

© Коллектив авторов

КОНСЕНСУСНАЯ АНСАМБЛЕВАЯ НЕЙРОСЕТЕВАЯ МУЛЬТИТАРГЕТНАЯ МОДЕЛЬ RAGE-ИНГИБИРУЮЩЕЙ АКТИВНОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ

П.М. Васильев*, А.А. Спасов, А.Н. Кочетков, М.А. Перфильев, А.Р. Королева

Волгоградский государственный медицинский университет,
400131 Волгоград, пл. Павших борцов, 1; *эл. почта: pvassiliev@mail.ru

Передача активации рецепторов RAGE по сигнальному пути RAGE-NF-κB является одним из механизмов возникновения воспалительных реакций, вызывающих тяжёлые осложнения при сахарном диабете. Ингибиторы RAGE являются перспективными фармакологическими соединениями, что требует разработки новых предсказательных моделей. На основе методологии искусственных нейронных сетей построена консенсусная ансамблевая нейросетевая мультитаргетная модель, описывающая зависимость уровня RAGE-ингибирующей активности от аффинности соединений к 34 белкам-мишеням сигнального пути RAGE-NF-κB. Для этого на основе ранее созданной базы данных по трёхмерным моделям релевантных биомишеней была сформирована расширенная база данных по валидным трёхмерным моделям белков-мишеней сигнальной цепочки RAGE-NF-κB. Выполнен ансамблевый молекулярный докинг известных RAGE-ингибиторов из верифицированной базы данных в сайты моделей белков-мишеней, определены минимальные энергии докинга для каждого соединения в отношении каждой мишени и сформирована расширенная обучающая выборка для нейросетевого моделирования. С использованием семи вариантов обучения на основе искусственных многослойных перцептронных нейронных сетей построены три ансамбля классификационных решающих правил для прогноза трёх уровней RAGE-ингибирующей активности по расчётной аффинности соединений к значимым белкам-мишеням сигнального пути RAGE-NF-κB. С применением простого консенсуса второго уровня выполнена оценка прогностической способности созданной модели, показана её высокая точность и статистическая значимость. С помощью полученной консенсусной ансамблевой нейросетевой мультитаргетной модели проведен виртуальный скрининг новых соединений различных химических классов. Перспективные вещества синтезированы и направлены на экспериментальное изучение.

Ключевые слова: RAGE-ингибиторы; мультитаргетная аффинность; молекулярный докинг; искусственные нейронные сети; консенсусная ансамблевая модель; виртуальный скрининг

DOI: 10.18097/PBMC20216703268

ВВЕДЕНИЕ

По данным ВОЗ на 15 апреля 2021 г. сахарный диабет (СД) занимает девятое место в мире как причина смерти. В 2019 г. диабет был причиной смертности 1,4 млн. человек [1]. Приводящие к смерти пациента тяжёлые осложнения СД (нефро-, кардио-, ретино-, нейро- и ангиопатии) возникают вследствие хронической гипергликемии, которая порождает неферментативное гликирование белков, в результате чего образуются конечные продукты гликирования (AGE), взаимодействующие со своими специфическими рецепторами RAGE [2]. Передача активации RAGE по сигнальному пути RAGE-NF-κB [3] является одним из механизмов возникновения воспалительных реакций, вызывающих тяжёлые осложнения СД. NF-κB является транскрипционным фактором гена *AGER*, экспрессия которого приводит к синтезу новых AGE рецепторов, что за счёт положительной обратной связи существенно усиливает воспалительные процессы [4, 5]. В настоящее время обусловленная активацией RAGE связь между развитием сахарного диабета и болезни Альцгеймера является достоверно установленным фактом [5]. Мультитаргетные ингибиторы RAGE с пострецепторной активностью, действующие как на сам рецептор, так и на киназы, вовлечённые в передачу сигнала по пути RAGE-NF-κB, являются новыми перспективными фармакологическими соединениями для лечения указанных патологий.

В нашем предыдущем исследовании [6] была построена мультитаргетная модель зависимости RAGE-ингибирующей активности от расчётной аффинности соединений к 22 значимым белкам-мишеням сигнального пути RAGE-NF-κB, основанная на использовании единичных нейросетей для прогноза каждой из трёх градаций активности.

Целью настоящего исследования была разработка более мощной консенсусной ансамблевой нейросетевой мультитаргетной модели для прогноза уровня RAGE-ингибирующей активности по расчётной аффинности соединений уже к 34 значимым белкам-мишеням сигнальной цепи RAGE-NF-κB и виртуальный скрининг с использованием этой модели новых высокоактивных RAGE-ингибиторов.

Для достижения указанной цели необходимо было решить следующие задачи:

1. сформировать расширенную базу данных по валидированным 3D-моделям белков-мишеней, релевантных сигнальному пути RAGE-NF-κB;
2. выполнить молекулярный докинг известных RAGE-ингибиторов в сайты дополнительно валидированных 3D-моделей релевантных белков-мишеней, сформировать расширенную обучающую выборку;
3. построить консенсусную модель зависимости RAGE-ингибирующей активности соединений от их расчётной аффинности к релевантным



белкам-мишеням в виде трёх ансамблей искусственных нейронных сетей, оценить точность полученной модели;

4. с использованием построенной консенсусной ансамблевой нейросетевой мультитаргетной модели выполнить виртуальный скрининг соединений различных химических классов с прогнозируемой высокой RAGE-ингибирующей активностью, и выявить наиболее перспективные для экспериментального изучения вещества.

МЕТОДИКА

Формирование верифицированной базы данных по структуре и уровню активности известных RAGE-ингибиторов

Данный обязательный этап по подготовке обучающей выборки был выполнен ранее согласно методике, детально описанной в нашем предыдущем исследовании [6]. На основе информации, полученной из ChEMBL (<https://www.ebi.ac.uk/chembl/>), BindingDB (<https://www.bindingdb.org/bind/index.jsp>) и PubChem (<https://pubchem.ncbi.nlm.nih.gov/>), сформирована база данных, включающая 183 известных соединения с проверенными структурными формулами и унифицированным уровнем RAGE-ингибирующей активностью, которая содержит 38 структур с высокой активностью (метка градации High), 61 структуру с умеренной активностью (метка Moderate), 39 структур с низкой активностью (Low) и 45 неактивных структур (Inactive). На указанную базу данных получено свидетельство о государственной регистрации [7].

Формирование расширенной базы данных по валидированным 3D-моделям белков-мишеней, релевантных сигнальному пути RAGE-NF-κB

Ранее нами было найдено [8], что в передаче сигнала по 14 основным узлам цепочки RAGE-NF-κB участвуют 34 белка-мишени, соотносимые с осложнениями СД или с болезнью Альцгеймера. В нашей предыдущей работе [6] с использованием специально разработанной методики была сформирована валидированная база данных по 3D-моделям 22 основных белков-мишеней, релевантных пути RAGE-NF-κB.

Для 12 других биомишеней NF-κB2, RelA, AKT3, RASA1, PIK3CB, PLCB1, PLCG1, PLCD1, DIAPH1, NOX1, NOX2, NOX4 по этой же методике [6] были найдены ещё 36 валидных 3D-моделей. Ранее эти мишени не были включены в расчёты из-за недостатка информации о локализации специфических сайтов и о структуре референсных ингибиторов (препаратов сравнения), необходимых для валидации 3D-моделей.

Дополнительная валидация включала следующие основные этапы.

1. В базах данных PDBe (<http://www.ebi.ac.uk/pdbe/>) и ModBase (<https://modbase.compbio.ucsf.edu/modbase.cgi/>) с проверкой на адекватность в программе VMD 1.9.2 [9] были найдены 3D-модели указанных 12 белков-мишеней *Homo sapiens*.

2. Для каждого из этих 12 видов биомишеней были определены местоположение специфического ингибирующего сайта связывания и ключевые связывающие аминокислоты, ответственные за фиксацию лигандов в этом сайте. При этом учитывали: 1) найденные в доступных литературных источниках экспериментальные данные об активности изоформ с различными точечными мутациями в специфическом сайте ингибирования; 2) полученные с помощью программ LigPlot+ 1.4.5 [10], LigandScout 4.2.1 [11] и BestSites 18.3.19 результаты анализа экспериментальных рентгеноструктурных 3D-моделей комплексов конкретного белка с его известными ингибиторами; 3) взятые из открытых публикаций данные по молекулярному моделированию связывания лигандов с этим белком.

3. В поисковых системах IUPHAR/BPS (<http://www.guidetopharmacology.org/>) и ChEMBL (<https://www.ebi.ac.uk/chembl/>) для каждого белка-мишени были найдены референсные ингибиторы, для которых с помощью системы Microcosm BioS 18.1.9 [12] было найдено такое же число структурно наиболее сходных с ними неактивных соединений.

4. Для всех найденных референсных ингибиторов и неактивных соединений были построены оптимизированные 3D-модели. Методами молекулярной механики в программе MarvinSketch 17.1.23 [13] для каждой структуры были сформированы по 10 конформеров с наименьшей энергией, которые затем были оптимизированы с помощью программы MORAC2016 [14] с использованием полуэмпирического квантово-химического метода PM7 и для каждого соединения отобрано по одному конформеру с наименьшей общей энергией.

5. Для каждой из найденных 3D-моделей дополнительно вводимых биомишеней выполняли с помощью программы AutoDock Vina 1.1.2 [15] ансамблевый молекулярный докинг в ингибирующий сайт белка соответствующих референсных и неактивных соединений. Докинг выполняли с пятикратным повтором для каждого лиганда, каждый раз в десяти биологически активных конформациях, вычисляя по 50 полученным значениям минимальную энергию докинга ΔE. Отдельно для каждой 3D-модели по всем препаратам сравнения и по всем неактивным соединениям вычисляли два средних значения минимальных энергий докинга ΔE. Метрикой валидности 3D-модели служила разность этих двух средних значений ΔE для референсных и неактивных соединений.

В итоге была сформирована расширенная валидированная база данных по 102 3D-моделям 34 белков-мишеней, релевантных сигнальному пути RAGE-NF-κB.

Молекулярный докинг известных RAGE-ингибиторов в сайты дополнительно валидированных 3D-моделей белков-мишеней, релевантных сигнальному пути RAGE-NF-κB

Оптимальные 3D-конформации 183 известных RAGE-ингибиторов были рассчитаны ранее при проведении предыдущего исследования

с использованием программ MarvinSketch 17.1.23 [13] и MORAC2016 [14]. Использованная методика кратко изложена выше и детально описана в нашей работе [6].

Ансамблевый докинг 3D-моделей известных RAGE-ингибиторов в сайты 3D-моделей 22 релевантных белков-мишеней был также выполнен ранее с помощью программы AutoDock Vina 1.1.2 [15]. Использованная методика кратко изложена выше и детально описана в нашей работе [6].

В настоящем исследовании для каждого из 12 дополнительно вводимых в моделирование белков-мишеней в каждую из трёх соответствующих ему валидных 3D-моделей по этой же методике выполняли ансамблевый молекулярный докинг в ингибирующий сайт всех известных RAGE-ингибиторов.

По массиву из 933300 значений энергий докинга ΔE была сформирована матрица аффинности 183 известных RAGE-ингибиторов в отношении 34 релевантных белков-мишеней сигнальной цепочки RAGE-NF- κ B, включающая 6222 минимальных значения ΔE .

Построение консенсусной ансамблевой нейросетевой мультитаргетной модели зависимости RAGE-ингибирующей активности соединений от их расчетной аффинности к белкам-мишеням сигнального пути RAGE-NF- κ B

Образующие ансамбль решающих правил классификационные модели для прогноза уровня RAGE-ингибирующей активности были построены по методике искусственных многослойных перцептронных нейронных сетей с использованием модуля Statistica Neural Networks программы Statistica 8.0 [16] по описанной ниже схеме.

Была расширена сформированная в нашем предыдущем исследовании [6] обучающая выборка, включающая для известных RAGE-ингибиторов три переменных с метками наличия/отсутствия объединённых уровней активности: h/nh — есть высокая активность / нет высокой активности; hm/nhm — есть высокая или умеренная активность / нет высокой или умеренной активности; a/na — есть активность / нет активности; а также 22 переменных, содержащих значения ΔE аффинности этих соединений к релевантным белкам-мишеням. При расширении добавлены ещё 12 переменных, отражающих значения аффинности ΔE для известных RAGE-ингибиторов к дополнительно вводимым в моделирование белкам-мишеням и введены семь переменных сэмпинга. В машинном обучении сэмпинг — это процесс формирования из исходного набора данных выборок, представляющих интерес для анализа. В настоящем исследовании каждая переменная сэмпинга задаёт семь отличных друг от друга вариантов из обучающей, тестовой и валидационной подвыборок (в соотношении 5:1:1), используемых для обучения нейронных сетей.

Полученная в итоге расширенная классификационная обучающая выборка по уровню активности и значениям аффинности

известных RAGE-ингибиторов включает данные по 183 соединениям, для каждого из которых определено 45 переменных: шифр, метки трёх объединённых градаций активности (выступали как выходные нейроны), энергии докинга в 34 мишени (служили входными нейронами), семь переменных сэмпинга.

В качестве исходной архитектуры нейронной сети был выбран двухслойный перцептрон MLP $k-m-2$ с узким горлом. Здесь k — число входных нейронов, в данном случае 34; m — число скрытых нейронов, устанавливается программой от 3 до 33, поскольку $2 < m < k$. В программе Statistica 8.0 [16] при построении перцептронных сетей для скрытого и выходного слоев используются пять наиболее распространённых активационных функций Identity, Logistic, Tanh, Exponential, Sine, попарный перебор которых также осуществляется этой программой.

Обучение нейросетей проводили для каждой пары объединённых градаций с использованием алгоритма обратного распространения ошибок, перебором двух видов функции ошибок (среднеквадратичного отклонения и кросс-энтропии), пяти различных активационных функций для скрытого и выходного слоя нейронов, с использованием семи вариантов сэмпинга.

С целью достижения наилучшего результата обучения число сетей для каждой из трёх пар градаций активности и каждого из семи вариантов сэмпинга было установлено в 4000 с автоматическим отбором в каждом цикле обучения 1000 наилучших сетей. После окончания обучения для заданного уровня активности и заданного варианта сэмпинга из 1000 лучших отобранных программой нейросетей по совокупности характеристик точности обучения, тестирования и валидации вручную отбирали 100 лучших, из которых с использованием ROC-анализа вручную отбирали одну наилучшую. Всего было обучено более 100 тыс. нейронных сетей.

В результате проведенных расчетов для трёх уровней RAGE-ингибирующей активности была сформирована консенсусная прогнозная модель, состоящая из трёх ансамблей по семь нейросетей в каждом (всего 21 нейросеть).

Оценка точности консенсусной ансамблевой нейросетевой мультитаргетной модели RAGE-ингибирующей активности

Консенсусную оценку уровня RAGE-ингибирующей активности с использованием полученного ансамбля нейронных сетей выполняли следующим образом.

Для прогнозируемого соединения по каждому уровню активности с использованием семи нейросетей выполняли расчёт наличия/отсутствия данного уровня активности, в результате чего получали спектр из трёх наборов альтернативных прогнозных оценок. Обработку этого спектра первичных оценок проводили с использованием простого невзвешенного консенсуса, где результатом голосования по семи исходам принимался такой, для которого наблюдалось четыре или более совпадающих оценок.

Например, для известного высокоактивного RAGE-ингибитора с шифром RAGE-0024 спектр первичных прогнозных оценок будет выглядеть так: “h h nh nh h h h”, “hm hm hm hm hm hm”, “a a a a a a”, что соответствует высокой предсказанной активности. В рамках консенсуса 1-го уровня наличие высокой активности будет соответствовать консенсус 5 положительных оценок из 7, а наличие выраженной активности и просто наличие активности — 7 положительных оценок из 7.

Суммарные консенсусные оценки первого уровня должны быть непротиворечивы. В данном примере это правило выполняется. В соответствии с консенсусными оценками первого уровня “h hm a”, в рамках консенсуса второго уровня вещество RAGE-0024 по прогнозу должно обладать высокой активностью с порядком консенсуса пять (минимальное число совпадающих для целевого уровня активности оценок).

Для известного умеренно активного RAGE-ингибитора с шифром RAGE-0001 получен следующий спектр первичных прогнозных оценок: “nh nh nh nh nh nh”, “hm hm hm hm hm hm”, “a a a a a a”. Консенсус 1-го уровня будет “nh hm a” и в соответствии с консенсусом 2-го уровня вещество RAGE-0024 по прогнозу должно обладать умеренной активностью с порядком консенсуса семь.

Низкоактивному соединению соответствует прогнозный консенсус 1-го уровня “nh nhm a”, неактивному соединению — “nh nhm na”. Все остальные наборы консенсусных оценок 1-го уровня противоречивы и при прогнозе игнорируются — например, “h nhm na”.

В соответствии с описанной схемой, на объединённой выборке для уровней активности High, High or Moderate и Active для каждой нейросети в отдельности и для консенсусного

решающего правила была выполнена оценка общей точности прогноза F_0 , точности прогноза активных соединений F_a (чувствительность), точности прогноза неактивных соединений F_n (специфичность) и по данным ROC-анализа рассчитана площадь под кривой AUC.

Виртуальный скрининг с использованием консенсусной ансамблевой нейросетевой мультитаргетной модели новых RAGE-ингибиторов с высокой активностью

Виртуальный скрининг высокой RAGE-ингибирующей активности был выполнен по вышеописанной методике для 472 соединений — производных 25 различных химических классов. По результатам прогноза были отобраны перспективные структуры с непротиворечивыми оценками наличия высокой или умеренной активности.

Все расчёты были выполнены на суперкомпьютерных вычислительных комплексах гибридной архитектуры общей пиковой производительностью ~8,9 TFLOPS.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Сигнальный путь RAGE-NF-κB с 14 основными узлами приведен на рисунке 1. Общая схема сигнального пути была взята из KEGG [3], содержательная детализация числа релевантных биомешеней в узлах была выполнена в нашем предыдущем исследовании [8].

Для этих 14 узлов ранее нами было найдено 34 биомешени, релевантных осложнениям СД и болезни Альцгеймера [8]: NF-κB1, NF-κB2, RelA, ERK1, ERK2, MAPK13, MAPK14, JNK1, JNK2, JNK3, AKT1, AKT2, AKT3, RASA1, CDC42, RAC1, PRKCA, PRKCB, PRKCG, PRKCD, PRKCE, PRKCH, PRKCCQ, ROS1, PIK3CA, PIK3CB, PLCB1, PLCG1, PLCD1, DIAPH1, NOX1, NOX2, NOX4, RAGE.

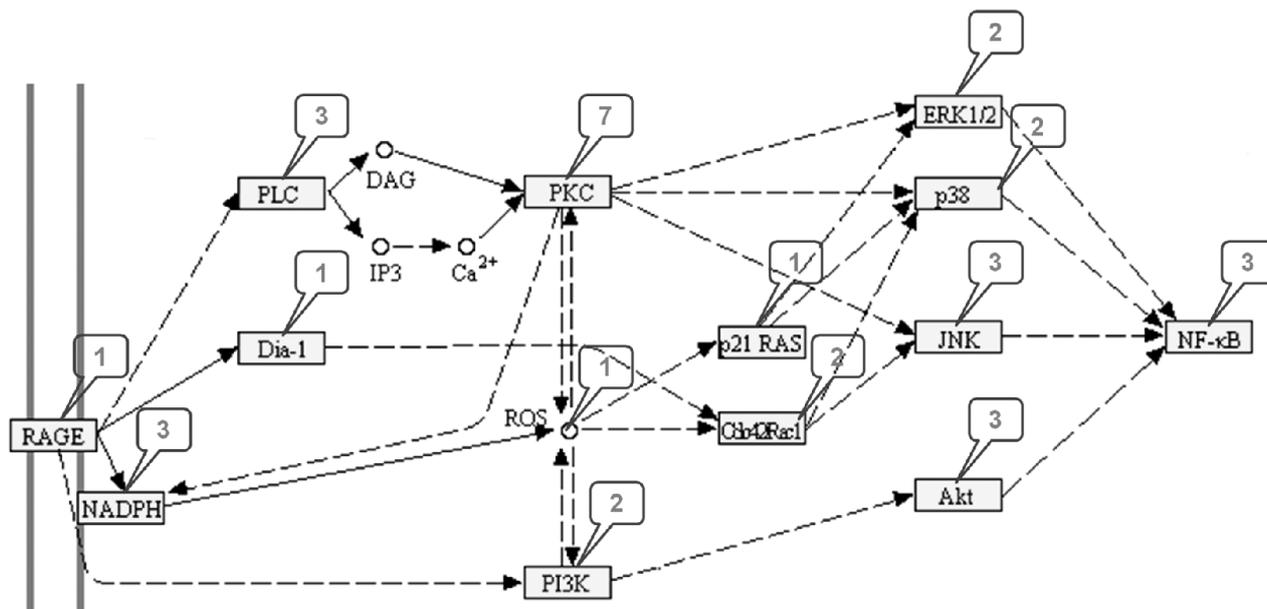


Рисунок 1. Сигнальный путь RAGE-NF-κB: цифрами указано число релевантных белков-мишеней для данного узла сигнальной цепи.

КОНСЕНСУСНАЯ МУЛЬТИТАРГЕТНАЯ МОДЕЛЬ RAGE-ИНГИБИРОВАНИЯ

Ранее в нашем предыдущем исследовании [6] были валидированы 3D-модели для 22 из 34 перечисленных выше биомишеней.

В настоящем исследовании дополнительно валидированы 3D-модели 12 релевантных биомишеней, для которых всего было найдено 20 экспериментальных и 179 теоретических адекватных 3D-моделей, 147 референсных лигандов и 147 структурно сходных с ними неактивных соединений (табл. 1).

В результате проведенной валидации было дополнительно отобрано 36 валидных 3D-моделей (по три модели на каждую вновь вводимую мишень).

Обобщённые результаты ансамблевого молекулярного докинга 183 RAGE-ингибиторов в эти 36 3D-моделей приведены в таблице 2.

Как видно из таблицы 2, для добавленных в моделирование 12 мишеней средние энергии докинга ΔE соединений с различным уровнем RAGE-ингибирующей активности сопоставимы, что хорошо согласуется с результатами, полученными в нашем предыдущем исследовании для докинга в 22 мишени [6]. Для всех попарных сравнений по классам активности для всех 34 мишеней различия по критерию Манна-Уитни статистически не достоверны. Таким образом, энергии докинга соединений в сайты 34 релевантных биомишеней не могут по отдельности, без учёта взаимного влияния этих биомишеней друг на друга, служить достаточно адекватной метрикой для расчётной оценки их экспериментальной RAGE-ингибирующей активности.

Таблица 1. Число дополнительно валидированных 3D-моделей релевантных белков-мишеней сигнального пути RAGE-NF- κ B

Шифр ¹	Название ²	Число 3D-моделей ³		Число соединений ⁴
		PDBe	ModBase	
NF- κ B2	Nuclear factor NF- κ -B p100(p52) subunit	3	19	68
RelA	Transcription factor p65	7	10	20
AKT3	RAC-gamma serine/threonine-protein kinase	—	17	24
RASA1	Ras GTPase-activating protein 1	5	11	6
PIK3CB	Phosphatidylinositol 4,5-bisphosphate 3-kinase C beta	—	16	34
PLCB1	1-Phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-1	—	19	38
PLCG1	1-Phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1	4	13	26
PLCD1	1-Phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-1	—	20	10
DIAPH1	Protein diaphanous homolog 1	—	14	0
NOX1	NADPH oxidase 1	—	7	32
NOX2	NADPH oxidase 2	1	24	24
NOX4	NADPH oxidase 4	—	9	12
Всего		20	179	294

Примечание: 1 — общепринятые сокращенные обозначения белков по данным UniProt; 2 — номенклатурное название, рекомендуемое UniProt; 3 — число моделей: экспериментальных из PDBe и теоретических из ModBase; 4 — число референсов для валидации моделей равно числу структурно сходных с ними неактивных соединений.

Таблица 2. Результаты ансамблевого молекулярного докинга известных RAGE-ингибиторов в сайты дополнительно валидированных 3D-моделей белков-мишеней, релевантных сигнальному пути RAGE-NF- κ B

Шифр	ΔE , ккал/моль ¹				
	High ²	Moderate ³	Low ⁴	Inactive ⁵	δ ⁶
NF- κ B2	-7,56	-7,51	-7,52	-7,59	0,04
RelA	-8,48	-8,57	-8,40	-8,45	0,09
AKT3	-8,01	-7,83	-7,77	-7,96	0,14
RASA1	-8,87	-8,88	-8,61	-8,67	0,17
PIK3CB	-8,21	-8,13	-8,01	-8,10	0,11
PLCB1	-8,73	-8,63	-8,44	-8,69	0,16
PLCG1	-8,59	-8,60	-8,40	-8,40	0,13
PLCD1	-9,40	-9,19	-9,10	-9,05	0,19
DIAPH1	-9,53	-9,31	-9,22	-9,20	0,18
NOX1	-9,53	-9,31	-9,22	-9,20	0,18
NOX2	-7,21	-7,22	-7,13	-7,16	0,05
NOX4	-7,49	-7,37	-7,19	-7,23	0,17

Примечание: 1 — средняя энергия докинга; 2 — для класса соединений с градацией активности High; 3 — для класса соединений с градацией активности Moderate; 4 — для класса соединений с градацией активности Low; 5 — для класса соединений с градацией активности Inactive; 6 — средняя разность значений ΔE при попарном сравнении классов.

В таблице 3 приведены характеристики составляющих прогнозные ансамбли наилучших нейронных сетей, полученных для трёх уровней RAGE-ингибирующей активности при обучении на семи вариантах сэмплинга.

По совокупности показателей точности все нейросети статистически очень высоко достоверны, $p < 0,001$ по биномиальному критерию [17].

Таким образом, следует ожидать, что при совместном использовании этих нейросетей точность консенсусного прогноза может существенно возрасти за счёт взаимной компенсации ошибок, что и было подтверждено последующими расчётами.

В таблице 4 приведены сравнительные характеристики точности отдельных нейронных сетей и консенсусной ансамблевой нейросетевой мультитаргетной модели RAGE-ингибирующей активности при тестировании на объединённой выборке.

Видно, что консенсусный прогноз существенно превышает по точности прогноз с использованием любой из отдельных нейросетей, достигая для двух уровней активности максимально возможной 100%-ной точности. Минимальная значимость для точности прогноза 94,7% составляет по биномиальному критерию [17] $z=8,53$, $p < 1 \cdot 10^{-15}$.

Полученная консенсусная ансамблевая нейросетевая мультитаргетная модель была использована для скрининга соединений с высокой или выраженной RAGE-ингибирующей активностью в базе данных 472 производных 25 различных химических классов.

В качестве примера на рисунке 2 приведены структуры синтезированных производных трёх различных скаффолдов, которые ранее в проведённом нами предварительном исследовании, в соответствии с полученными прогнозными оценками, были позиционированы как источники веществ с высокой RAGE-ингибирующей активностью [18].

Соединение **5** относится к производным урацила [19], соединение **143** — к производным имидазобензимидазола [20], а соединение **462** — к производным дифенилоксида [21].

Результаты ансамблевой нейросетевой оценки спектра прогнозных показателей активности этих трёх соединений приведены в таблице 5.

В рамках консенсуса первого уровня наличие у соединений **5**, **143**, **462** высокой и выраженной активности и просто наличие активности соответствуют от 4 до 6 совпадающих положительных оценок.

Таблица 3. Характеристики ансамблей нейросетей, отражающих мультитаргетную зависимость RAGE-ингибирующей активности соединений от их расчётной аффинности к белкам-мишеням сигнального пути RAGE-NF-κB

Лучшая нейросеть		Точность, %			
Номер ¹	Архитектура	Train ²	Test ³	Val ⁴	AUC ⁵
Уровень активности High					
1	MLP 34-16-2 Tanh Softmax	100,0	84,6	76,9	95,6
2	MLP 34-17-2 Tanh Softmax	98,5	92,3	74,1	87,2
3	MLP 34-14-2 Tanh Softmax	100,0	81,5	73,1	87,8
4	MLP 34-19-2 Tanh Softmax	100,0	84,6	69,2	91,0
5	MLP 34-12-2 Tanh Softmax	95,4	80,8	76,9	85,1
6	MLP 34-10-2 Logistic Softmax	97,7	84,6	84,6	89,0
7	MLP 34-17-2 Tanh Softmax	100,0	88,5	80,8	93,5
Уровень активности High or Moderate					
1	MLP 34-13-2 Tanh Softmax	100,0	88,5	80,8	95,6
2	MLP 34-17-2 Logistic Softmax	100,0	96,2	55,6	93,8
3	MLP 34-17-2 Logistic Softmax	100,0	81,5	73,1	97,0
4	MLP 34-21-2 Tanh Softmax	100,0	76,9	80,8	94,8
5	MLP 34-11-2 Tanh Softmax	100,0	84,6	61,5	92,3
6	MLP 34-21-2 Exponential Softmax	75,6	57,7	61,5	78,0
7	MLP 34-16-2 Tanh Softmax	100,0	80,8	73,1	92,2
Уровень активности Active					
1	MLP 34-9-2 Logistic Softmax	99,2	84,6	80,8	95,3
2	MLP 34-15-2 Tanh Softmax	100,0	92,3	88,9	97,0
3	MLP 34-14-2 Tanh Softmax	100,0	88,9	84,6	95,8
4	MLP 34-8-2 Tanh Softmax	100,0	92,3	76,9	90,8
5	MLP 34-19-2 Logistic Softmax	100,0	80,8	80,8	91,5
6	MLP 34-22-2 Tanh Softmax	99,2	84,6	92,3	95,3
7	MLP 34-13-2 Tanh Softmax	100,0	96,2	88,5	95,1

Примечание: 1 — номер сэмплинга для обучения сетей; 2 — при обучении; 3 — в независимом тестировании; 4 — при валидации; 5 — по данным ROC-анализа на объединённой выборке.

КОНСЕНСУСНАЯ МУЛЬТИТАРГЕТНАЯ МОДЕЛЬ RAGE-ИНГИБИРОВАНИЯ

Таблица 4. Сравнительная точность консенсусной ансамблевой нейросетевой модели зависимости RAGE-ингибирующей активности соединений от их расчётной аффинности к белкам-мишеням сигнального пути RAGE-NF-κB

Номер сети ¹	Точность, %			
	F ₀ ³	F _a ⁴	F _n ⁵	AUC ⁶
Уровень активности High				
1	94,7	94,5	94,5	95,6
2	84,2	96,6	94,0	87,2
3	76,3	97,9	93,4	87,8
4	78,9	97,2	93,4	91,0
5	68,4	96,6	90,7	85,1
6	76,3	98,6	94,0	89,0
7	89,5	97,2	95,6	93,5
Консенсус ²	94,7	100,0	98,9	97,4
Уровень активности High or Moderate				
1	98,0	92,9	95,6	95,6
2	90,9	95,2	92,9	93,8
3	91,9	95,2	93,4	93,5
4	97,0	90,5	94,0	94,8
5	96,0	88,1	92,3	92,3
6	83,8	56,0	71,0	78,0
7	97,0	89,3	93,4	92,2
Консенсус ²	100,0	100,0	100,0	100,0
Уровень активности Active				
1	97,8	84,4	94,5	95,3
2	97,8	95,6	97,3	97,0
3	96,4	95,6	96,2	95,8
4	97,8	88,9	95,6	90,8
5	99,3	80,0	94,5	91,5
6	100,0	84,4	96,2	95,3
7	100,0	91,1	97,8	95,1
Консенсус ²	100,0	100,0	100,0	100,0

Примечание: 1 — номер лучшей сети, полученной на семи вариантах сэмплинга; 2 — простой консенсус результатов прогноза для семи лучших сетей; 3 — общая точность прогноза; 4 — точность прогноза активных соединений (чувствительность); 5 — точность прогноза неактивных соединений (специфичность); 6 — по данным ROC-анализа.

Таблица 5. Ансамблевый прогноз RAGE-ингибирующей активности производных перспективных скаффолдов

Соединение	Результат прогноза ¹							Консенсус ²
	1	2	3	4	5	6	7	
Уровень активности High								
5	nh	nh	h	h	h	h	nh	4h
143	nh	h	h	h	h	h	h	6h
462	h	nh	nh	nh	h	h	h	4h
Уровень активности High or Moderate								
5	nhm	hm	hm	nhm	hm	hm	nhm	4hm
143	nhm	nhm	hm	hm	hm	nhm	hm	4hm
462	hm	hm	nhm	hm	hm	nhm	hm	5hm
Уровень активности Active								
5	a	a	na	a	a	a	a	6a
143	a	a	a	a	na	a	a	6a
462	a	na	a	a	na	a	a	5a

Примечание: 1 — по лучшей сети, полученной для данного варианта сэмплинга; 2 — простой консенсус результатов прогноза для семи лучших сетей.

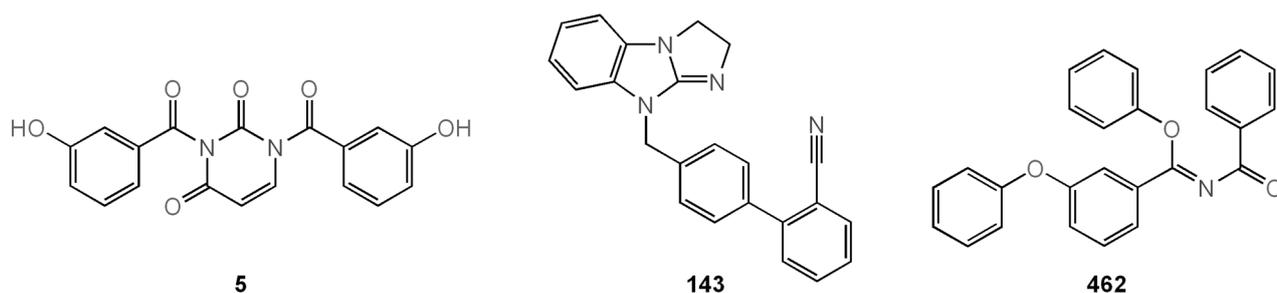


Рисунок 2. Структурные формулы высоко активных по прогнозу RAGE-ингибиторов — производных скаффолдов, перспективных как источники веществ с высокой RAGE-ингибирующей активностью.

Таблица 6. Консенсусный прогноз высокого уровня RAGE-ингибирующей активности производных перспективных скаффолдов

Соединение	Консенсус 1-го уровня ¹			Консенсус 2-го уровня ²	Непротиворечивая оценка активности ³
	High	High or Moderate	Active		
5	4h	4hm	6a	4	High
143	6h	4hm	6a	6	High
462	4h	5hm	5a	4	High

Примечание: 1 — простой консенсус результатов прогноза для семи лучших сетей; 2 — число положительных оценок высокой активности h; 3 — при наличии положительных консенсусных оценок по трём градациям активности.

Результаты прогноза высокого уровня RAGE-ингибирующей активности соединений **5**, **143**, **462** при использовании консенсуса второго уровня приведены в таблице 6.

Согласно данным таблицы 6, все три соединения **5**, **143**, **462** должны проявлять высокую RAGE-ингибирующую активность: спектр прогнозных оценок непротиворечив, для градации High наблюдается консенсус второго уровня, которому соответствуют от 4 до 6 совпадающих положительных оценок.

По результатам виртуального скрининга среди 472 новых производных отобраны 37 соединений с возможной высокой RAGE-ингибирующей активностью и 81 соединение с возможной умеренной активностью. Из них 56 веществ синтезировано и направлено на экспериментальное изучение.

ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

Методом искусственных нейронных сетей построена консенсусная ансамблевая нейросетевая мультитаргетная модель зависимости уровня RAGE-ингибирующей активности соединений от их расчётной аффинности к 34 белкам-мишеням сигнального пути RAGE-NF-κB. Модель включает три ансамбля по семь нейросетей в каждом ансамбле для градаций RAGE-ингибирующей активности “высокая”, “выраженная”, “наличие активности”.

Точность прогноза RAGE-ингибирующей активности с применением классификационной двухуровневой консенсусной ансамблевой модели является статистически достоверной и превышает точность отдельных нейросетей, составляющих прогнозные ансамбли.

С использованием сформированной консенсусной ансамблевой нейросетевой мультитаргетной модели выполнен виртуальный скрининг новых производных различных химических классов, найдены перспективные соединения, половина из которых синтезированы и направлены на экспериментальное тестирование RAGE-ингибирующей активности.

Таким образом, был выполнен полный расчётный комплекс работ по направленному поиску *in silico* системных полифункциональных мультитаргетных ингибиторов рецепторов конечных продуктов гликирования, которые станут основой создания принципиально новых лекарственных препаратов для лечения осложнений при СД и болезни Альцгеймера.

На примере RAGE-ингибирующей активности разработана новая методология построения классификационных моделей для прогноза мультитаргетной активности химических соединений, основанная на двухуровневом консенсусе прогнозных оценок, полученных с использованием нескольких ансамблей нейронных сетей.

Разработанная консенсусная ансамблевая нейросетевая методология может быть использована для поиска высокоактивных соединений с другими видами фармакологической активности, в том числе, имеющими общесистемный характер, такими, как гипогликемическая, психотропная, противовоспалительная и другие.

ФИНАНСИРОВАНИЕ

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект 18-015-00499).

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая статья не содержит описания проведённых авторами исследований с участием людей или использованием животных в качестве объектов.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

ЛИТЕРАТУРА

- World Health Organization (2021) <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- Ansari N.A., Rasheed Z. (2009) Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry, **3**, 335. DOI: 10.1134/S1990750809040027.
- KEGG: AGE-RAGE signaling pathway in diabetic complications — *Homo sapiens* (2017) https://www.genome.jp/kegg-bin/show_pathway?map=hsa04933&show_description=show
- Khan M.I., Rath S., Adhami V.M., Mukhtar H. (2018) Seminars Cancer Biology, **49**, 75-82.
- Kong Y., Wang F., Wang J., Liu C., Zhou Y., Xu Z., Zhang C., Sun B., Guan G. (2020) Frontiers Aging Neurosci., **12**, 217. DOI: 10.3389/fnagi.2020.00217.
- Васильев П.М., Спасов А.А., Яналиева Л.Р., Кочетков А.Н., Ворфоломеева В.В., Клочков В.Г., Анпазова Д.Т. (2019) Биомедицинская химия, **65**(2), 91-98. [Vassiliev P.M., Spasov A.A., Yanaliyeva L.R., Kochetkov A.N., Vorfolomeyeva V.V., Klochkov V.G., Appazova D.T. (2019) Biomeditsinskaya Khimiya, **65**(2), 91-98.
- Васильев П.М., Яналиева Л.Р., Спасов А.А., Кочетков А.Н., Ворфоломеева В.В., Клочков В.Г. (2019) Свидетельство о государственной регистрации базы данных № 2019620160. Ингибиторы рецепторов конечных продуктов гликирования (24.01.2019), Офиц. бюлл. Программы для ЭВМ. БД. ТИМС, **2**, RU 2019620160. [Vassiliev P.M., Yanaliyeva L.R., Spasov A.A., Kochetkov A.N., Vorfolomeyeva V.V., Klochkov V.G. (2019) Ingibitory retseptorov konechnykh produktov glikirovaniya (Advanced Glycation End Product Receptor Inhibitors), Certificate of State Registration of the Database no. 2019620160. (24.01.2019), Offits. Bull. Programmy dlya EVM, baz dannykh i topologiy integral'nykh mikroskhem ETIN (Programs for Computers, Databases and Topologies of Integrated Circuits), **2**, RU 2019620160.].
- Васильев П.М., Яналиева Л.Р., Кочетков А.Н., Ворфоломеева В.В., Клочков В.Г. (2018) Вестник ВолгГМУ, **3**(67), 133-138. [Vassiliev P.M., Yanaliyeva L.R., Kochetkov A.N., Vorfolomeyeva V.V., Klochkov V.G. (2018) Vestnik VolgGMU, **3**(67), 133-138.]
- VMD (2016) <https://www.ks.uiuc.edu/Research/vmd/>
- Laskowski R.A., Swindells M.B. (2011) J. Chem. Inf. Modeling, **51**, 2778-2786.
- Inteligand (2018) <http://www.inteligand.com/>
- Vasilyev P.M., Luzina O.A., Babkov D.A., Appazova D.T., Salakhutdinov N.F., Spasov A.A. (2019) J. Struct. Chem., **60**(11), 1827-1832.
- ChemAxon: Marvin (2017) <https://chemaxon.com/products/marvin>
- МОРАС (2016) <http://openmopac.net/>
- Trott O., Olson A.J. (2010) J. Comput. Chem., **31**, 455-461.
- Нейронные сети. Statistica Neural Networks: Методология и технология современного анализа данных (2008) Горячая линия - Телеком, М. [Neironnye seti. Statistica Neural Networks: Metodologiya i tekhnologiya sovremennogo analiza dannykh (Neural networks. Statistica Neural Networks: Methodology and Technology of Modern Data Analysis) (2008), Goryachaya liniya - Telekom, Moscow.].
- Глотов Н.В., Животовский Л.А., Хованов Н.В., Хромов-Борисов Н.Н. (1982) Биометрия, Изд-во Ленингр. ун-та, Л. [Glotov N.V., Zhivotovskij L.A., Hovanov N.V., Hromov-Borisov N.N. (1982) Biometriya, Izd-vo Leningr. un-ta, L.]
- Vassiliev P.M., Spasov A.A., Babkov D.A., Litvinov R.A., Kochetkov A.N. (2021) Russ. Chem. Bull., **70**(3), 562-566.
- Spasov A.A., Brel A.K., Litvinov R.A., Lisina S.V., Kucheryavenko A.F., Budaeva Yu.N., Salaznikova O.A., Rashchenko A.I., Shamshina D.D., Batrakov V.V., Ivanov A.V. (2018) Russ. J. Bioorganic Chem., **44**(6), 769-777.
- Babkov D.A., Zhukovskaya O.N., Borisov A.V., Babkova V.A., Sokolova E.V., Brigadirova A.A., Litvinov R.A., Kolodina A.A., Morkovnik A.S., Sochnev V.S., Borodkin G.S., Spasov A.A. (2019) Bioorganic Med. Chem. Lett., **29**(17), 2443-2526.
- Spasov A.A., Popov Yu.V., Lobasenko V.S., Korchagina T.K., Vassiliev P.M., Kuznetsova V.A., Brigadirova A.A., Rashchenko A.I., Babkov D.A., Kochetkov A.N., Kovaleva A.I., Efremova O.S. (2017) Russ. J. Bioorganic Chem., **43**(2), 163-169.

Поступила в редакцию: 15. 04. 2021.

После доработки: 28. 04. 2021.

Принята к печати: 10. 05. 2021.

**THE CONSENSUS ENSEMBLE NEURAL NETWORK MULTITARGET MODEL
OF RAGE INHIBITORY ACTIVITY OF CHEMICAL COMPOUNDS**

P.M. Vassiliev, A.A. Spasov, A.N. Kochetkov, M.A. Perfilev, A.R. Koroleva*

Volgograd State Medical University,
1 Pavshikh Bortsov sq., Volgograd, 400131 Russia; *e-mail: pvassiliev@mail.ru

RAGE signal transduction via the RAGE-NF- κ B signaling pathway is one of the mechanisms of inflammatory reactions that cause severe complications in diabetes mellitus. RAGE inhibitors are promising pharmacological compounds that require the development of new predictive models. Based on the methodology of artificial neural networks, consensus ensemble neural network multitarget model has been constructed. This model describes the dependence of the level of the RAGE inhibitory activity on the affinity of compounds for 34 target proteins of the RAGE-NF- κ B signal pathway. For this purpose an expanded database of valid three-dimensional models of target proteins of the RAGE-NF- κ B signal chain was created on the basis of a previously created database of three-dimensional models of relevant biotargets. Ensemble molecular docking of known RAGE inhibitors from a verified database into the sites of added models of target proteins was performed, and the minimum docking energies for each compound in relation to each target were determined. An extended training set for neural network modeling was formed. Using seven variants of sampling by the method of artificial multilayer perceptron neural networks, three ensembles of classification decision rules were constructed to predict three level of the RAGE-inhibitory activity based on the calculated affinity of compounds for significant target proteins of the RAGE-NF- κ B signaling pathway. Using a simple consensus of the second level, the predictive ability of the created model was assessed and its high accuracy and statistical significance were shown. The resultant consensus ensemble neural network multitarget model has been used for virtual screening of new derivatives of different chemical classes. The most promising substances have been synthesized and sent for experimental studies.

Key words: RAGE inhibitors; multitarget affinity; molecular docking; artificial neural networks; consensus ensemble model; virtual screening

Funding. This work was supported by the Russian Foundation for Basic Research (project no. 18-015-00499).

Received: 15.04.2021, revised: 28.04.2021, accepted: 10.05.2021.