

*Ye.B. Shustov*

## FROM RESEARCH *IN VIVO* AND *IN VITRO* TO METHODOLOGY *IN SILICO*

**Yevgeny Shustov** – Deputy Director for Science, Scientific Centre for Biomedical Technologies of Federal Medical-Biological Agency, corresponding member of Russian Academy of Natural Sciences, Doctor of Medicine, professor, recipient of the State Prize of the Russian Federation, St. Petersburg; **e-mail: shustov-msk@mail.ru.**

*We analyze methodological basics of computing experiments in silico in modern molecular pharmacology. We demonstrate that the research of probable intermolecular interaction of chemical compounds and molecular targets of medicines enables us to reveal new promising drugs as well as forecast probable biological activity of new compounds.*

**Keywords:** *molecular design; docking of molecule; computing experiments in pharmacology; molecular pharmacology; adenosine receptors; serine protease.*

*Е.Б. Шустов*

## ОТ ИССЛЕДОВАНИЙ *IN VIVO* И *IN VITRO* К МЕТОДОЛОГИИ *IN SILICO*

**Евгений Борисович Шустов** – заместитель директора по научной работе ФГБУН «Научный центр биомедицинских технологий» Федерального медико-биологического агентства, член-корреспондент РАЕН, доктор медицинских наук профессор, лауреат Государственной премии РФ по науке и технике, г. Санкт-Петербург; **e-mail: shustov-msk@mail.ru.**

*Анализируются методологические основы вычислительных экспериментов in silico в современной молекулярной фармакологии. Показано, что исследования вероятного межмолекулярного взаимодействия химических соединений и молекулярных мишеней лекарственных средств позволяют как выявлять новые перспективные лекарства, так и прогнозировать возможную биологическую активность новых соединений.*

**Ключевые слова:** *молекулярный дизайн; докинг молекул; вычислительные эксперименты в фармакологии; молекулярная фармакология; аденозиновые рецепторы; сериновые протеазы.*

В современной фармакологии задачи синтеза новых лекарственных веществ давно перешли из области стихийного синтеза в область направленного синтеза молекул с заданными свойствами. При этом и свойства молекул, и прогнозируемые для них эффекты и механизмы действия являются результатом математического моделирования межмолекулярного взаимодействия.

В настоящее время достижения современной медицинской генетики, геномики и протеомики позволили сделать переход от расшифровки молекулярных механизмов действия лекарственных средств к их конструированию на основе технологий компьютерной визуализации и вычислительных экспериментов по взаимодействию молекул-мишеней и молекул потенциальных лекарственных средств. Молекулярными мишенями при этом

чаще всего выступают такие макромолекулы, как белки (структурные белки, рецепторные белки, ферменты), полинуклеотидные комплексы (фрагменты ДНК, РНК), а также генные ансамбли, кодирующие синтез различных белков.

Для реализации задач конструирования используются специализированные базы данных по структуре макромолекулярных мишеней, с которыми должны взаимодействовать исследуемые соединения. Свойства анализируемых веществ, в первую очередь – молекулярные структуры, фармакоформные группировки и пространственные характеристики распределения зарядов, должны быть формализованы и описаны в структурах и форматах специализированных баз данных. К настоящему времени разработаны программы компьютерной визуализации, которые на основе

технологии вычислительного докинга определяют, возможно ли взаимодействие молекулы-мишени с анализируемым соединением, и если да – то по каким группировкам оно будет происходить.

Докингом называется процесс «подстыковки» лиганда в лиганд-связывающий центр рецептора с целью поиска как наиболее выгодных положений и ориентаций лигандов, так и выявления факторов, изменение которых может привести к улучшению лиганд-рецепторного взаимодействия. Поскольку определение пространственных структур комплексов лиганд–рецептор с помощью различных физико-химических методов для многих лигандов затруднено, задачей докинга является предсказание связывания химических соединений с рецепторами на основании только виртуального эксперимента. Метод молекулярного докинга позволяет предсказывать пространственную структуру комплекса рецептор–лиганд и свободную энергию его образования, исходя из данных о пространственной структуре рецептора, известной с разрешением в несколько ангстрем (например, полученной с помощью рентгеноструктурного анализа) и химической структуре лиганда.

Компьютерная программа выискивает в своей базе данных соединение, для которого известно аналогичное взаимодействие, и рассчитывает вероятностные коэффициенты наличия или отсутствия для исследуемого соединения эффектов, характерных для соединений-аналогов. В результате формируется список возможных эффектов при лиганд-рецепторном взаимодействии молекулы-мишени и анализируемого вещества, на основе которого затем делается вывод о молекулярном механизме его действия.

Подобные вычислительные эксперименты могут отталкиваться как от молекулы-мишени (тогда идет поиск способного к взаимодействию с ней химической структуры), так и от соединения (тогда идет поиск молекулярной мишени, с которой сможет вступить во взаимодействие исследуемое вещество). Значимость такого виртуального эксперимента стала настолько большой, что в классификации типов экспериментальных медико-биологических исследований, кроме исследований *in vitro* (в пробирке) и *in vivo* (в живом организме), появился новый тип – *in silico* (в кремне, то есть – компьютерное моделирование).

Одной из программ, широко используемых для вычислительных экспериментов виртуального межмолекулярного взаимодействия, является программа PASS.

Компьютерная программа PASS [6] позволяет по структурной формуле «лекарственно подобного» органического соединения (молекулярная масса которого находится в интервале 50 – 1250 а.е.м.) оценить вероятный

профиль его биологической активности. Оценка основана на анализе взаимосвязей «структура–активность» для обширной обучающей выборки, включающей в себя субстанции лекарственных препаратов; «кандидаты в препараты», находящиеся на различных стадиях клинических и доклинических исследований; фармакологические вещества и биохимические реагенты («chemical probes»); вещества, для которых имеется информация о специфической токсичности [5].

В программе PASS биологическая активность представлена качественным образом (активно/неактивно). Алгоритм построения моделей взаимосвязи «структура–активность» на основе веществ обучающей выборки и прогнозирования активности для новых (не включенных в обучающую выборку) веществ основан на Байесовских оценках [5]. Средняя точность прогноза, рассчитанная по скользящему контролю с исключением по одному для всех веществ обучающей выборки и всех представленных для этих веществ в обучающей выборке видов биологической активности, составляет около 95%.

В качестве выходной информации пользователь получает список прогнозируемых видов активности с оценками вероятности наличия каждого вида активности  $P_a$  и вероятности отсутствия каждого вида активности  $P_i$ , которые могут принимать значения в диапазоне от нуля до единицы. Вероятности  $P_a$  и  $P_i$  являются также оценками вероятностей ошибок первого и второго рода соответственно. Их можно рассматривать и как меры принадлежности прогнозируемого соединения к нечетким классам «активных» и «неактивных» веществ. Все эти интерпретации вероятностей  $P_a$  и  $P_i$  являются эквивалентными и полезными для анализа результатов прогноза. На их основе можно сконструировать самые разные критерии анализа результатов прогноза, соответствующие решению конкретных практических задач.

Вероятность  $P_a$  отражает, прежде всего, сходство структуры молекул данного вещества со структурами молекул наиболее типичных в соответствующем подмножестве «активных» веществ в обучающей выборке. Поэтому никакой прямой корреляции величин  $P_a$  с количественными характеристиками активности, как правило, нет. Действительно, активное, но имеющее нетипичную для обучающей выборки структуру молекулы вещество может иметь в прогнозе низкое значение  $P_a$ . Другой важный аспект интерпретации результатов прогноза связан с новизной анализируемого соединения. Если ограничиться только видами активности, прогнозируемыми с наиболее высокими значениями  $P_a$ , то отобранные на основе прогноза вещества могут оказаться аналогами известных фармакологических веществ. Например, при  $P_a > 0,7$  шансы

обнаружить активность в эксперименте довольно высоки, но найденные вещества могут оказаться близкими структурными аналогами известных лекарств.

Если выбрать в качестве критерия  $0,5 < Pa < 0,7$ , шансы обнаружения активности в эксперименте будут ниже, но соединения будут в меньшей степени похожи на известные фармакологические вещества. Для  $Pi < Pa < 0,5$  шансы обнаружения активности в эксперименте еще ниже, но если прогноз подтвердится, найденное соединение может оказаться родоначальником нового химического класса для рассматриваемого вида биологической активности (New Chemical Entity).

Так, например, методами информационного анализа была предсказана биологическая активность янтарного эфира диметиламиноэтанола – его яблочнокислой соли Антилактат [7]. В этом исследовании в качестве наиболее вероятных видов активности были представлены (в скобках – вероятность наличия соответствующей активности): агонист гормона роста (93%), стимулятор аденилатциклазы (87,8%), радиопротектор (78,4%), агонист глицинового сайта NMDA-рецептора (76,5%), ингибитор ацетилтрансферазы аминокислот (75,2%), ингибитор НАДФН-оксидазы (74%), ингибитор креатиназы (66,1%), ингибитор 2-оксоглутаратдекарбоксилазы (66,6%), ингибитор лизин-6-трансаминазы (65,4%), агонист ГАМК-С рецептора (65,3%), антигипоксанта (64%), ГАМК-аминотрансферазы ингибитор (60,9%) и др.

Современная версия программы PASS (PASS 2014) прогнозирует 7158 видов биологической активности, включая 480 терапевтических эффектов и 3818 механизмов действия со средней точностью около 95% (скользящий контроль с исключением по одному).

Анализ взаимосвязей между предсказанными программой PASS эффектами и механизмами действия далее выполняются с использованием компьютерной программы Pharma Expert [4], которая содержит данные о более чем 10 тысячах установленных соотношений между фармакотерапевтическими эффектами и механизмами действия. Если для анализируемого соединения одновременно предсказываются как фармакотерапевтический эффект, так и связанные с ним механизмы действия, вероятность обнаружения данного фармакотерапевтического эффекта у этого соединения возрастает.

Для прогноза взаимодействия химических соединений с белками-мишенями человека была создана специальная версия программы PASS – PASS Targets, где в качестве обучающей выборки была использована база данных ChEMBLdb 17 [8]. PASS Targets прогнозирует взаимодействие с 1488 белками человека со средней точностью, рассчитанной по скользящему контролю.

Для целей виртуального фармакологического эксперимента используются следующие международные базы данных:

- *Chemical Structure Lookup Service (CSLS)*. База данных CSLS (URL: <http://cactus.nci.nih.gov/cgi-bin/lookup/>) содержит информацию о 74 млн структурных формул, содержащихся в более чем 100 различных информационных ресурсах (46 млн уникальных структур).

- *PubChem*. База данных Pub Chem (URL: <https://pubchem.ncbi.nlm.nih.gov/>) поддерживается Национальным центром биотехнологической информации (NCBI) США и представляет собой крупнейший публично доступный источник информации о структуре и свойствах химических соединений. Pub Chem содержит в настоящее время данные о 115 млн образцов химических соединений, 46 млн уникальных структур, 200 млн результатов тестирования биологической активности химических соединений. В базе данных Pub Chem возможен поиск не только идентичных структур органических соединений, но и структурных аналогов.

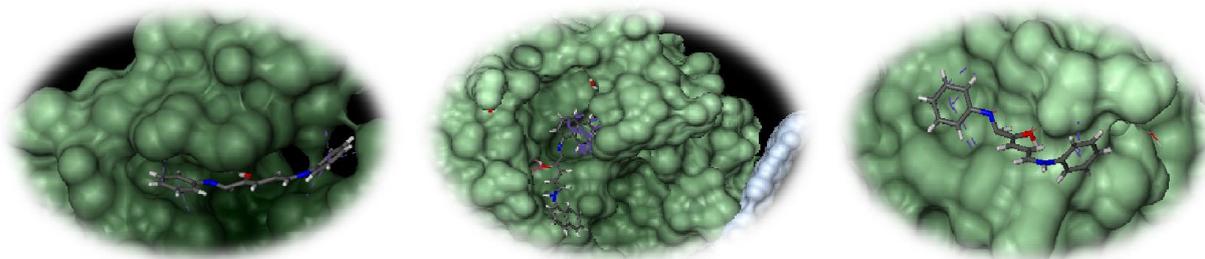
- *Thomson Reuters Integrity*. Базы данных, доступные через портал Integrity (URL: [integrity.thomson-pharma.com](http://integrity.thomson-pharma.com)), содержат информацию о 380000 биологически активных соединениях и 2000 фармакологических мишенях. Имеется возможность поиска по структурному сходству, которое оценивается на основе коэффициента Танимото. Коэффициент Танимото рассчитывается на основе «отпечатков пальцев» (fingerprints) для структуры, используемой в качестве запроса, и для всех структур, содержащихся в базе данных Integrity.

В работе Н.И. Барановой, П.А. Алексеевой, К.С. Чистякова и соавт. [2] продемонстрирована эффективность совместного применения этих двух методов, позволяющая оперативно экспериментально проверять результаты вычислительных экспериментов и корректировать анализируемые выборки из баз данных. Эффективность информационно-молекулярного скрининга при этом существенно повышается [3].

В качестве примера (рис. 1) приведем результаты докинг-анализа в системе свертывания.

При анализе пространственной структуры ключевого фермента образования тромбина – активного фактора X (фактора Хагемана) в ходе вычислительных экспериментов была установлена возможность докинга (пространственного взаимодействия с активным центром фермента) для пентасахаридов (рис. 2).

На уровне регуляции процессов микроциркуляции перспективной мишенью для разработки новых лекарственных средств становится структура аденозиновых рецепторов, в частности, подтипы A3 и A4. Пространственная структура A3-аденозинового рецептора



FXIIa

FXa

Pa (тромбин)

Рис 1. Белок-лигандные комплексы некоторых ферментативно активных факторов свертывания

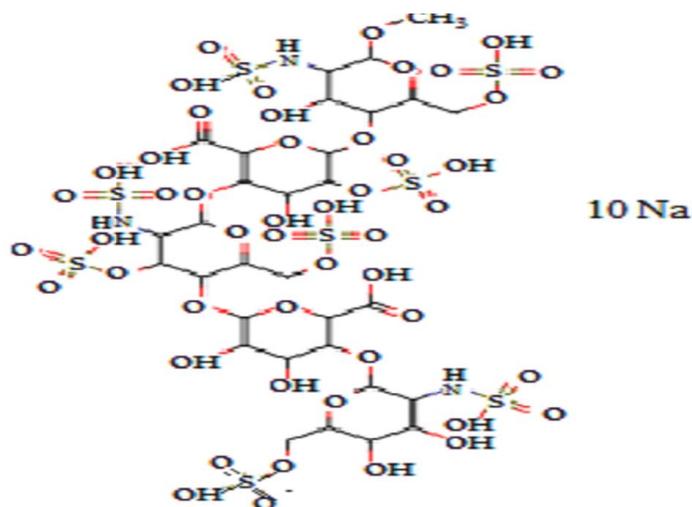


Рис. 2. Структура потенциального ингибитора фактора Хагемана (по результатам вычислительных экспериментов)

известна (включена в общедоступные базы данных), он относится к сопряженным с G-белком рецепторам и состоит из семи трансмембранных  $\alpha$ -спиралей, попарно соединенных тремя внешними и тремя внутриклеточными гидрофильными петлями (рис. 3).

У-образный активный сайт  $A_3$  рецептора состоит из трех гидрофобных доменов ( $P_1$ ,  $P_2$  и  $P_3$ ) (рис. 4). Домен  $P_1$  главным образом состоит из трех неполярных боковых цепей, а именно Leu90, Trp243 и Leu246, причем взаимодействие с Trp243 является наиболее значительным для активации рецептора [4; 5]. Домен  $P_2$  включает различные типы гидро-

фобных аминокислот: Met172, Ile253 и Trp254. Домен  $P_3$  в основном включает остатки Phe182, Ile186 и Val178, взаимодействие с ними для разных лигандов-антагонистов может варьироваться.

Оценка того, с какими аминокислотами взаимодействуют антагонисты  $A_3$ , показала, что в связывании могут принимать участие: TRP243, ILE186, LEU90, LEU246, MET172, TYR254, PHE182, ASN250, SER247, PHE168, GLN167, HIS95, LEU244. Эффективные сайты связывания  $A_3$  антагонистов с рецептором встречаются обычно в верхней области трансмембранных  $\alpha$ -спиралей TM3, TM5, TM6 и TM7 (рис. 5). Все антагонисты демон-

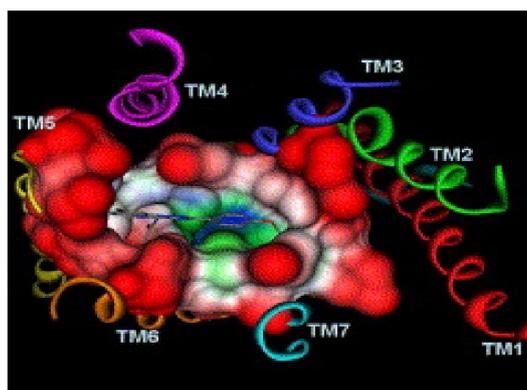


Рис. 3. Модель  $A_3$ -аденозинового рецептора

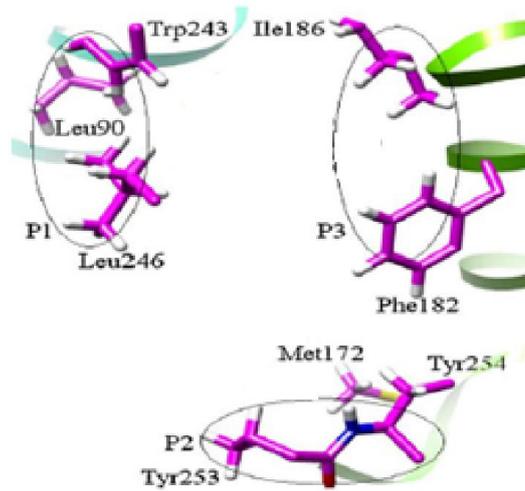


Рис. 4. Y-образный активный сайт  $A_3$  рецептора

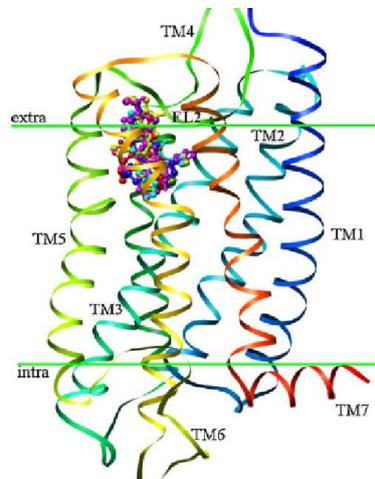


Рис. 5. Эффективный сайт связывания  $A_3$  антагонистов с рецептором

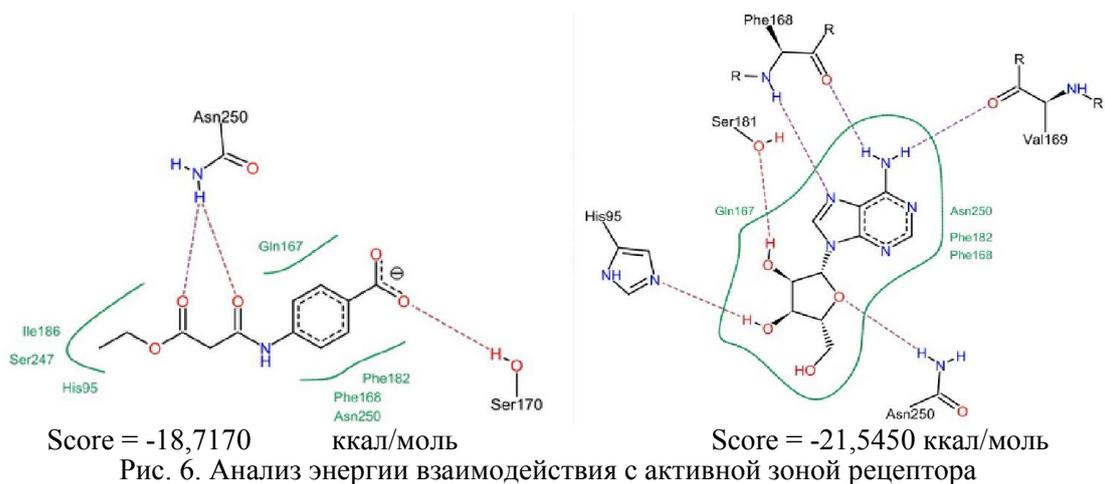


Рис. 6. Анализ энергии взаимодействия с активной зоной рецептора

стрируют некоторые общие взаимодействия с рецептором (рис. 6), а именно – образуют водородные связи с Asn250 и взаимодействуют с остатками Phe168 или Phe182.

Анализ результатов докинга показал, что пространственным характеристикам возможного взаимодействия соответствует такое соединение, как смешанный сложный эфир-амид малоновой кислоты, формула которого

представлена на рис. 7.

4-[(3-этокси-3-оксопропаноил)амино]бензойная кислота в домене P<sub>1</sub> взаимодействует с LEU246 и TRP243, играющими значимую роль для активации рецептора; ILE186 и PHE182 домена P<sub>3</sub> образуют водородные связи с ASN250, SER247, а также взаимодействуют с остатками HIS95 и PHE168 (домен P<sub>2</sub>). Таким образом, лиганд стабилизирован в трех

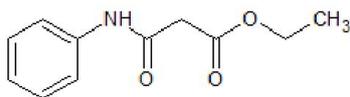


Рис. 7. Структурная формула 4-[(3-этокси-3-оксопропаноил)амино] бензойной кислоты

доменах рецептора P<sub>1</sub>, P<sub>2</sub> и P<sub>3</sub>, демонстрируют общие взаимодействия известных ранее селективных антагонистов с рецептором.

Таким образом, на основании вычислительного моделирования исследования было выявлено новое соединение (4-[(3-этокси-3-оксопропаноил)амино]бензойная кислота), которое может рассматриваться как перспективное средство, корректирующее процессы микроциркуляции, в том числе – в зоне ишемии миокарда [1].

#### ЛИТЕРАТУРА

1. Авенирова Е.Л., Алексеева П.А., Баранова Н.И. [и др.]. Молекулярные аспекты создания лекарственных препаратов: использование методов компьютерного моделирования с целью создания нового противоишемического средства // Биомедицина. 2014. № 1. С. 4–10.

2. Баранова Н.И., Алексеева П.А., Чистяков К.С. [и др.]. Поиск синтетических антикоагулянтов – ингибиторов сериновых протеаз: сочетание *in silico* и *in vitro* методов // Биомедицина. 2014. № 32. С. 17–24.

3. Баскин И.И., Палюлин В.А., Зефирова Н.С. Молекулярное моделирование рецепторов физиологически активных веществ для целей медицинской химии // Успехи химии.

2009. Т. 78. № 6. С. 539–557.

4. Лагунин А.А., Поройков В.В., Филимонов Д.А., Глориозова Т.А. Свидетельство об официальной регистрации программы для ЭВМ Pharma Expert № 2006613590 от 16 октября 2006 г. / Федеральная служба по интеллектуальной собственности, патентам и товарным знакам. М., 2006.

5. Филимонов Д.А., Поройков В.В. Прогноз спектров биологической активности органических соединений // Российский химический журнал. 2006. Т. 50. № 2. С. 66–75.

6. Филимонов Д.А., Поройков В.В., Глориозова Т.А., Лагунин А.А. Свидетельство об официальной регистрации программы для ЭВМ PASS № 2006613275 от 15 сентября 2006 г. / Федеральная служба по интеллектуальной собственности, патентам и товарным знакам. М., 2006.

7. Drachuk O., Semenov S., Oliynik S. Synthesis, structure analysis and biological activity spectrum prediction of Antilactate // Вісник проблем біології і медицини (Україна). 2011. Вип. 2. Т. 1. С. 120–125.

8. Lagunin A., Filimonov D.A., Poroikov V.V. Multi-targeted natural products evaluation based on biological activity prediction with PASS // Cur. Phar. Des. 2010. № 16 (15). P. 1703–1717.