

В. Р. Хайруллина¹, А. Я. Герчиков¹, А. А. Лагунин^{2, 3}, Ф. С. Зарудий⁴

QSAR-МОДЕЛИРОВАНИЕ ИНГИБИТОРОВ ТИМИДАЛАСИНТАЗЫ В РЯДУ ПРОИЗВОДНЫХ ХИНАЗОЛИНА

¹ Башкирский государственный университет, химический факультет, Россия, 450076, Уфа, ул. Заки Валиди, 32; факс: +7(347)229-9707; e-mail: gerchikov@inbox.ru, Veronika1979@yandex.ru

² Российский национальный исследовательский медицинский университет им. Н. И. Пирогова, Россия, 117997, Москва, ул. Островитянова, 1; факс: +7(495)434-1422.

³ Научно-исследовательский институт биомедицинской химии им. В. Н. Ореховича, Россия, 119121, Москва, ул. Погодинская, 10, стр. 8; факс: +7(499)245-0857; e-mail: alexey.lagunin@ibmc.msk.ru

⁴ Башкирский государственный медицинский университет, Россия, 450000, Уфа, ул. Ленина, 3; факс: +7(347)272-3751; e-mail: zarudii.f@yandex.ru

Тимидилатсинтаза является лекарственной мишенью противоопухолевых препаратов класса антиметаболитов. Несмотря на использование таких препаратов в клинике, они обладают рядом сильно выраженных побочных эффектов и накапливаются в тканях. Поэтому необходим поиск и создание новых менее токсичных ингибиторов тимидилатсинтазы. С использованием программы GUSAR 2013 изучена количественная связь “структура — активность” в ряду ингибиторов тимидилатсинтазы антифолатного типа в диапазоне значений величины IC_{50} 0,52 – 24800,00 нмоль/л. На основе MNA- и QNA-дескрипторов с использованием самосогласованной регрессии, построены статистически значимые модели QSAR для прогноза численных значений IC_{50} , которые характеризуются высокой точностью прогноза для структур обучающих и тестовых выборок (R^2_{train} : 0,855 – 0,922; R^2_{train} : 0,810 – 0,895; R^2_{test} : 0,734 – 0,790; R^2_{test} : 0,800 – 0,835).

Ключевые слова: ингибиторы тимидилатсинтазы антифолатного типа; QSAR; GUSAR 2013; QNA- и MNA-дескрипторы; анализ взаимосвязи структура — активность.

Тимидилатсинтаза (ТС, EC2.1.1.45) представляет собой бисубстратный фермент, в котором происходит синтез деокситимидинмонофосфата — нуклеотида, необходимого для синтеза ДНК [1 – 4]. Повышенная активность ТС наблюдается в опухолевых клетках вследствие их высокой скорости роста и развития [3, 4], поэтому этот фермент является лекарственной мишенью противоопухолевых препаратов из класса антиметаболитов. Однако используемые в настоящее время в медицинской практике противоопухолевые препараты, прямо или косвенно снижающие активность ТС, как например, метатрексат и ралтитрексид, обладают рядом сильно выраженных побочных эффектов. Кроме того, эти препараты и их аналоги накапливаются в тканях, что дополнительно усиливает их токсические свойства [5 – 7]. В связи с этим поиск биологически активных веществ, способных ингибировать активность ТС, а, следовательно, замедлить процессы биосинтеза ДНК в опухолевых клетках, является важной практической задачей медицинской химии, направленной на разработку эффективных лекарственных средств. Решение данной задачи исключительно на основе эмпирического анализа биологических данных без привлечения методов компьютерной химии представляет собой трудновыполнимую задачу и требует значительных временных и материальных затрат [8]. Вместе с тем привлечение методов виртуального скрининга, базирующихся на анализе взаимосвязей “структура — активность” — (Q)SAR (Quantitative Structure — Activity Relationships) [9 – 12], позволит на доклинической стадии отобрать соединения-хиты с заданным профилем активности из вирту-

альных библиотек и баз данных для биологических испытаний в условиях *in vivo* [11, 12]. Этот подход может существенно сократить временные и материальные затраты на поиск и разработку потенциальных ингибиторов ТС антифолатного типа.

В этой связи целью настоящей работы было построение и валидация QSAR-моделей селективных ингибиторов ТС для производных хиназолина с общими структурными формулами I – VI, изображенными на рис. 1, на основе двухмерного представления их структурной формулы. Эти соединения обладают высоким структурным сходством с коферментом ТС, 5,10-метилентетрагидрофолатом, благодаря чему перспективны для разработки на их основе ингибиторов данного фермента.

Экспериментальная часть

Количественный анализ взаимосвязей “структура — активность” для ингибиторов ТС проводили с использованием компьютерной программы GUSAR 2013 (General Unrestricted Structure Activity Relationships) [9 – 11]. Построение моделей QSAR выполняли в несколько этапов по методике, приведенной ниже и подробно описанной в работах [13, 14].

Обучающие (OB1 – OB3), а также внутренние тестовые выборки (TB1 – TB2) для построения моделей QSAR (M1 – M9) и их валидации сформированы на основе литературных данных [15 – 20], представленных в выборках B1 – B3 в соответствии со схемой, изображенной на рис. 2.

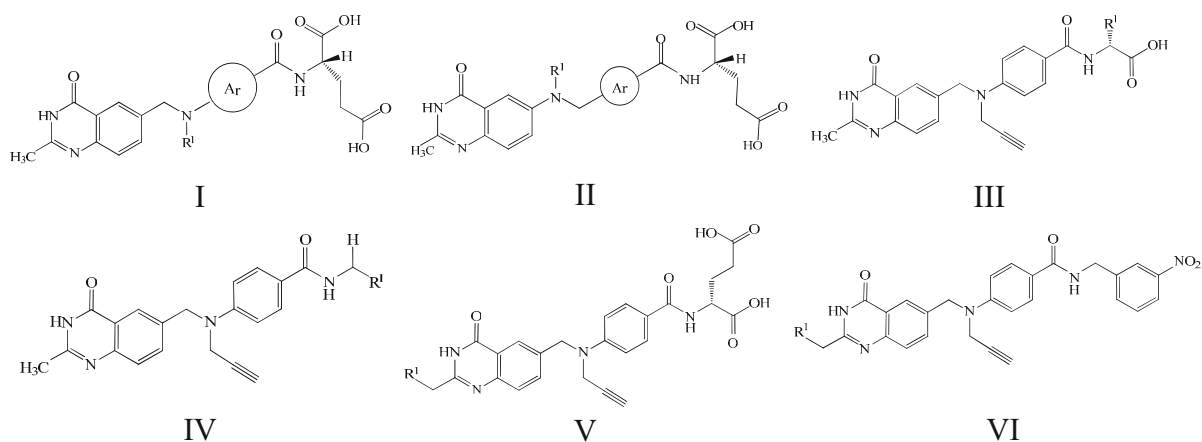


Рис. 1. Общие структурные формулы моделируемых ингибиторов ТС антифолатного типа.

Обучающая выборка OB1, предназначенная для построения моделей QSAR M1 – M3, сформирована на основе выборки B1 и включает в себя 196 структур ингибиторов ТС антифолатного типа. Численные значения ингибиторной активности IC_{50} для этих соединений экспериментально измерены авторами работ [15 – 20]. Выборки B2 и B3 получены в результате разбиения выборки B1, предварительно ранжированной по возрастанию значений IC_{50} , в соотношении 2:1. В дальнейшем на основе выборок B2 и B3 формировали

обучающую выборку OB2 и тестовую выборку TB1 соответственно. Выборка B3 содержит 66 соединений, лежащих в широком диапазоне активности, что позволяет ее использовать в качестве обучающей выборки OB3 для формирования моделей M7 – M9. В этом случае в качестве тестовой выборки TB2 для валидации моделей M7 – M9 выбрали структуры выборки B2. Дополнительно прогностическую способность моделей M1 – M3 оценивали на структурах внешней тестовой выборки TB3, которая содержит 16 производных хиναзолина, близких по строению к соединениям выборок B1 – B3. Ингибиторная активность соединений тестовой выборки TB3 экспериментально изучена в работе [21] в тех же условиях, что и для соединений, вошедших в обучающие выборки OB1 – OB3. Для построения моделей QSAR M1 – M9 данные IC_{50} в выборках B1 – B3 (в моль/л) трансформировали в величины pIC_{50} по формуле:

$$pIC_{50} = -\log_{10}(IC_{50}).$$

Для создания моделей QSAR с целью выполнения по ним в дальнейшем количественного прогноза ингибиторной активности производных хинаязолина в отношении ТС использовали программу GUSAR 2013. Идеология построения моделей QSAR с использованием данной программы детально описана в работах [11 – 14]. Для описания структур соединений и построения моделей QSAR использует 2 типа атомоцентрических дескрипторов: подструктурные — MNA (Multilevel Neighbourhoods of Atoms) и электро-топологические — QNA (Quantitative Neighbourhoods of Atoms) [9 – 14]. Эти типы дескрипторов рассчитывались автоматически из структурных формул химических соединений, представленных в 2D-формате, с учетом валентности и парциальных зарядов содержащихся в них атомов. Учет специфики типов связи с точки зрения стереохимии не предусмотрен. В данном исследовании использовали 3 варианта построения консенсус-моделей QSAR: 1) комбинация полного набора регрессионных уравнений, построенных на осно-

Таблица 1
Статистические характеристики и оценка показателей точности предсказаний значений pIC_{50} для ингибиторов ТС по консенсус-моделям M1 – M9

Обучающая выборка	Модель	N	R_{OB}^2	Q_{OB}^2	F	S. D.	V	R_{TB}^2
QSAR-модели, построенные на основе QNA-дескрипторов								
OB1	M1	196	0,895	0,869	65,386	0,363	21	-
OB2	M4	130	0,857	0,821	42,917	0,420	14	0,800 ^a
OB3	M7	66	0,867	0,818	30,872	0,420	10	0,734 ^b
QSAR-модели, построенные на основе MNA-дескрипторов								
OB1	M2	196	0,896	0,872	54,243	0,364	24	-
OB2	M5	130	0,867	0,832	37,718	0,408	16	0,812 ^a
OB3	M8	66	0,855	0,810	26,404	0,449	9	0,755 ^b
QSAR-модели, построенные на основе QNA- и MNA-дескрипторов								
OB1	M3	196	0,917	0,895	65,352	0,325	25	-
OB2	M6	130	0,895	0,868	47,100	0,364	17	0,835 ^a
OB3	M9	66	0,922	0,891	47,501	0,329	11	0,790 ^b

Примечание. N — число структур в обучающей выборке; R_{OB}^2 — коэффициент детерминации, рассчитанный для соединений из обучающей выборки; R_{TB}^2 — коэффициент детерминации, рассчитанный для соединений из тестовой выборки; Q_{OB}^2 — коэффициент корреляции, рассчитанный на обучающей выборке при скользящем контроле с исключением по одному; F — критерий Фишера; S.D. — стандартное отклонение; V — число переменных в конечном регрессионном уравнении; тире — означает отсутствие данных; ^a — результат прогноза значений pIC_{50} для структур ингибиторов ТС, содержащихся в тестовой выборке TB1; ^b — результат прогноза значений pIC_{50} для структур ингибиторов ТС, содержащихся в тестовой выборке TB2.

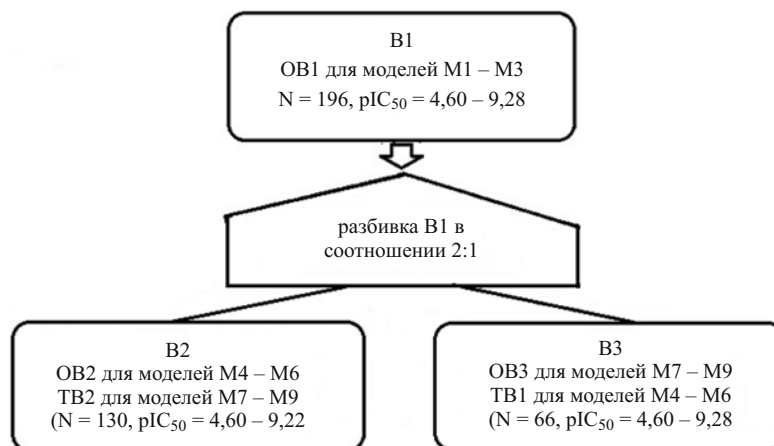


Рис. 2. Схема формирования обучающих и тестовых выборок для построения QSAR моделей M1 – M9, где OB — обучающая выборка; TB — тестовая выборка, N — число соединений, вошедших в соответствующие выборки B1 – B3, обучающие выборки OB1 – OB3 и тестовые выборки TB1 – TB2.

ве QNA-дескрипторов; 2) сочетание полного набора регрессионных уравнений, построенных на основе MNA-дескрипторов; 3) объединение всех регрессионных уравнений, построенных на основе как на QNA-, так и на MNA-дескрипторах. В качестве математического метода для определения оптимального набора дескрипторов и построения моделей QSAR использовали метод самосогласованной регрессии [9 – 11]. Программа позволяет автоматически создавать модели QSAR как на базе единичных регрессионных уравнений, так и на основе набора регрессионных зависимостей, объединенных в единую консенсус-модель. Нами при построении консенсус-моделей учитывались все 360 создаваемых частных регрессионных зависимостей (по 180 для каждого типа дескрипторов). Точность построенных моделей QSAR оценивали по результатам прогноза величины IC_{50} для структур обучающих и тестовых выборок OB1 – OB3, TB1 – TB3. В качестве внутренней валидации использовали скользящий контроль со случайным 20-кратным исключением 20 % структур обучающей выборки. Итоговый прогноз активности для конкретного соединения формировали усреднением спрогнозированных величин IC_{50} отдельных моделей QSAR, входящих в консенсус-модель, что позволяет уменьшить варибельность результатов. Дополнительно по консенсус-модели M3, содержащей 196 структур производных хиназолина, проводили оценку вклада атомов в величину активности ингибиторов TC антифолатного типа. Как сообщалось в работах [13, 14], эта процедура реализуется в программе GUSAR 2013 автоматически при построении моделей QSAR на основе QNA-дескрипторов.

Результаты и их обсуждение

С использованием консенсус-подхода, реализованного в программе GUSAR 2013, проведено моделирование количественной взаимосвязи между строением и активностью ингибиторов TC самцов белых мышей

производными хиназолина с общими структурными формулами I – VI (рис. 1). При этом в зависимости от типа использованных при расчетах дескрипторов (MNA- и QNA-дескрипторов) для каждой из обучающих выборок получено по 3 консенсус модели. Статистические параметры этих моделей, а также характеристики точности предсказанных по ним значений pIC_{50} для ингибиторов TC, входивших в обучающие и тестовые выборки, представлены в табл. 1.

Данные этой таблицы позволяют заключить, что все построенные консенсус-модели QSAR обладают приемлемой предсказательной способностью при их использовании на структурах обучающих выборок OB1 – OB3 ($R^2 > 0,8$; $Q^2 > 0,7$) и внутренних тестовых выборок TB1 и TB2 ($R_{TB}^2 > 0,7$) в диапазоне величины $\Delta pI_{50} = 4,67$ единиц. Среди моделей QSAR M4 – M9, построенных на основе обучающих выборок OB2 – OB3, наибольшей прогностической способностью отличаются модели QSAR M6 и M9. Этот факт объясняется тем, что идеология объединения консенсус-моделей, построенных с использованием либо QNA-, либо MNA-дескрипторов, в единую модель QSAR, в ряде случаев, как например для моделей M3, M6, M9, позволяет снизить варибельность результатов предсказания отдельных моделей. В пользу этого заключения свидетельствуют также результаты наших предыдущих исследований и данные других авторов [9, 11, 22]. Кроме того, очевидным доказательством корректности применения программы GUSAR 2013 для моделирования взаимосвязи между строением и ингибиторной активностью производных хиназолина в отношении TC является достоверный прогноз целевого свойства для соединений тестовой выборки TB2, включающей 130 соединений, по моделям M7 – M9, построенных на основе обучающей выборки OB3, которая содержит 66 структур ингибиторов TC.

В дальнейшем модели QSAR M1 – M3, построенные на основе максимального набора ингибиторов TC, использованы для прогноза численных значений pIC_{50} для структур внешней тестовой выборки TB3, содер-

жащей структурные аналоги производных хиназолина с общими структурными формулами I–VI (рис. 1). Как видно из данных табл. 2, консенсус-модели QSAR M1–M3 показали довольно высокую прогностическую способность на структурах внешней тестовой выборки ТВЗ. Расхождения в экспериментальных и предсказанных по этим моделям численных значений IC_{50} не превышают одного порядка ($RMSE < 0,55$). Эти модели могут быть использованы для виртуального скрининга виртуальных библиотек и баз данных с целью поиска новых ингибиторов ТС антифолатного типа на основе производных хиназолина.

Дополнительно был выполнен структурный анализ производных хиназолина с общими структурными формулами I–VI (рис. 1). Анализ влияния вклада различных структурных фрагментов в активность ингибиторов ТС проводили на основании результатов сравнительных исследований, а также данных, полученных с использованием программы GUSAR 2013. Установлено, что в большинстве случаев результаты

анализа взаимосвязи “структура — активность” для соединений с общими структурными формулами I–VI, полученные с использованием программы GUSAR 2013, удовлетворительно согласуются с экспериментальными данными, детально описанными в работах [15–20]. Исключением являются азот-, фтор- и хлорсодержащие соединения, а также структуры ингибиторов ТС с объемными гетероциклическими фрагментами. Мы полагаем, что расхождение результатов структурного анализа, выполненного с использованием GUSAR 2013, для галоген- и азотсодержащих ингибиторов ТС в сравнении с экспериментальными данными обусловлены рядом факторов. Так, например, помимо физико-химических факторов, на основе которых в используемой программе происходит расчет QNA-дескрипторов и оценка их вклада в целевое свойство, на активность ингибиторов ТС большое значение оказывают структурные и стерические факторы. В частности, авторами работ [15–20] показано, что на активность ингибиторов ТС существенное

Таблица 2

Прогноз численных значений pIC_{50} для структур внешней тестовой выборки ТВЗ по консенсус-моделям QSAR M1–M3

Структурная формула соединения	pIC_{50exp}	pIC_{50pred}	Структурная формула соединения	pIC_{50exp}	pIC_{50pred}
	7,00	6,84 ^a 6,91 ^b 7,08 ^c		6,85	7,29 ^a 7,29 ^b 7,17 ^c
	6,80	6,90 ^a 6,84 ^b 6,72 ^c		6,65	6,48 ^a 6,67 ^b 6,63 ^c
	6,65	7,01 ^a 6,82 ^b 6,78 ^c		6,59	6,51 ^a 6,57 ^b 6,52 ^c
	6,47	6,28 ^a 6,55 ^b 6,62 ^c		6,43	6,42 ^a 6,69 ^b 6,57 ^c
	6,32	7,00 ^a 7,16 ^b 7,07 ^c		6,24	7,24 ^a 6,83 ^b 7,07 ^c
	6,21	7,25 ^a 7,09 ^b 7,00 ^c		6,19	6,38 ^a 6,29 ^b 6,25 ^c
	6,11	6,32 ^a 6,35 ^b 6,29 ^c		6,04	6,37 ^a 6,36 ^b 6,28 ^c
	5,89	6,69 ^a 6,32 ^b 6,60 ^c		5,75	6,51 ^a 6,29 ^b 6,52 ^c

^a – Прогноз численных значений IC_{50} выполнен по консенсус-модели QSAR M1;

^b – прогноз численных значений pIC_{50} выполнен по консенсус-модели QSAR M2;

^c – прогноз численных значений pIC_{50} выполнен по консенсус-модели QSAR M3.

влияние оказывает ориентация ароматических фрагментов в активном центре данного фермента, природа ациклического линкера, ароматических фрагментов Ar и концевых заместителей R¹, в том числе связанных и с асимметрическим атомом углерода (рис. 1). Немаловажное значение на активность ингибиторов ТС оказывает способность заместителей в ароматическом фрагменте и других концевых групп взаимодействовать в активном центре фермента с близлежащими аминокислотными остатками. Как показано в работах [15 – 20], нарушение планарного расположения ароматического фрагмента за счет введения в него объемных функциональных групп будет способствовать снижению целевого свойства. Все эти факторы, как и оптическая активность моделируемых молекул, полностью не учитываются в программе GUSAR 2013. Мы полагаем, что эти факторы оказывают существенное влияние на активность ингибиторов ТС в ряду производных хиначолина. Однако важным достоинством программы GUSAR 2013 является заложенная в ней способность учитывать влияние природы структурных заместителей и способа их сочетания друг с другом. Так, консенсус-модели QSAR, построенные на основе MNA-дескрипторов, которые относятся к классу структурных дескрипторов, характеризуются довольно высокой точностью, а этот факт, в свою очередь, является очевидным доказательством того, что природа структурных фрагментов вносит значительный вклад в активность.

Используемый подход позволяет с высокой степенью надежности моделировать активность ингибиторов ТС в ряду производных хиначолина с целью разработки новых ингибиторов данного фермента антифолатного типа. Для виртуального скрининга предпочтительнее использовать модель M3, так как она построена на основе наибольшего количества структур (196 структур) производных хиначолина с выраженным ингибирующим действием в отношении тимидилатсинтазы. В то же время достаточно точный прогноз значений IC₅₀ для выборок B2 – B3 по моделям M4 – M9 свидетельствует о хорошей моделируемости целевого свойства с помощью подхода, заложенного в программе GUSAR 2013.

Работа выполнена при финансовой поддержке выполняемого в рамках госзадания и Программы фундаментальных исследований государственных академий наук на 2013 – 2020 гг. (Лагунин А. А.).

ЛИТЕРАТУРА

1. Liu, J. C. Schmitz, X. Lin, et al., *Biochim. Biophys. Acta*, **1587**(2 – 3), 174 – 182 (2002).
2. P. R. Subbarayan, K. Lee, B. Ardalani, *Anticancer Res.*, **30**(4), 1157 – 1162 (2010).
3. J. A. van der Zee, C. H. J. van Eijck, H. van Dekken, et al., *Eur. J. Surg. Oncol. (EJSO)*, **38**(11), 1058 – 1064 (2012).
4. O. M. H. Salo-Ahen, A. Tochowicz, Pozzi C., D. Cardinale, et al., *J. Med. Chem.*, **58**(8), 3572 – 3581 (2015).
5. J. Walling, *Invest. New Drugs*, **24**, 37 – 77 (2006).
6. E. Chu, *J. Biol. Chem.*, **265**(15), 8470 – 8478 (1990).
7. V. J. Chen, *British J. Cancer*, **78**(3), 27 – 34 (1998).
8. Д. А. Филимонов, В. В. Поройков, *Рос. хим. ж.*, **50**(2), 66 – 75 (2006).
9. A. V. Zakharov, A. A. Lagunin, D. A. Filimonov, et al., *Chem. Res. Toxicol.*, **25**(11), 2378 – 2385 (2012).
10. И. А. Таипов, В. Р. Хайруллина, А. Я. Герчиков и др., *Вестн. Башкир. универ.*, **17**(2), 886 – 891 (2012).
11. D. A. Filimonov, A. V. Zakharov, A. A. Lagunin, et al., *SAR and QSAR in Environmen. Res.*, **20**(7 – 8), 679 – 709 (2009).
12. A. V. Zakharov., E. V. Varlamova, A. A. Lagunin, et al., *Mol. Pharmaceut.*, **13**(2), 545 – 556 (2016).
13. В. Р. Хайруллина, А. Я. Герчиков, А. А. Лагунин и др., *Биохимия*, **80**(1), 96 – 110 (2015).
14. В. Р. Хайруллина, А. Я. Герчиков, Ф. С. Зарудий и др., *Вестн. Башкир. универ.*, **19**(2), 417 – 422 (2014).
15. P. J. Marsham, A. L. Jackman, J. Oldfield, et al., *J. Med. Chem.*, **33**(11), 3072 – 3078 (1990).
16. P. J. Marsham, L. R. Hughes, A. L. Jackman, et al., *J. Med. Chem.*, **34**(5), 1594 – 1605 (1991).
17. P. J. Marsham, A. L. Jackman, A. J. Hayter, et al., *J. Med. Chem.*, **34**(7), 2209 – 2218 (1991).
18. P. J. Marsham, A. L. Jackman, A. J. Barker, et al., *J. Med. Chem.*, **38**(6), 994 – 1004 (1995).
19. P. J. Marsham, J. M. Wardleworth, F. T. Boyle, et al., *J. Med. Chem.*, **42**(19), 3809 – 3820 (1999).
20. L. F. Hennequin, F. T. Boyle, J. M. Wardleworth, et al., *J. Med. Chem.*, **39**, 695 – 704 (1996).
21. L. R. Hughes, A. L. Jackman, J. Oldfield, et al., *J. Med. Chem.*, **33**(11), 3060 – 3067 (1990).
22. J. C. Dearden, M. T. Cronin, K. L. Kaiser, *SAR QSAR Environ Res.*, **20**(3 – 4), 241 – 266 (2009).

Поступила 15.01.16

QSAR MODELLING OF THYMIDYLATE SYNTHASE INHIBITORS IN A SERIES OF QUINAZOLINE DERIVATIVES

V. R. Khairullina¹, A. Ya. Gerchikov¹, A. A. Lagunin^{2,3}, and F. S. Zarudii⁴

¹ Department of Chemistry, Bashkir State University, Ufa, Bashkortostan, 450076 Russia

² N. I. Pirogov National Research Medical University, Moscow, 117997 Russia

³ V. N. Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, 119121 Russia

⁴ Bashkir State Medical University, Ufa, Bashkortostan, 450008 Russia

* e-mail: Veronika1979@yandex.ru

Thymidylate synthase (TS) is a target for anticancer drugs of the antimetabolite class. Although these drugs are used in clinics, they exhibit a number of strongly pronounced side effects and are accumulated in tissues. Therefore, a search for and creation of new less toxic inhibitors of TS are necessary. For this purpose, analysis of quantitative “structure – activity” relationship (QSAR) in a series of well-known TS inhibitors of the antifolate type with IC₅₀ values in a range of 0.52 – 24800.00 nmol/L was performed using GUSAR 2013 software for establishing their important structural features. Statistically significant QSAR models characterized by high accuracy of predicting IC₅₀ values in the training and test sets (Rtrain2: 0.855 – 0.922; Qtrain2: 0.810 – 0.895; Rtest12: 0.734 – 0.790; Rtest22: 0.800 – 0.835) were constructed using self-consistent regression based on MNA and QNA descriptors.

Keywords: thymidylate synthase inhibitors; antifolate type inhibitors; quantitative structure – activity relationship; GUSAR 2013; QNA descriptors; MNA-descriptors.