Биомедицинская химия, 2019 том 65, вып. 2, с. 103-113.

©Коллектив авторов

QSAR-МОДЕЛИРОВАНИЕ ИНГИБИТОРОВ ДЕЗОКСИУРИДИНТРИФОСФАТАЗЫ В РЯДУ НЕКОТОРЫХ ПРОИЗВОДНЫХ УРАЦИЛА

Ю.З. Мартынова[1], В.Р. Хайруллина[1]*, А.Р. Гимадиева[2], А.Г. Мустафин[1,2]

[1]Башкирский государственный университет,
Республика Башкортостан, 450076, Уфа, ул. З. Валиди, 32; *эл. почта: Veronika1979@yandex.ru
[2]Уфимский Институт химии УФИЦ РАН, Республика Башкортостан, 450054, Уфа, пр. Октября, 69

С использованием программы GUSAR выполнен количественный анализ взаимосвязи между структурой и активностью 135 ингибиторов дезоксиуридинтрифосфатазы человека на основе производных урацила, в интервале значений IC_{50}=30÷185000 нмоль/л. Построено шесть консенсусных моделей QSAR, обладающих достаточно высокой описательной точностью и предсказательной способностью. В результате виртуального скрининга базы данных ChEMBL с использованием построенных консенсусных моделей выявлено 10 потенциальных ингибиторов дезоксиуридинтрифосфатазы, которые рекомендованы для экспериментального тестирования.

Ключевые слова: ингибиторы дезоксиуридинтрифосфатазы; моделирование QSAR; программа GUSAR; MNA- и QNA-дескрипторы

DOI: 10.18097/PBMC20196502103

ВВЕДЕНИЕ

Фермент дезоксиуридинтрифосфатаза (dUTPase, КФ 3.6.1.23) относится к семейству гидролаз. Этот фермент, присутствующий в различных живых системах, включая простейшие, бактерии, вирусы, а также организм животных и человека, катализирует реакцию гидролиза дезоксиуридинтрифосфата с образованием дезоксиуридинмонофосфата и пирофосфат-аниона. Данная биохимическая реакция выполняет две важные функции: 1) снижает внутриклеточную концентрацию дезоксиуридинтрифосфата и тем самым способствует снижению уровня мутаций в молекулах ДНК; 2) продукт гидролиза – дезоксиуридинмонофосфат – является субстратом для другого жизненно важного фермента – тимидилатсинтазы, непосредственно участвующей в синтезе дезокситимидинмонофосфата [1-11]. В дальнейшем дезокситимидинмонофосфат в результате фосфорилирования превращается в дезокситимидинтрифосфат, представляющий собой одно из мономерных звеньев молекул ДНК [3-11]. Таким образом, dUTPase требуется не только для биосинтеза дезокситимидинтрифосфата, но также для поддержания целостности геномной ДНК.

В настоящее время dUTPase рассматривается как перспективная мишень для разработки лекарственных средств широкого спектра действия: противоопухолевых, противовирусных (в том числе и против вируса иммунодефицита человека), антибактериальных, противомалярийных. Предполагается, что терапевтическое действие этих препаратов будет осуществляться благодаря подавлению ими биосинтеза ДНК в различных вирусах, бактериях и простейших, а также в опухолевых клетках. Кроме того, актуальность поиска ингибиторов dUTPase обусловлена и тем фактом, что этот фермент способен значительно снижать эффективность химиотерапии при лечении пациентов со злокачественными опухолями 5-фторурацилом, увеличивая риск летального исхода. Данный эффект обусловлен тем, что наряду с дезоксиуридинтрифосфатом, dUTPase может гидролизовать и близкий к нему по строению метаболит 5-фторурацила – 5-фтордезоксиуридинтрифосфат с образованием 5-фтордезоксиуридинмонофосфата, тем самым снижая концентрацию исходного вещества. Эта реакция протекает аналогично реакции гидролиза дезоксиуридинтрифосфата, но является нежелательной, поскольку по мере включения dUTPase в метаболизм 5-фтордезоксиуридинмонофосфата у онкологических больных со временем возникает резистентность к 5-фторурацилу. Таким образом, поиск эффективных ингибиторов dUTPase является не только фундаментальной, но и практически важной задачей медицинской химии [1, 12-14].

Однако рациональный поиск новых лекарств без применения методов виртуального скрининга невозможен как с экономической точки зрения, так и из-за высоких временных затрат. Методы анализа количественных взаимосвязей структура-активность (QSAR) позволяют осуществлять на ранних стадиях исследования отбор наиболее перспективных молекул для экспериментального тестирования. Практическая реализация такого подхода требует наличия достаточно точных и обладающих предсказательной способностью моделей QSAR [15-22]. Поэтому целями настоящей работы стали построение и валидация моделей QSAR, ориентированных на оценку численных значений полуэффективной концентрации IC_{50} для ингибиторов dUTPase в ряду некоторых производных урацила.

МЕТОДИКА

Изучение количественной взаимосвязи "структура – ингибиторная активность" для ингибиторов dUTPase проводили в ряду производных урацила с общей структурной формулой I (рис. 1) с использованием компьютерной программы GUSAR (General Unrestricted Structure-Activity Relationships) [15-22]. Модели QSAR строили в несколько этапов, схематично представленных на рисунке 2. Методика формирования и отбора дескрипторов для построения консенсусных моделей описана в работах [15-22].

Формирование обучающих и тестовых выборок

Обучающие выборки OB1-OB2, а также внешние и внутренние тестовые выборки TB1-TB2 для построения моделей M1-M6 формировали на основе массива структур MS1 в соответствии со схемой, представленной на рисунке 3. Массив MS1 формировали на основе литературных данных. Он содержит численные значения IC_{50} для 135 структур ингибиторов dUTPase, полученные экспериментально, измеренные авторами работ [12-14]. Детальная характеристика обучающих выборок OB1-OB2 и тестовых выборок TB1-TB2 представлена в таблицах 1 и 2, соответственно. Как видно из рисунка 3 и данных таблицы 1, анализируемые соединения обладают разной степенью ингибиторного действия в отношении dUPhase человека.

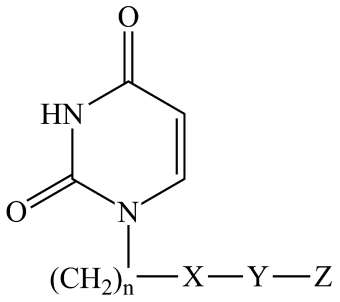


Рисунок 1. Общая структурная формула ингибиторов dUTPhase, вошедших в обучающие массивы OB1 и OB2, где n = 1÷3; X – связь с атомом кислорода, серы или т.п.; Y – линейная или разветвлённая алкиленовая группа, имеющая от 1 до 8 атомов углерода, или т.п.; и Z – $-SO_2NR^1R^2$ или $-NR^3SO_2-R^4$, где R^1 и R^2 – алкильная группа, имеющая от 1 до 6 атомов углерода, аралкильная группа, которая необязательно замещена, или т.п.; R^3 – алкильная группа, имеющая от 1 до 6 атомов углерода, или т.п.; и R^4 – ароматическая углеводородная группа, ненасыщенная гетероциклическая группа или т.п.

Обучающая выборка OB1 предназначена для построения моделей QSAR M1-M3 и включает в себя 116 ингибиторов dUThase. Для проверки предсказательной способности моделей M1-M3 использовали внешнюю тестовую выборку TB1. Обе эти выборки получены в результате разделения исходного массива данных MS1 в соотношении 6:1 посредством того, что в TB1 из MS1 переносилось каждое седьмое соединение. Предварительно все структуры массива данных MS1 были ранжированы по возрастанию численного значения IC_{50}. Аналогичным образом из выборки OB1 были получены обучающая выборка OB2 и внутренняя тестовая выборка TB2.

Предварительно перед построением моделей значения IC_{50} из нмоль/л пересчитывали в моль/л и затем переводили их в величины pIC_{50} по формуле:

$pIC_{50} = -lgIC_{50}$, где IC_{50} (в моль/л).

Проверку предсказательной способности итоговых моделей M1-M6 проводили на структурах тестовых выборок TB1-TB2, характеристика которых представлена в таблице 2. Приведённые в таблицах 1 и 2 данные позволяют заключить, что диапазон активности моделируемых соединений $\Delta pIC_{50}>1,5$, что удовлетворяет необходимым условиям для построения качественных моделей QSAR [23].

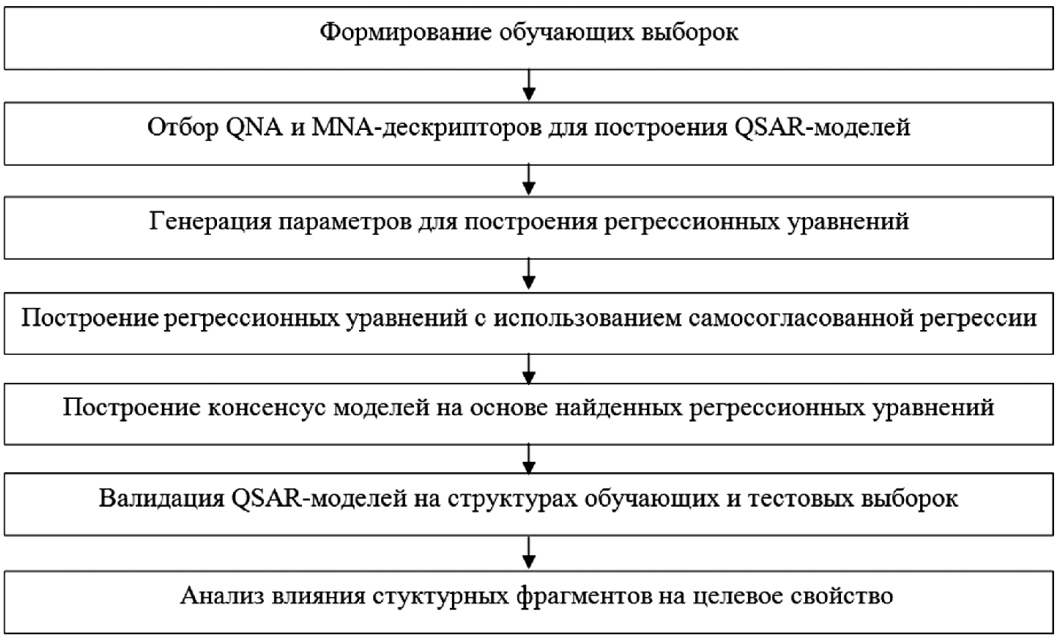


Рисунок 2. Общая идеология построения моделей QSAR M1-M6.

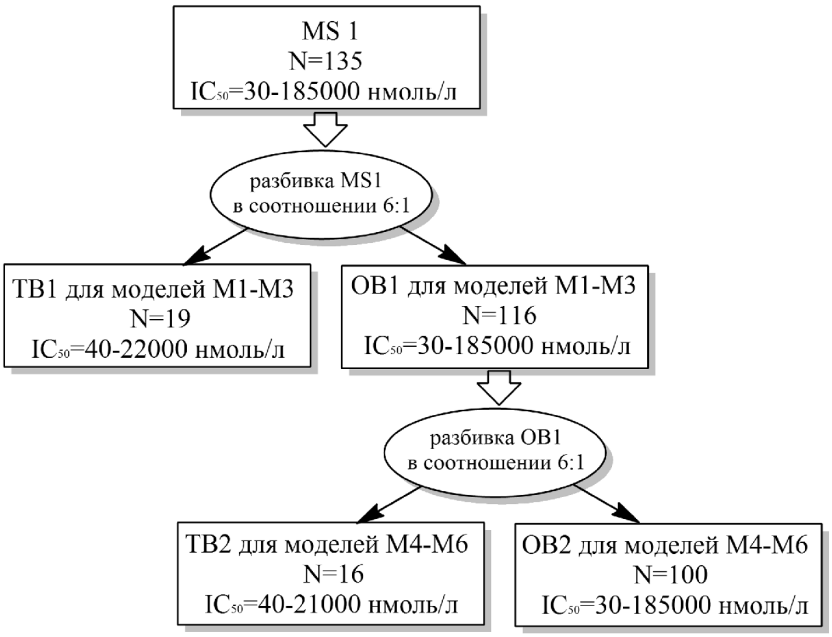


Рисунок 3. Формирование обучающих и тестовых выборок для построения моделей QSAR M1-M6.

Таблица 1. Характеристика обучающей выборки OB_i

Комментарий	Характеристика параметров OB_i*	OB1*	OB2*
Количество соединений	N	116	100
Среднее значение pIC_{50}	$\overline{pIC_{50}}$	5,63	
Диапазон значения pIC_{50}	ΔpIC_{50}	3,79	
Распределение наблюдаемых значений OB_i вокруг среднего значения TB (в %)	$\overline{pIC_{50OB}}$±0,5%	31,90	32,00
	$\overline{pIC_{50OB}}$±1,0%	72,41	72,00
	$\overline{pIC_{50OB}}$±1,5%	88,79	88,00
	$\overline{pIC_{50OB}}$±2,0%	100,00	100,00
	$0,10\times\Delta pIC_{50}$	0,38	
	$0,15\times\Delta pIC_{50}$	0,57	
	$0,20\times\Delta pIC_{50}$	0,76	
	$0,25\times\Delta pIC_{50}$	0,95	

Примечание. Здесь и в таблице 2: * - цифрами 1 и 2 обозначены номера обучающих выборок.

Таблица 2. Характеристика тестовой выборки TB_i

Комментарий	Характеристика параметров OB_i*	TB1*	TB2*
Количество соединений	N	19	16
Среднее значение pIC_{50}	$\overline{pIC_{50}}$	6,09	6,07
Диапазон значения pIC_{50}	ΔpIC_{50}	2,74	2,72
Распределение наблюдаемых значений TB_i вокруг среднего значения TB (в %)	$\overline{pIC_{50TB}}$±0,5%	47,37	50,00
	$\overline{pIC_{50TB}}$±1,0%	78,95	81,25
	$\overline{pIC_{50TB}}$±1,5%	100,00	100,00
	$\overline{pIC_{50TB}}$±2,0%	100,00	100,00
Распределение наблюдаемых значений TB_i вокруг среднего значения OB (в %)	$\overline{pIC_{50OB}}$±0,5%	36,84	31,25
	$\overline{pIC_{50OB}}$±1,0%	78,95	75,00
	$\overline{pIC_{50OB}}$±1,5%	89,47	93,75
	$\overline{pIC_{50OB}}$±2,0%	100,00	100,00

Построение моделей QSAR

Модели QSAR M1-M6 строили с использованием программы GUSAR на основе двух типов дескрипторов атомных окрестностей: QNA (Quantitative Neighbourhoods of Atoms) и MNA (Multilevel Neighbourhoods of Atoms) [15-21], расчёт которых осуществлялся автоматически на основе структурных формул химических соединений. В GUSAR проблема возможной мультиколлинеарности дескрипторов решается двумя способами: 1) алгоритмами формирования QNA-дескрипторов; 2) использованием для отбора признаков в итоговую модель QSAR метода самосогласованной регрессии, SCR. В нее включен метод регуляризации, который препятствует отбору коррелируемых дескрипторов. Детально описание идеологии расчёта QNA- и MNA-дескрипторов представлено в работах [17-19, 21, 24]. В дополнение к этому, для повышения описательной и прогностической способности моделей использовали такие дескрипторы, как топологическая длина, топологический объём и липофильность, которые отражают интегральные характеристики молекулы.

В связи с тем, что структуры моделируемых ингибиторов dUThase весьма похожи, для отбора оптимального числа дескрипторов при построении QSAR-моделей в программе GUSAR использовали метод самосогласованной регрессии SCR [16-20]. Благодаря этому удалялись переменные, коррелирующие между собой, а также плохо описывающие целевое свойство. Кроме того, сообщалось, что этот метод устойчив к недостаточно высокому качеству экспериментальных данных [21].

Программа GUSAR позволяет строить как частные регрессионные зависимости, так и консенсусные модели на их основе. В настоящем исследовании с целью снижения вариабельности конечных результатов мы использовали для построения моделей QSAR M1-M6 консенсусный подход. Консенсусные модели строили в программе GUSAR автоматически, исходя из принципа общего сходства отдельных регрессионных зависимостей [21]. Консенсус-модели M1 и M4 строили независимо друг от друга на основе QNA-дескрипторов и вышеупомянутых трёх интегральных дескрипторов. Модели M2 и M5 строили на основе всех четырёх типов дескрипторов. Таким образом, каждая из четырех итоговых моделей QSAR M1-M2, M4-M5 включала в себя по 10 регрессионных зависимостей, объединенных автоматически, исходя из принципа их общего сходства. Модели M3 и M6 строили по тому же принципу, они включали в себя по 100 частных регрессионных зависимостей. При этом каждая из этих 100 отдельных моделей, входящих в модели M3 и M6, была построена независимо от других на основании либо набора дескрипторов QNA + 3 интегральных дескриптора, либо набора дескрипторов MNA + 3 интегральных дескриптора.

Регрессионные уравнения, построенные на основе QNA- и MNA-дескрипторов, не поддаются однозначной физической интерпретации, и по этой причине регрессионные уравнения, построенные на их основе, в явном виде в программе GUSAR не отображаются. Итоговый прогноз численного значения pIC_{50} для конкретного соединения по моделям M1-M6 формировался на основе результатов усреднения прогнозных значений pIC_{50} частных регрессионных моделей QSAR, входящих в используемую консенсусную модель.

Оценка точности и предсказательной способности моделей QSAR

Точность полученных моделей QSAR M1-M6 оценивали по результатам прогноза параметра IC_{50} для структур обучающих выборок OB1-OB2 с использованием двух категорий метрик: 1) по метрикам на основе коэффициентов детерминации R^2 (R^2_{train}, $R_0{}^2_{train}$, $R'_0{}^2_{train}$, $\overline{R^2}_m$, CCC). 2) по метрикам, основанным на оценке ошибок прогноза значений pIC_{50} для структур обучающих выборок OB1-OB2 (среднеквадратичная ошибка (RMSE), средняя абсолютная ошибка (MAE), стандартное отклонение (SD)) [23, 25-26]. Предсказательную способность этих моделей оценивали аналогично по результатам прогноза параметра pIC_{50} для структур внешней и внутренней тестовых выборок TB1 и TB2 с использованием тех же категорий метрик. Расчёт этих статистических параметров проводили с использованием программы Xternal Validation Plus 1.2 [28].

Дополнительно проверку адекватности построенных моделей M1-M6 осуществляли с использованием процедуры скользящего контроля с 20-ти кратным случайным удалением 20% соединений из обучающих выборок OB1 и OB2. Обе процедуры в программе GUSAR осуществляются автоматически [15-22].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

На основе консенсусного подхода, реализованного в программе GUSAR, проведено моделирование количественной взаимосвязи между строением и активностью 135 ингибиторов dUTPase с общей структурной формулой I (рис. 1), содержащимися в обучающих выборках OB1-OB2. В результате, в зависимости от типа использованных при расчётах дескрипторов, для каждой из этих обучающих выборок построено по три модели QSAR. Описательная и прогностическая способность этих моделей оценена на структурах обучающих выборок OB1-OB2 при скользящем контроле, а также путём прогноза для структур тестовых выборок TB1-TB2. В таблице 3 представлены статистические характеристики, полученные с использованием программы GUSAR при оценке точности прогноза значений pIC_{50} для ингибиторов dUPhase, входивших в обучающие выборки OB1-OB2, по моделям M1-M6. При этом коэффициенты детерминации, величины стандартного отклонения и критерия Фишера, представленные в таблице 3, представляют собой усреднённые значения, полученные с учётом всех частных регрессионных моделей, входящих в модель M_i.

Таблица 3. Статистические характеристики и оценка показателей точности предсказаний значений pIC_{50} для ингибиторов dUPhase по консенсус-моделям M1-M6

Обучающая выборка	Модель	N	N_{mod}	$\overline{R^2_{ові}}$	$\overline{Q^2_{ові}}$	$\overline{F_{ові}}$	$\overline{SD_{ові}}$	V
Модели QSAR, построенные на основе QNA-дескрипторов								
OB1	M1	116	10	0,706	0,642	21,453	0,433	10
OB2	M4	100	10	0,696	0,638	23,096	0,445	8
Модели QSAR, построенные на основе MNA-дескрипторов								
OB1	M2	116	10	0,786	0,722	24,586	0,374	12
OB2	M5	100	10	0,773	0,706	20,627	0,390	11
Модели QSAR, построенные на основе QNA- и MNA-дескрипторов								
OB1	M3	116	100	0,783	0,729	25,119	0,377	11
OB2	M6	100	100	0,763	0,705	22,438	0,399	10

Примечание: N – число структур в обучающей выборке; N_{mod} – число частных регрессионных моделей, входящих в модель M_i; $R^2_{ов}$ – коэффициент детерминации, рассчитанный для соединений из обучающей выборки; $Q^2_{ов}$ – коэффициент корреляции, рассчитанный на обучающей выборке при скользящем контроле с исключением по одному; F – критерий Фишера; SD – стандартное отклонение; V – число переменных в конечном регрессионном уравнении.

Дополнительно для оценки точности и предсказательной способности моделей M1-M6 использовали программу Xternal Validation Plus 1.2. Расчёт статистических параметров моделей QSAR M1-M6 в этой программе осуществляется на основе сопоставления экспериментальных данных pIC_{50}, с усреднёнными данными pIC_{50pred}, которые были ранее предсказаны с использованием программы GUSAR. Эта процедура проводилась двукратно без усреднения итоговых результатов [25]:

1) для полного набора соединений каждой обучающей и тестовой выборки;

2) для каждой обучающей и тестовой выборки, содержащей 95% исходных данных.

В соответствии с современными рекомендациями в области QSAR-моделирования [23, 25-26], точность и предсказательную способность модели M_i оценивали как высокую, если при её тестировании на соединениях обучающих и тестовых выборок OB_i и TB_i одновременно выполнялись следующие четыре условия: 1) численные значения коэффициентов детерминации R^2, R^2_0, $R^{2'}_0$, Q^2_{F1}, Q^2_{F2} и критерия CCC были близки между собой и стремились к единице; 2) численное значение критерия $R^2_m>0,5$ при $\Delta R^2_m<0,2$; 3) численное значение средней абсолютной ошибки MAE не превышало 10% от интервала активности соединений ΔpIC_{50} моделируемой обучающей выборки OB_i; 4) величина параметра MAE+3·SD (где SD – стандартное отклонение) не превышала 20% от интервала активности соединений ΔpIC_{50}, моделируемой обучающей выборки OB_i.

Точность и предсказательную способность модели M_i оценивали как низкую, если при её тестировании на соединениях обучающих и тестовых выборок OB_i и TB_i одновременно выполнялись следующие условия: 1) численные значения коэффициентов детерминации R^2, R^2_0, $R^{2'}_0$, Q^2_{F1}, Q^2_{F2} и критерия CCC не превышали порогового значения 0,5 для TB_i и не превышали порогового значения 0,75 для обучающей выборки OB_i; 2) численное значение критерия $R^2_m\leq0,5$ при $\Delta R^2_m>0,2$; 3) численное значение средней абсолютной ошибки MAE превышало 15% от интервала активности соединений ΔpIC_{50} моделируемой обучающей выборки OB_i; 4) величина параметра MAE+3·SD превышала 25% от интервала активности соединений ΔpIC_{50} моделируемой обучающей выборки OB_i. Во всех остальных случаях точность и предсказательная способность моделей M1-M6 по вышеописанным критериям оценивались как умеренные.

В ходе выполненных исследований установлено, что все модели QSAR M1-M6 обладали умеренной описательной и прогностической способностью. В качестве примера на лепестковой диаграмме, изображенной на рисунках 4(a,b) и 5(a,b), представлены статистические параметры, оцененные по результатам прогноза параметра pIC_{50} для ингибиторов dUTPase, содержащихся в обучающих выборках OB1-OB2 и тестовых выборках TB1 и TB2, с использованием моделей QSAR M1-M6. При этом корректность выбора типа оптимального набора дескрипторов для построения этих моделей подтверждена отсутствием систематической ошибки прогноза значения pIC_{50} для ингибиторов dUTPase, содержащихся во внутренних и внешних тестовых выборках TB1-TB2. В пользу того, что все консенсусные модели обладали довольно высокой точностью, свидетельствуют и данные таблицы 3.

Расхождения в численных значениях коэффициентов детерминации R^2 для структур обучающих выборок OB1-OB2, полученных при оценке точности и предсказательной способности консенсус-моделей M1-M6 в программах GUSAR 2013 и Xternal Validation Plus 1,2 и представленных в таблице 3 и на рисунке 4, обусловлены разной идеологией расчёта этих параметров. Так, в программе GUSAR численное значение активности для тестируемого соединения при использовании любой консенсусной модели

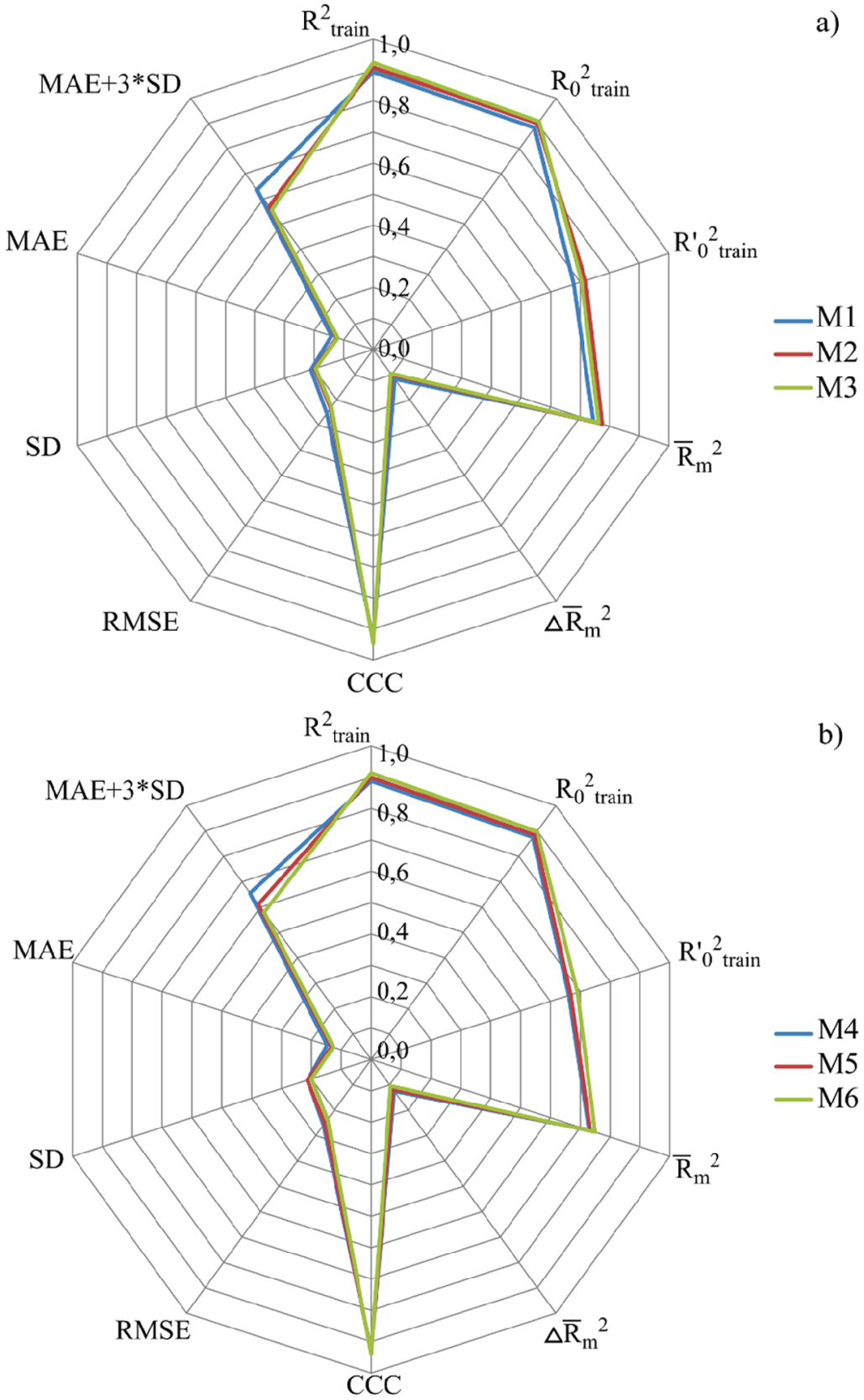


Рисунок 4. Статистические параметры, оцененные по результатам прогноза параметра pIC_{50} для ингибиторов dUTPase, содержащихся в обучающих выборках OB1 (a) и OB2 (b), по моделям QSAR M1-M6, где R^2_{train}, $R_0{}^2_{train}$, $R'_0{}^2_{train}$, R^2_m, ΔR^2_m, CCC – коэффициенты детерминации, рассчитанные на основе сопоставления экспериментальных и расчётных данных pIC_{50} для структур обучающих выборок OB1(a) и OB2 (b) [23] для соединений OB_i; RMSE – среднеквадратическая ошибка; MAE – средняя абсолютная ошибка; SD – стандартное отклонение.

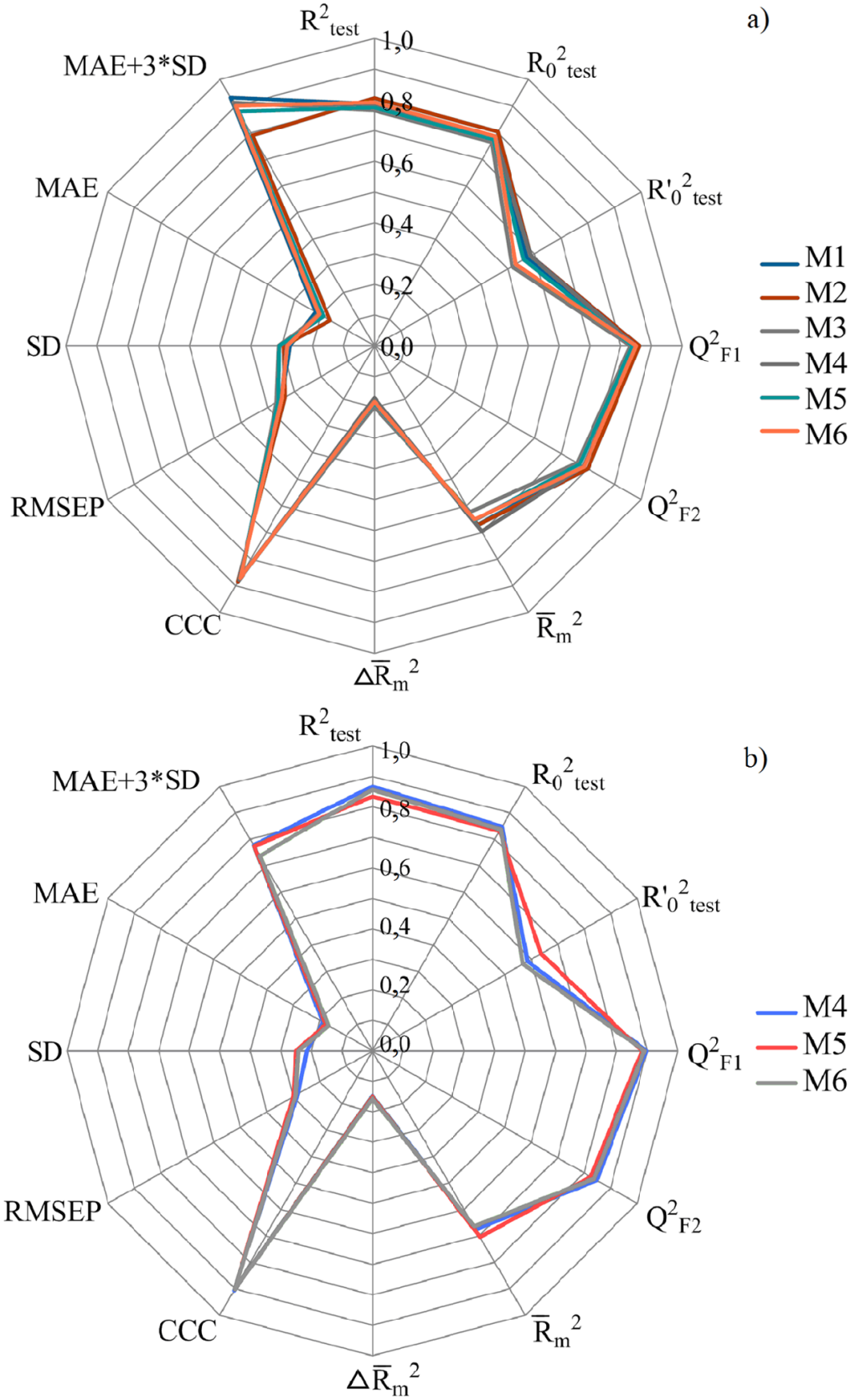


Рисунок 5. Статистические параметры, оцененные по результатам прогноза параметра pIC_{50} для ингибиторов dUTPase, содержащихся в тестовых выборках TB1 (a) и TB2 (b), по моделям QSAR M1-M6, где Q^2_{F1}, Q^2_{F2}, R^2_{train}, $R_0{}^2_{train}$, $R'_0{}^2_{train}$, $\overline{R^2}_m$, $\overline{\Delta R^2}_m$, CCC – коэффициенты детерминации, рассчитанные на основе сопоставления экспериментальных и расчётных данных pIC_{50} для структур обучающих выборок OB1(a) и OB2 (b) [23] для соединений OB_i; RMSEP – среднеквадратическая ошибка; MAE – средняя абсолютная ошибка; SD – стандартное отклонение.

представляет собой средневзвешенное прогнозное значение величины pIC_{50}, которое оценивается по всем регрессионным зависимостям, входящим в данную модель. В частности, для выборки OB1 при прогнозе с использованием модели M1 получается набор из 10 предсказанных значений pIC_{50pred} и 10 наборов разных критериев внутренней валидации: R^2_{OBi}, Q^2_{OBi}, F_{OBi}, SD_{OBi}. В дальнейшем эти данные усредняются с учётом всех 10 частных моделей QSAR. В то же время, как упоминалось выше, в программе Xternal Validation Plus 1,2 статистические параметры рассчитываются двухкратно, исходя из результатов сопоставления экспериментальных данных pIC_{50} с усреднёнными данными pIC_{50pred}, предсказанными с использованием программы GUSAR: для 100% и для 95% соединений обучающей и тестовой выборки без усреднения конечных результатов [28].

Как видно из рисунка 5, значения ошибок MAE при прогнозе pIC_{50} для структур тестовых выборок TB1-TB2 не превышают 5% от диапазона ингибиторной активности pIC_{50} структур обучающих выборок OB1-OB2. Доля соединений тестовых выборок TB1 и TB2, для которых ошибка прогноза значений ΔpIC_{50} не превышает интервал $0{,}25\times\Delta pIC_{50}$ составляет 5,3%. Таким образом, все построенные нами модели характеризуются достаточно высокой точностью и умеренной предсказательной способностью независимо от типа выбранных дескрипторов.

Для решения задачи поиска новых потенциальных ингибиторов dUTPase человека среди производных урацила с общей структурной формулой I (рис. 1) в базе данных ChEMBL нами были выбраны модели M3 и M6, так как они включают в себя по 100 частных регрессионных моделей и построены на основе максимального набора структур и дескрипторов. Виртуальный скрининг проводили среди соединений-лидеров и субстанций лекарственных препаратов различного фармакологического профиля, как минимум, на 70% схожих по строению с соединениями, входящими в обучающие выборки OB1 и OB2 (по индексу Танимото) с шифрами ChEMBL3664439 и ChEMBL2057909 (https://www.ebi.ac.uk). Виртуальный скрининг проводили для 163 производных урацила, для которых было экспериментально установлено наличие противоопухолевого, антибактериального или противовирусного действия. Тем не менее, только 10 отобранных нами соединений ранее не были известны как ингибиторы dUTPhase и соответствовали области применимости моделей M3 и M6. Полный список структур и предсказанных значений IC_{50pred} моделями M3 и M6 для этих соединений представлен в таблице 4. Согласно прогнозу, эти структуры могут обладать выраженной ингибиторной активностью в отношении dUTPase человека, что позволяет рассматривать их в качестве перспективных соединений для экспериментального тестирования.

Таблица 4. Потенциально активные ингибиторы dUTPhase, отобранные из базы данных ChEMBL на основе виртуального скрининга

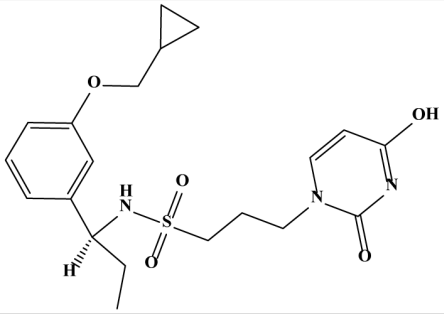
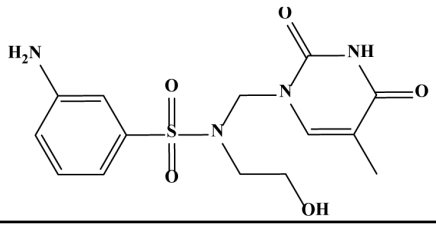
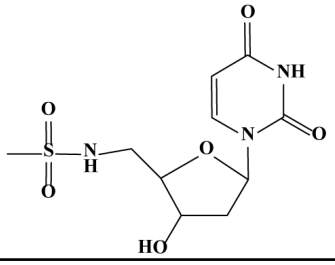
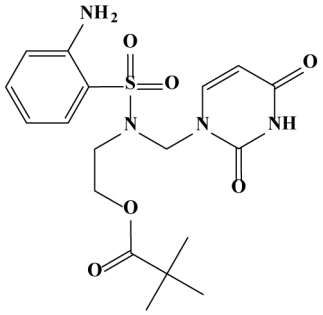
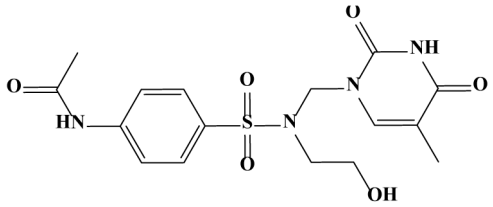
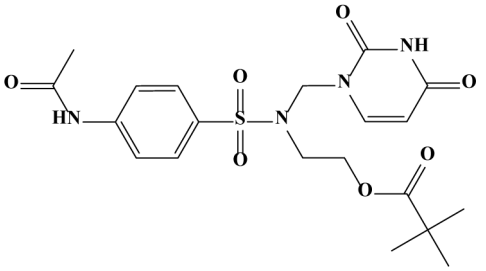
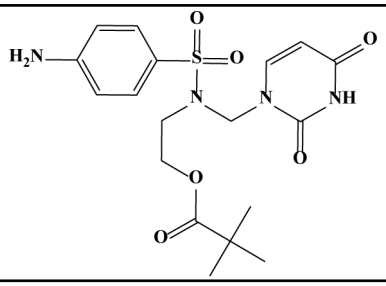
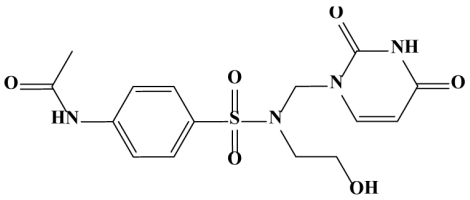
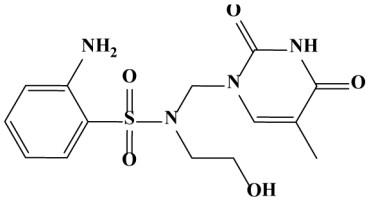
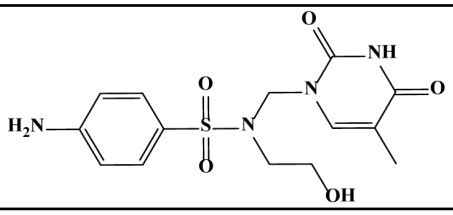
№	Шифр соединения	Структурная формула	IC_{50pred}, нмоль/л	
			M3	M6
1	CHEMBL3920919		120	131
2	CHEMBL516567		12170	16252
3	CHEMBL194845		4804	11479

Таблица 4. Потенциально активные ингибиторы dUTPhase, отобранные из базы данных ChEMBL на основе виртуального скрининга (продолжение)

№	Шифр соединения	Структурная формула	IC_{50pred}, нмоль/л	
			М3	М6
4	CHEMBL546285		8736	12114
5	CHEMBL491212		12218	11082
6	CHEMBL490591		5954	6038
7	CHEMBL478238		9292	10009
8	CHEMBL476347		15535	13369
9	CHEMBL476345		11918	19182
10	CHEMBL476344		13059	15823

ЗАКЛЮЧЕНИЕ

С использованием программы GUSAR построены консенсусные модели QSAR, характеризующие количественные соотношения “структура – ингибиторная активность” в отношении dUTPase человека, обладающие достаточно высокой точностью и предсказательной способностью. В результате виртуального скрининга базы данных ChEMBL с использованием наиболее качественных полученных нами моделей QSAR выявлено 10 потенциальных ингибиторов dUTPase, которые рекомендованы для экспериментальных исследований.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена при поддержке гранта №19-73-20073 Российского научного фонда.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая статья не содержит каких-либо исследований с участием людей или с использованием животных в качестве объектов.

ЛИТЕРАТУРА

1. *Tóth J., Varga B., Kovács M., Málnási-Csizmadia A., Vértessy B.G.* (2007) J. Biol. Chem., **282**(4), 33572-33582.
2. *Shlomai J., Kornberg, A.* (1978) J. Biol. Chem., **253**(9), 3305-3312.
3. *Gadsden M.H., McIntosh E.M., Game J.C., Wilson P.J., Haynes R.H.* (1993) EMBO J. **12**(11), 4425-4431.
4. *el-Hajj H.H., Zhang H., Weiss B.* (1988) J. Bacteriol., **170**(3), 1069-1075. DOI: 10.1128/jb.170.3.1069-1075.1988
5. *Traut T.W.* (1994) J. Mol. Cell. Biochem., **140**(1), 1-22.
6. *Pearl L.H., Savva R.* (1996) J. Nat. Struct. Biol., **3**(6), 485-487.
7. *Pugacheva E.N., Ivanov A.V., Kravchenko J.E., Kopnin B.P., Levine A.J., Chumakov P.M.* (2002) J. Oncogene, **21**(30), 4595-4600.
8. *Chano T., Mori K., Scotlandi K., Benini S., Lapucci C., Manara M.C., Serra M., Picci P., Okabe H., Baldini N.* (2004) J. Oncol. Rep., **11**(6), 1257-1263.
9. *Ladner R.D., Lynch F.J., Groshen S., Xiong Y.P., Sherrod A., Caradonna S.J., Stoehlmacher J., Lenz H.J.* (2000) J. Cancer Res., **60**, 3493-3503.
10. *Romeike B.F., Bockeler A., Kremmer E., Sommer P., Krick C., Grasser F.* (2005) J. Pathol. Res. Pract., **201**(11), 727-732.
11. *Vassylyev D.G., Morikawa K.* (1996) J. Structure, **4**(12), 1381-1385.
12. *Miyakoshi H., Miyahara S., Yokogawa T., Chong Kh.T., Taguchi J., Endoh K., Yano W., Wakasa T., Ueno H., Takao Y., Nomura M., Shuto S., Nagasawa H., Fukuoka M.* (2012) J. Med. Chem., **55**(7), 2960-2969.
13. *Miyahara S., Miyakoshi H., Yokogawa T., Chong Kh.T., Taguchi J., Muto T., Endoh K., Yano W., Wakasa T., Ueno H., Takao Y., Fujioka A., Hashimoto A., Itou K., Yamamura K., Nomura M., Nagasawa H., Shuto S., Fukuoka M.* (2012) J. Med. Chem., **55**(7), 2970-2980.
14. *Fukuoka M., Yokogawa T., Miyahara S., Miyakoshi H., Yano W., Taguchi J., Takao Y.* (2013) Patent №8530490 B2 US, Appl. №12/996,079, United States Patent, 10.09.2013.
15. *Хайруллина В.Р., Акбашева Ю.З., Гимадиева А.Р., Мустафин А.Г.* (2017) Вестник Башкирск. ун-та, **19**(2), 960-965.
16. *Хайруллина В.Р., Акбашева Ю.З., Янгирова К.Р., Гимадиева А.Р., Мустафин А.Г.* (2018) Вестник Башкирск. ун-та, **23**(4), 987-993.
17. *Zakharov A.V., Lagunin A.A., Filimonov D.A., Poroikov V.V.* (2012) Chem, Res. Toxicol., **25**(11) 2378-2385.
18. *Lagunin A., Zakharov A., Filimonov D., Poroikov V.* (2011) Mol. Inf., **30**(2-3), 241-250.
19. *Filimonov D.A., Zakharov A.V., Lagunin A.A., Poroikov V.V.* (2009) J. SAR QSAR Environ. Res., **20**(7-8), 679-709.
20. *Masand V.H., Devidas T., Mahajan D.T., Patil K.N., Hadda T.B., Youssoufi M.H., Jawarkar R.D., Shibi I.G.* (2013) J. Chem. Biol. Drug Des., **81**(4), 527-536.
21. *Zakharov A.V., Peach M.L., Sitzmann M., Nicklaus M.C.* (2014) J. Chem. Inf. Model., **54**(3), 713-719.
22. *Хайруллина В.Р., Герчиков А.Я., Лагунин А.А., Зарудий Ф.С.* (2017) Хим.-фарм. журнал, **51**(10), 33-37.
23. *Dearden J.C., Cronin M.T.D., Kaiser K.L.E.* (2009) J. SAR QSAR Environ. Res., **20**(3-4) 241-266.
24. *Lagunin A.A., Geronikaki A., Eleftheriou P., Pogodin P.V., Zakharov A.V.* (2019) J. Chem. Inf. Model., 1-61.
25. *Roy K., Mitra I., Kar S., Ojha P.K., Das R.N., Kabir H.* (2012) J. Chem. Inf. Model., **52**(2), 396-408.
26. *Roy P.P., Paul S., Mitra I., Roy K.* (2009) J. Mol., **14**(5), 1660-1701.
27. *Roy K., Das R.N., Ambure P., Aher R.B.* (2016) J. Chemom. Intell. Lab. Syst., **152**, 18-33.
28. Xternal Validation Plus, version 1.2, 2016, DTC Lab., Kolkata, India.

Поступила в редакцию: 19. 01. 2019.
После доработки: 18. 02. 2019.
Принята к печати: 18. 03. 2019.

QSAR-MODELING OF DESOXYURIDINE TRIPHOSPHATASE INHIBITORS IN A SERIES OF SOME DERIVATIVES OF URACIL

Yu.Z. Martynova[1], V.R. Khairullina[1]*, A.R. Gimadieva[2], A.G. Mustafin[1,2]

[1]Bashkir State University,
32, Z. Validi str., Ufa, Republic of Bashkortostan, 450076 Russia; *e-mail: Veronika1979@yandex.ru
[2]Ufa Institute of Chemistry of RAS, 69, Oktyabrya ave., Ufa, Republic Bashkortostan, 450054 Russia

Due to the widespread prevalence, deoxyuridine triphosphatase (UTPase) is considered by modern biochemists and physicians as a promising target for the development of drugs with a wide range of activities. The therapeutic effect of these drugs will be due to suppression of DNA biosynthesis in various viruses, bacteria and protozoa. In order to rationalize the search for new dUTPase inhibitors, domestic and foreign researchers are actively using the QSAR methodology at the selection stage of hit compounds. However, the practical application of this methodology is impossible without existence of valid QSAR models. With the use of the GUSAR 2013 program, a quantitative analysis of the relationship between the structure and efficacy of 135 dUTPase inhibitors based on uracil derivatives was performed in the IC_{50} range of 30÷185000 nmol/L. Six statistically significant valid consensus models, characterized by high descriptive ability and moderate prognostic ability on the structures of training and test samples, are constructed. To build valid QSAR models for dUTPase inhibitors can use QNA or MNA descriptors and their combinations in a consensus approach.

Key words: deoxyuridine triphosphatase inhibitors; QSAR modeling; GUSAR 2013 program; MNA and QNA descriptors