity. One of the promising compounds for the fight against SARS-CoV-2 coronavirus is ivermectin – a complex of semi-synthetic derivatives of natural avermectins, which have been effectively used in medicine, veterinary medicine, and agriculture as antiparasitic agents for many years. There are also several successful studies on the use of avermectins as antitumor drugs. However, despite the worldwide recognition, high physiological significance and prospects, there are still no reliable data for most individual avermectins on their main ecotoxicological characteristics, which in accordance with the current legislation of the European Union is a prerequisite for substances produced or imported over one ton per year. Using the OCHEM web-platform (https://ochem.eu) adequate models of Quantitative Structure – Property Relationship (QSPR) were constructed, which allowed us to evaluate the influence of the structure of organic compounds on the value of the bioconcentration factor (BCF). QSPR models were developed using various fragmentary, topological, physico-chemical molecular descriptors, as well as machine learning methods such as Random Forest (RF) and Associative Neural Networks (ASNN). To quantify the ability of individual avermectins to bioconcentrate, a consensus QSPR model has been developed, which is freely available on the Internet at: https://ochem.eu/model/20673575. In comparative QSPR modeling, molecular descriptors were used, including fractal ones, calculated using the HYBOT program. The developed models have a comparable predictive ability and can be useful in determining the strategy of synthesis, testing of new drugs based on individual avermectins, including the selection of lead compounds that are potent in inhibition of SARS-CoV-2 replication. When conducting structural interpretation by the method of molecular pairs, the most common molecular transformations that increase and decrease the bioconcentration factor have been determined, which can be taken into account in the rational molecular design of new physiologically active compounds.

Keywords: macrolides, avermectins, bioconcentration factor, molecular descriptors, machine learning.

ВВЕДЕНИЕ

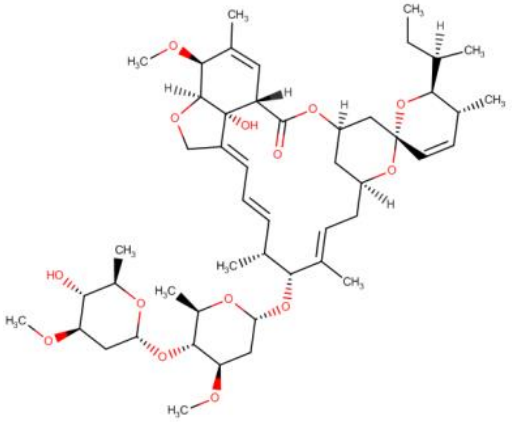
На сегодняшний день одним из наиболее перспективных ингибиторов коронавируса SARS-CoV-2 является ивермектин, который в ходе исследований *in vitro* позволил снизить долю вирусной РНК на 99,98% в течение 48 ч при концентрации 5 мкМ, которая безопасна для клеток [1]. Ранее ивермектин также снижал репликацию вирусов ВИЧ-1 и Денге [2]. Инвермектин представляет собой комплекс ивермектина B1a (≈90%) и ивермектина B1b (≈10%), которые являются полусинтетическими производными природных авермектинов.

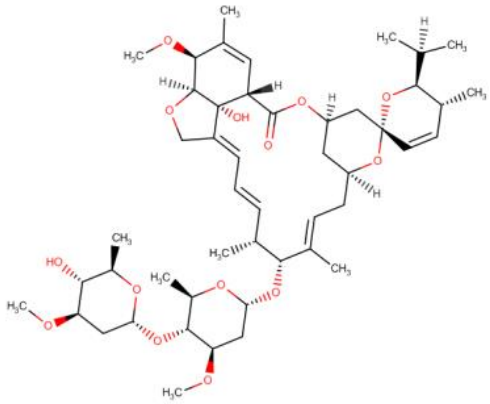
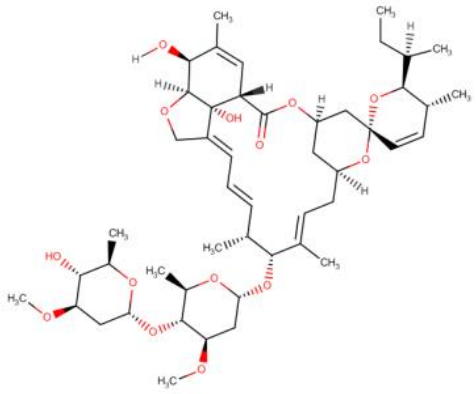
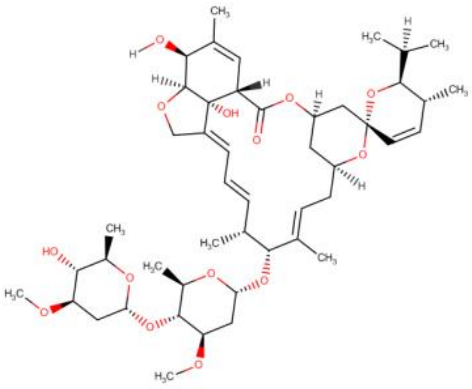
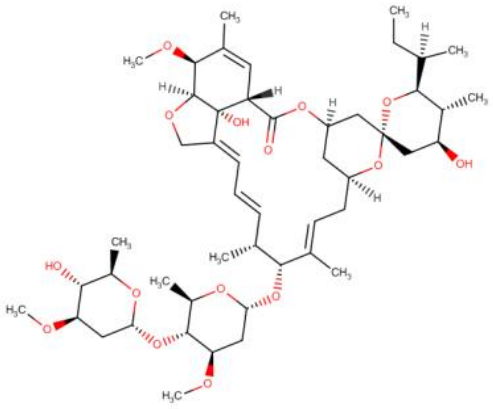
Природные авермектины, продуцируемые культурой *Streptomyces avermitilis*, а также их полусинтетические производные, в том числе ивермектин, чрезвычайно эффективно используются против эндо- и эктопаразитов растений, животных и человека [3]. Высокая важность авермектинов для человечества была подтверждена вручением в 2015 году половины Нобелевской премии по физиологии и медицине Сатоси Омури и Уильяму Кэмпбеллу за их новаторские исследования, направленные на создание нового класса лекарств [4]. Авермектины также продемонстрировали противоопухолевую активность [5].

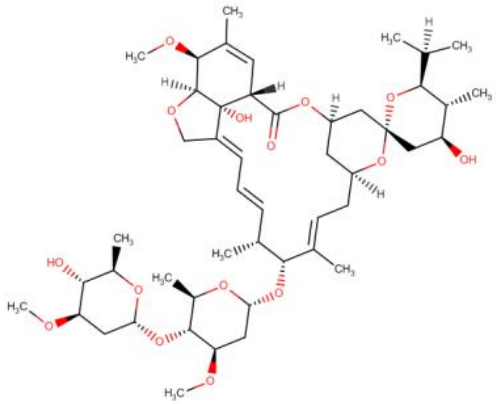
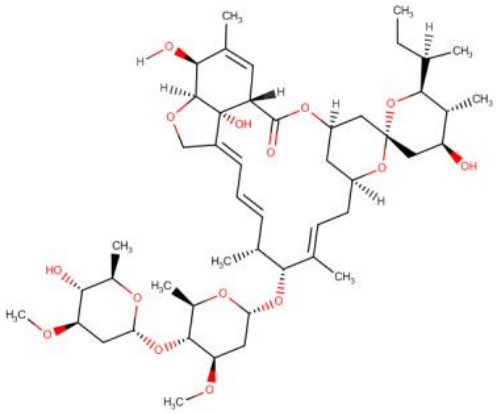
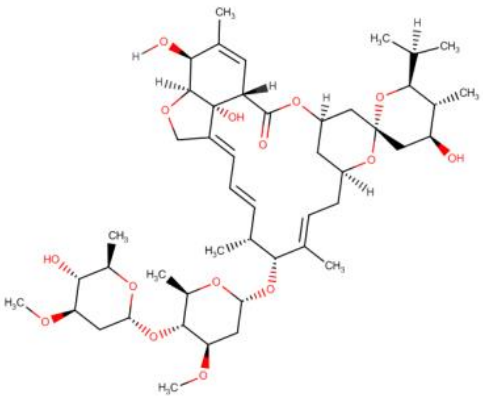
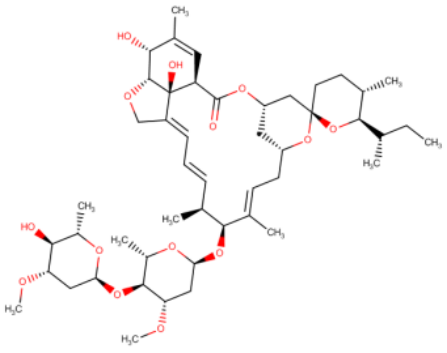
Природные авермектины, которые представляют собой комплекс из четырех основных (A1a, A2a, B1a, B2a) и четырех второстепенных (A1b, A2b, B1b, B2b) 16-членных макролидов, продуцируются культурой *Streptomyces avermitilis*. Химические структуры природных и полусинтетических авермектинов представлены в таблице 1.

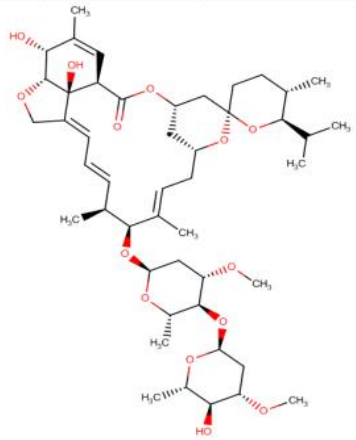
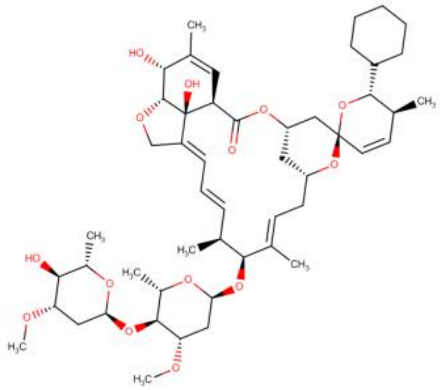
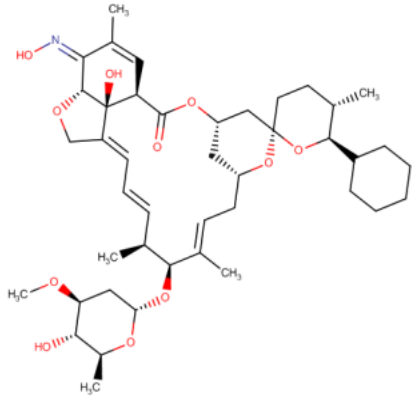
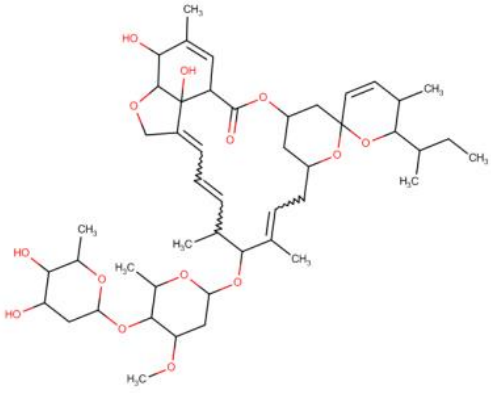
Таблица 1. Индивидуальные авермектины

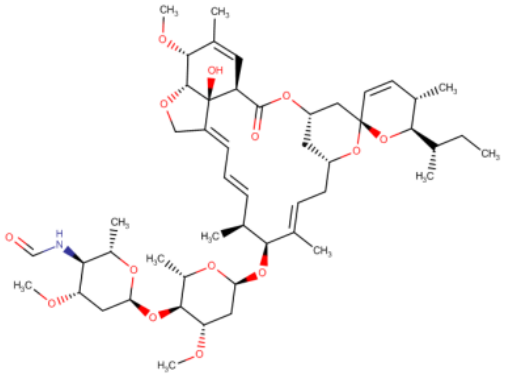
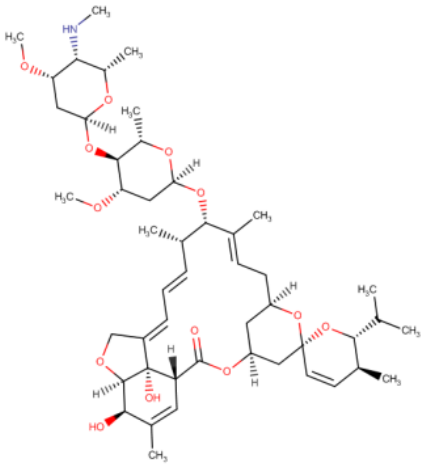
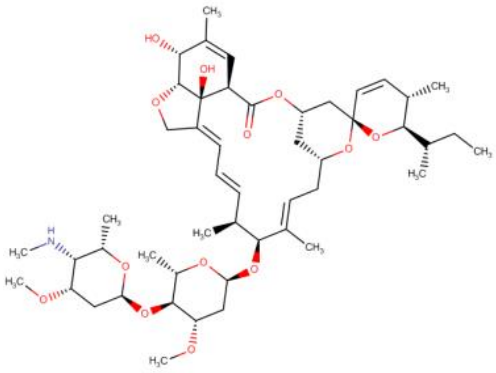
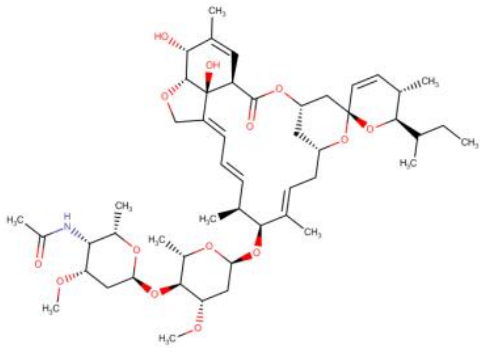
Table 1. Individual avermectins

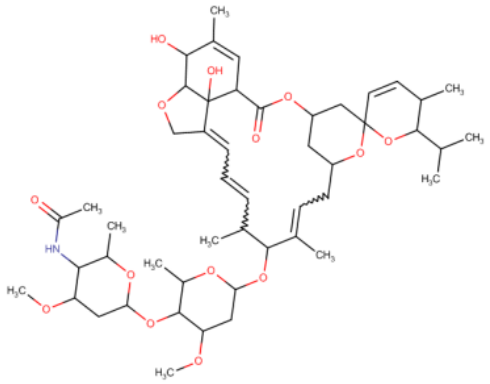
Название	Структура	номер CAS
Природные авермектины		
Авермектин A1a		65195-51-9

Название	Структура	номер CAS
Природные авермектины		
Авермектин A1b		65195-52-0
Авермектин B1a		65195-55-3
Авермектин B1b		65195-56-4
Авермектин A2a		65195-53-1

Название	Структура	номер CAS
Природные авермектины		
Авермектин A2b		65195-54-2
Авермектин B2a		65195-57-5
Авермектин B2b		65195-58-6
Полусинтетические авермектины		
Ивермектин B1a		71827-03-7

Название	Структура	номер CAS
Полусинтетические авермектины		
Ивермектин B1b		70209-81-3
Дорамектин		117704-25-3
Селамектин		220119-17-5
3"-O- Десметилавермектин B1a		99965-56-7

Название	Структура	номер CAS
Полусинтетические авермектины		
4"-(ацетиламино)-5-О-диметил-4"-диокси-Авермектин A1a		172926-97-5
Эмамектин B1B		121424-52-0
Эмамектин B1A		121124-29-6
Эприномектин B1a		133305-88-1

Название	Структура	номер CAS
Полусинтетические авермектины		
Эприномектин B1b		133305-89-2

Для всесторонней оценки возможности использования отдельных авермектинов в качестве лекарственных веществ чрезвычайно важна информация об их экотоксикологических характеристиках. В соответствии с Европейским регламентом о регистрации, оценке, разрешении и ограничении химических веществ (REACH) [6], перед началом производства и торговли требуется обязательная предварительная оценка различных экотоксикологических характеристик химических веществ, включая их способность к биоконцентрированию, которая оценивается с использованием коэффициента (фактора) биоконцентрации (BioConcentration Factor, BCF), представляющим собой отношение концентрации химического вещества в организме или биоте к его концентрации в воде.

В обзорной публикации [7] по экологическим свойствам авермектинов отмечается, что, несмотря на распространенность этих веществ во всем мире и их эффективное использование, к сожалению, экспериментальных данных по экотоксикологической оценке этих соединений крайне мало.

На сегодняшний день в связи со значительными финансовыми и временными затратами при оценке экспериментальных показателей химических веществ, а также с учетом этического отношения к животным, создаются математические модели «Количественная Связь Структура – Активность/Свойство» (QSAR/QSPR) [8–10], которые также поддерживаются REACH.

Принимая во внимание актуальность оценки BCF, было предложено значительное количество удовлетворительных моделей QSPR, некоторые из которых интегрированы в известные программы, например, T.E.S.T.v.4.2. [11], CompTox Chemistry Dashboard [12]. Предлагаемые модели QSPR с учетом области применимости (applicability domain, AD) позволяют надежно прогнозировать BCF для большого количества органических соединений. Область применимости модели это химическое, структурное, молекулярное и/или связанное с биологическим механизмом действия пространство данных, описываемое моделью [10]. Если новые молекулы из тестовой выборки находятся за пределами области применимости модели, то прогноз с

использованием соответствующей модели QSPR для них будет менее надежным. Определение AD является обязательным условием при проведении QSAR/QSPR моделирования.

К сожалению, природные и полусинтетические авермектины при прогнозировании BCF не входят в области применимости моделей, реализованных в указанных выше программных продуктах, что, вероятно, связано с незначительным структурным разнообразием обучающих выборок, типом используемых дескрипторов и методами их отбора.

Необходимость расширения структурного разнообразия обучающих выборок при прогнозировании BCF отмечена в публикации [13], где проводится подробный анализ существующих QSPR моделей BCF.

Учитывая отсутствие экспериментальных значений BCF для отдельных авермектинов, а также моделей QSPR, способных надежно прогнозировать эту характеристику с учетом областей применимости, целью данного исследования является количественная оценка коэффициента биоконцентрации индивидуальных авермектинов.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Свободно доступный набор экспериментальных значений BCF (1333 соединения) был экспортирован из вышеупомянутой публикации [13], в которой подробно описаны процедуры проверки исходной информации и разделения общей выборки на обучающую (1129 соединений) и тестовую (204 соединения) выборки. Экспериментальные значения BCF, приведенные авторами исследования [13] в л/кг, были преобразованы в десятичный логарифм этой величины.

QSPR моделирование проводилось с использованием интернет-ресурса «OCHEM» [14], который содержит значительное количество данных о структурах химических соединений и различных видах биологической активности, а также позволяет разрабатывать, интерпретировать QSPR/QSAR модели с использованием различных методов машинного обучения. В настоящей работе при моделировании с помощью интернет-ресурса «OCHEM» использовались различные типы дескрипторов [15], среди которых можно выделить целочисленные фрагментные симплексные дескрипторы, топологические индексы. Так, топологические дескрипторы (индексы), рассчитываемые с помощью встроенного в «OCHEM» модуля «Dragon» [15], связаны с применением теории графов при описании молекулярной структуры. При расчете симплексных дескрипторов (Simplex representation of molecular structure, SIRMS) с помощью модуля «SIRMS» [16] молекула рассматривается как система различных симплексов – четырехатомных молекулярных фрагментов фиксированной структуры (рис. 1). Дескриптором в данном случае является количество симплексов определенного вида. На 2D уровне атомы (вершины симплекса) дифференцируют не только по природе «метке» атома, но и с учетом различных физико-химических свойств – частичного заряда на атоме, липофильности, рефракции, способности выступать донором или акцептором водорода при образовании водородной связи.

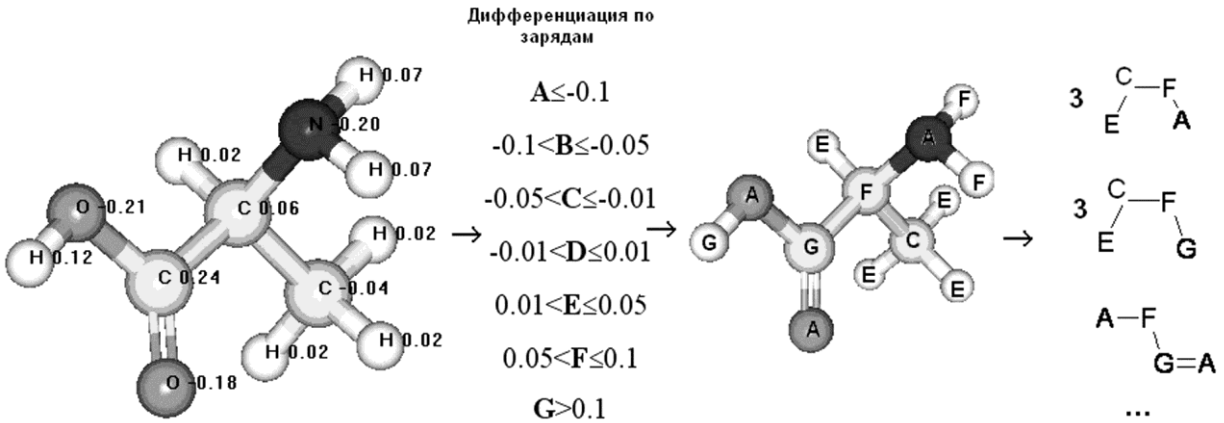


Рис. 1. Пример генерации симплексных дескрипторов для аланина на 2D уровне при использовании дифференциации атомов по их зарядам с помощью модуля SIRMS.

Fig. 1. Representative generation of simplex descriptors for alanine at 2D level using differentiation of atoms by their charges in accordance with SIRMS module.

Помимо этого, при моделировании с помощью «OCHEM» использовали встроенные модули «PyDescriptor» [17], «RDKIT» [18], «alvaDesc» [19], генерирующие наборы разнообразных топологических, фрагментарных, а также физико-химических, фармакофорных дескрипторов. Для построения моделей использовались методы Случайного Леса (Random Forest, RF) [20] и Ассоциативных Нейронных Сетей (Associative Neural Networks, ASNN) [21]. Построение консенсусной модели осуществлялось путем усреднения прогнозов отдельных моделей. Область применимости оценивалась с использованием концепции «расстояние до модели», в частности, подхода «CLASS-LAG». Для внутренней проверки использовалась процедура пятикратной кросс-валидации (5-folds). Для этого все соединения обучающей выборки случайным образом разделяются на 5 частей. Затем на 4 частях данных, объединенных в обучающую выборку, производится построение (обучение) QSPR модели, а оставшаяся часть данных используется как внутренняя тестовая выборка, т.е. на соединениях этой группы проверяется устойчивость и частично предсказательная способность модели. Данная процедура повторяется 5 раз, в результате чего каждая из 5 частей данных последовательно используется для тестирования. Важно подчеркнуть, что исследуемые соединения никогда не являются одновременно частью как обучающей, так и внутренней тестовой выборки. Более подробное описание использованных методов, дескрипторов, а также дополнительные ссылки на оригинальные работы даны в руководстве пользователя «OCHEM» [14].

Оценку точности и прогностической способности предложенных в работе моделей и их сравнение с другими QSPR моделями выполняли на основе коэффициента детерминации (Q^2) и среднеквадратичной ошибки (RMSE).

Учитывая, что вышеупомянутое QSPR моделирование потребовало привлечения большого количества дескрипторов (сотни и тысячи), для сравнения были сконструированы с помощью метода RF регрессионные модели, содержащие небольшое количество переменных, включая дескрипторы

HYBOT [22], топологические и фрактальные характеристики [23]. В основе HYBOT дескрипторов лежит оценка H-донорной и H-акцепторной способности молекул. В этом случае оценка AD проводилась интервальным методом. Дескрипторы с низкой вариабельностью и высоким коэффициентов взаимной корреляции (выше 0,9) были исключены при моделировании.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Результаты QSPR моделирования величин BCF приведены в таблице 2. Консенсусная модель находится в свободном доступе по ссылке, указанной в таблице 2. Все построенные модели имеют удовлетворительные статистические характеристики и демонстрируют предсказательную способность сопоставимую с существующими моделями QSPR, разработанными для исследуемой выборки [13].

Таблица 2. Статистические характеристики QSPR моделей BCF

Table 2. Statistical characteristics of QSPR BCF models

Метод	Дескрипторы	Выборка	Q^2	RMSE
Ассоциативные Нейронные Сети	Dragon	обучающая	0,70	0,74
		тестовая	0,74	0,80
	PyDescriptor	обучающая	0,69	0,76
		тестовая	0,76	0,78
	RDKIT	обучающая	0,67	0,78
		тестовая	0,74	0,80
Случайный Лес	alvaDesc	обучающая	0,78	0,64
		обучающая	0,80	0,70
	SIRMS	обучающая	0,72	0,72
		тестовая	0,78	0,74
	PyDescriptor	обучающая	0,75	0,68
		тестовая	0,79	0,73
	HYBOT	обучающая	0,67	0,78
		тестовая	0,71	0,85
Консенсусная модель (https://ochem,eu/model/20673575)		**обучающая**	0,77	0,65
		тестовая	0,80	0,70

Примечание: Q^2 – коэффициент детерминации; RMSE – среднеквадратичная ошибка

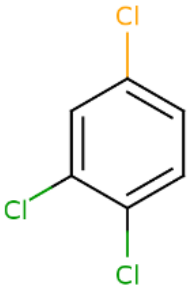
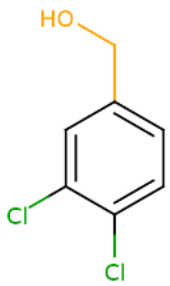
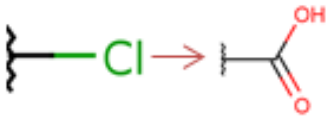
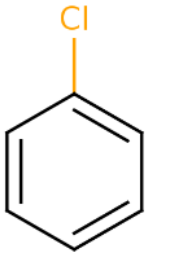
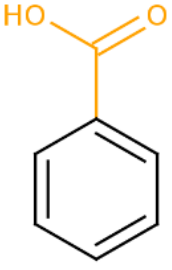
Для полученной консенсусной модели была проведена структурная интерпретация с помощью метода молекулярных пар, также реализованного в «OCHEM» [14]. Результаты интерпретации, в частности, молекулярные трансформации, которые в наибольшей степени увеличивают или уменьшают значения BCF для исследуемых 1129 соединений обучающей выборки приведены в таблице 3. Полный список молекулярных трансформаций, а также

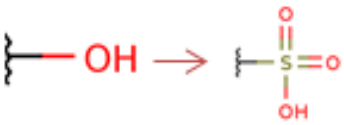
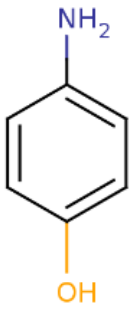
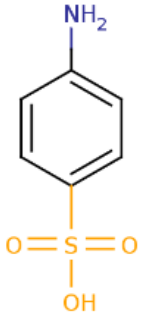
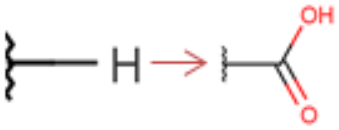
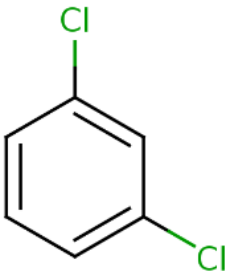
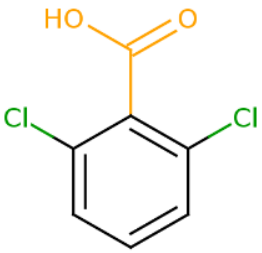
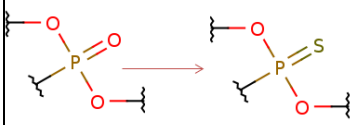
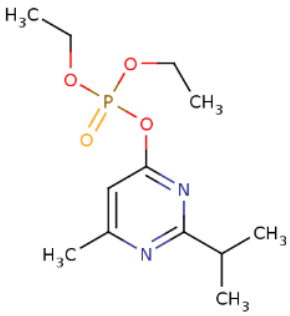
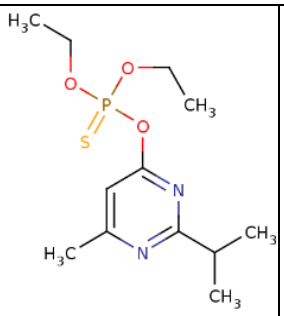
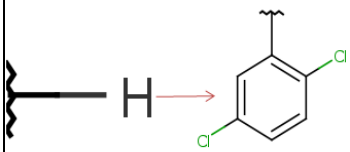
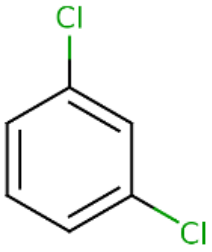
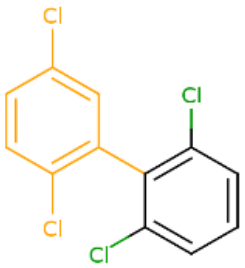
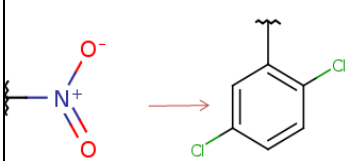
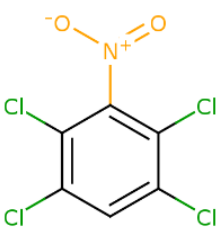
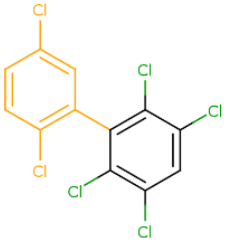
структуры соединений обучающей и тестовой выборок доступны по ссылке для консенсусной модели, приведенной в таблице 2.

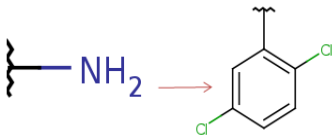
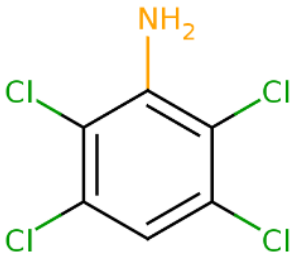
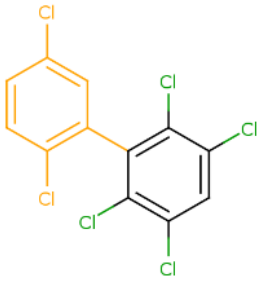
Как видно из таблицы 3, наибольшее снижение способности органических соединений к биоконцентрированию наблюдается при замещении атомов хлора гидроксильными и карбоксильными фрагментами. Также BCF уменьшается при замене гидроксильной группы на сульфогруппу $–SO_3H$ или водорода на карбоксильную группу. Замена фосфорной группы на тиофосфорную, а также аминогруппы, водорода, нитрогруппы на фрагмент, содержащий 2,6-дихлорбензол, напротив, значительно увеличивает BCF.

Таблица 3. Результаты структурной интерпретации консенсусной QSPR модели. Молекулярные трасформации, приводящие к повышению или понижению BCF для 1129 соединений обучающей выборки

Table 3. Results of structural interpretation of the consensus QSPR model. Molecular trasformations that lead to an increase or decrease in BCF for 1129 compounds of training set

Молекулярная трансформация	Пример пар соединений из обучающей выборки и их экспериментальные значения BCF, липофильности	N	Δmean
Молекулярные трасформации, **приводящие к понижению** BCF			
	 lgBCF = 3,04 lgP = 4,1  lgBCF = 0,45 lgP = 2,74	4	-1,7 ± 1,2
	 lgBCF = 2,43 lgP = 3,43  lgBCF = 1 lgP = 2,05	9	-1,6 ± 1,2

	 lgBCF = 1,62 lgP = 0,04	 lgBCF = 0,04 lgP = -2,16	4	-1,3 ± 0,66
	 lgBCF = 2,57 lgP = 3,53	 lgBCF = -0,15 lgP = 2,23	14	-1,1 ± 1,1
Молекулярные трасформации, **приводящие к повышению** BCF				
	 lgBCF = -1 lgP = 1,69	 lgBCF = 1,76 lgP = 3,3	5	2,0 ± 0,71
	 lgBCF = 2,57 lgP = 3,53	 lgBCF = 4,3 lgP = 5,62	9	2,1 ± 0,25
	 lgBCF = 3,26 lgP = 4,38	 lgBCF = 5,54 lgP = 6,64	8	3,0 ± 0,52

	 lgBCF = 2,78 lgP = 4,1  lgBCF = 5,54 lgP = 6,64	7	3,2 ± 0,4

Примечание: N – число молекулярных пар, для которых выявлена молекулярная трансформация; Δmean – средняя разница значений биоконцентрационного фактора при осуществлении молекулярной трансформации; lgP – липофильность веществ

Результаты интерпретации QSPR модели подтверждают ранее полученные данные [24], согласно которым липофильность соединений оказывает существенное влияние на их способность к биоконцентрированию: между этими двумя характеристиками существует линейная зависимость.

Используя консенсусную QSPR модель, были рассчитаны величины BCF для отдельных авермектинов, которые приведены в таблице 4. Все соединения вошли в область применимости, что может указывать на надежность прогнозируемых значений.

Таблица 4. Предсказанные QSPR-моделью значения BCF авермектинов
Table 4. The QSPR model predicted BCF values of avermectins

Название индивидуальных авермектинов	lgBCF
Авермектин A1a	1,35
Авермектин A1b	1,33
Авермектин B1a	1,32
Авермектин B1b	1,30
Авермектин A2a	1,43
Авермектин A2b	1,43
Авермектин B2a	1,44
Авермектин B2b	1,40
Ивермектин B1a	1,41
Ивермектин B1b	1,46
Дорамектин	1,36
Селамектин	1,50
3"-O-Десметилаверметин B1a	1,39
4"-(ацетиламино)-5-O-диметил-4"-диокси- Авермектин A1a	1,20
Эмамектин B1B	1,39
Эмамектин B1A	1,36
Эприномектин B1a	1,34
Эприномектин B1b	1,35

ЗАКЛЮЧЕНИЕ

В ходе вычислительных экспериментов с применением концептуально различных дескрипторов и методов машинного обучения были разработаны адекватные QSPR модели BCF. Используемые дескрипторы расширили области применимости моделей QSPR, что позволило сделать прогноз BCF для индивидуальных авермектинов более надежным. Полученные значения BCF авермектинов могут быть приняты во внимание при комплексной оценке безопасности авермектинов для окружающей среды.

Интерпретация QSPR моделей позволила определить влияние молекулярных трансформаций на изменение способности органических веществ к биоконцентрированию, что может быть использовано при разработке целевых соединений, а именно при молекулярном дизайне веществ с требуемым уровнем биоконцентрирования.

Сконструированные QSPR модели могут быть использованы для виртуального скрининга органических соединений с целью сокращения финансовых, временных и трудовых затрат при оценке их экотоксикологических параметров.

В будущих исследованиях потребуется анализ и прогноз других экотоксикологических параметров авермектинов, например, способности к биотрансформации и различных типов токсичности.

Список литературы / References:

1. Caly, L., Druce, J.D., Catton, M.G., Jans, D.A., & Wagstaff, K.M. (2020). The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro. *Antiviral Res.*, *178*, 104787. https://doi.org/10.1016/j.antiviral.2020.104787
2. Wagstaff, K.M., Sivakumaran, H., Heaton, S.M., Harrich, D., & Jans, D.A. (2012). Ivermectin is a specific inhibitor of importin α/β-mediated nuclear import able to inhibit replication of HIV-1 and dengue virus. *Biochem. J.*, *443*(3), 851 - 856. https://doi.org/10.1042/BJ20120150
3. Prichard, R., Ménez, C., & Lespine, A. (2012). Moxidectin and the avermectins: Consanguinity but not identity. *Int. J. Parasitol. Drugs & Drug Resist.*, *2*, 134 - 153. https://doi.org/10.1016/j.ijpddr.2012.04.001
4. The Nobel Prize in Physiology or Medicine 2015. https://www.nobelprize.org/prizes/medicine/2015/summary/ (accessed 28.09.2020).
5. Juarez, M., Schcolnik-Cabrera, A., & Dueñas-Gonzalez, A. (2018). The multitargeted drug ivermectin: from an antiparasitic agent to a repositioned cancer drug. *Am. J. Cancer Res.*, *8*(2), 317 - 331.
6. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02006R1907-20140410 (accessed 28.09.2020)
7. Bai, S.H., & Ogbourne, S. (2016). Eco-toxicological effects of the avermectin family with a focus on abamectin and ivermectin. *Chemosphere*, *154*, 204 - 214. https://doi.org/10.1016/j.chemosphere.2016.03.113
8. Piir, G., Kahn, I., García-Sosa, A.T., Sild, S., Ahte, P., & Maran, U. (2018). Best practices for QSAR model reporting: physical and chemical properties, ecotoxicity, environmental fate, human health, and toxicokinetics endpoints. *Environ. Health Perspect.*, *126*(12), 126001. https://doi.org/10.1289/EHP3264

9. Roy, K. (2020). *Ecotoxicological QSARs*. Methods in Pharmacology and Toxicology. New York: Springer US. P. 27. https://doi.org/10.1007/978-1-0716-0150-1
10. Muratov, E.N., Bajorath, J., Sheridan, R.P., Tetko, I.V., Filimonov, D., Poroikov, V., Oprea, T. I., Baskin, I.I., Varnek, A., Roitberg, A., Isayev, O., Curtarolo, S., Fourches, D., Cohen, Y., Aspuru-Guzik, A., Winkler, D.A., Agrafiotis, D., Cherkasov, A., & Tropsha, A. (2020). QSAR without borders. *Chem. Soc. Rev.*, *49*(11), 3525 - 3564. https://doi.org/10.1039/d0cs00098a
11. User's Guide for T.E.S.T. (version 4.2). https://www.epa.gov/chemical-research/users-guide-test-version-42-toxicity-estimation-software-tool-program-estimate (accessed 28.09.2020).
12. CompTox Chemistry Dashboard. https://comptox.epa.gov/dashboard (accessed 28.09.2020).
13. Lunghini, F., Marcou, G., Azam, P., Patoux, R., Enrici, M. H., Bonachera, F., Horvath, D., & Varnek, A. (2019). QSPR models for bioconcentration factor (BCF): are they able to predict data of industrial interest? *SAR QSAR Environ Res.*, *30*(7), 507 - 524. https://doi.org/10.1080/1062936X.2019.1626278
14. Web-based platform OCHEM. OCHEM user's manual http://docs.ochem.eu/display/MAN/OCHEM+Introduction (accessed 28.09.2020).
15. Todeschini, R., & Consonni V. (2000). *Handbook of Molecular Descriptors*. Weinheim: Wiley-VCH, 389 p.
16. Simplex representation of molecular structure - a chemoinformatic tool for calculation of simplex (fragment) descriptors. Available from: https://github.com/DrrDom/sirms (accessed 25.09.2020).
17. Masand, V.H., & Rastija V. (2017). PyDescriptor: A new PyMOL plugin for calculating thousands of easily understandable molecular descriptors. *Chemometrics and Intelligent Laboratory Systems*, 169, 12 - 18. https://doi.org/10.1016/j.chemolab.2017.08.003
18. rdkit.Chem.Descriptors module. http://rdkit.org/docs/source/rdkit.Chem.Descriptors.html (accessed 25.09.2020).
19. alvaDesc KNIME Plugin – tool for calculates alvaDesc molecular descriptors; software available at https://www.alvascience.com/knime-alvadesc/ (accessed 25.09.2020).
20. Breiman, L. RRforest software. http://www.stat.berkeley.edu/~breiman/RandomForests/reg_examples/RFR (accessed 28.09.2020).
21. Tetko, I.V. (2008). Associative neural network. *Methods Mol. Biol.*, *458*, 185 - 202. https://doi.org/10.1007/978-1-60327-101-1_10
22. Raevsky, O.A., Grigorev, V.Yu., Kireev, D.B., & Zefirov, N.S. (1992).Complete thermodynamic description of H-bonding in the framework of multiplicative approach. *Quant. Struct.-Act. Relat.*, *11*(1), 49 - 63. https://doi.org/10.1002/qsar.19920110109
23. Grigorev, V.Yu., & Grigoreva, L.D. (2016). Calculation and properties of fractal descriptors for C2-C9 alkanes. *Moscow Univ. Chem. Bull.*, *71*(3), 199 - 204. https://doi.org/10.3103/S0027131416030056
24. Meylan, W.M., Howard, P.H., Boethling, R.S., Aronson, D., Printup, H., & Gouchie, S. (1999). Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partition coefficient. *Environ. Toxicol. Chem.*, *18*, 664 - 672. https://doi.org/10.1002/etc.5620180412