

DOI: 10.24287/1726-1708-2019-18-4-136-152

© 2019 НИИЦ ДГОИ

Поступила 13.05.2019

Принята к печати 19.08.2019

# Современные методы разработки новых лекарственных средств, влияющих на систему гемостаза

А.В. Сулимов<sup>1,2</sup>, Д.К. Кутов<sup>1,2</sup>, А.С. Тащилова<sup>1,2</sup>, И.С. Ильин<sup>1,2</sup>, Н.А. Подоплелова<sup>3,4</sup>, М.А. Пантелеев<sup>3,4,5,6</sup>, И.В. Леденева<sup>7</sup>, Х.С. Шихалиев<sup>7</sup>, В.Б. Сулимов<sup>1,2</sup>

<sup>1</sup> Научно-исследовательский вычислительный центр, ФГБОУ ВО «Московский государственный университет им. М.В. Ломоносова», Москва  
<sup>2</sup> ООО «Димонта», Москва

<sup>3</sup> ФГБУ «Национальный медицинский исследовательский центр детской гематологии, онкологии и иммунологии им. Дмитрия Рогачева» Минздрава России, Москва

<sup>4</sup> ФГБУН «Центр теоретических проблем физико-химической фармакологии» РАН, Москва

<sup>5</sup> ФГБОУ ВО «Московский государственный университет им. М.В. Ломоносова», Москва

<sup>6</sup> ФГБОУ ВО «Московский физико-технический институт (государственный университет)», Москва

<sup>7</sup> ФГБОУ ВО «Воронежский государственный университет», Воронеж

**Контактная информация:**  
 Пантелеев Михаил Александрович,  
 д-р физ.-мат. наук, профессор,  
 зав. лабораторией клеточного  
 гемостаза и тромбоза НИИЦ детской  
 гематологии, онкологии и иммуноло-  
 гии им. Дмитрия Рогачева Минздрава  
 России.  
 Адрес: 117997, Москва, ГСП-7,  
 ул. Саморы Машела, 1  
 E-mail: mapanteleev@yandex.ru

Система свертывания крови играет важную роль в норме и патологии. Несмотря на то что в этой системе уже 30 лет не открывали новых молекул, за это время произошел переворот в представлениях о том, как она работает и какие из ферментов должны стать оптимальными мишенями для терапии. В то же время появились новые методы разработки лекарств, в первую очередь компьютерный докинг, которые идеально подходят для создания ингибиторов именно ферментов свертывания крови. В данном обзоре предпринята попытка соотнести линии развития новых представлений о механизмах свертывания, новых методов поиска лекарств и их сочетания, благодаря которому появляется все больше потенциально интересных молекул, способных в скором времени изменить возможности антикоагулянтной терапии. Кратко рассмотрены методы молекулярного моделирования, прежде всего докинг, которые все более широко применяются на начальной стадии разработки новых лекарственных препаратов, роль докинга на начальной стадии разработки новых ингибиторов, а также детально обсуждается структура активных центров факторов Ха и XIa, определяющая их взаимодействие с ингибиторами.

**Ключевые слова:** свертывание крови, тромбоз, антикоагулянты, ингибиторы факторов свертывания крови, компьютерный дизайн ингибиторов, молекулярный докинг

Сулимов А.В. и соавт. Вопросы гематологии/онкологии и иммунопатологии в педиатрии, 2019; 18 (4): 136–152.  
 DOI: 10.24287/1726-1708-2019-18-4-136-152

© 2019 by NMRC PHOI

Received 13.05.2019

Accepted 19.08.2019

## Modern methods for the development of new drugs that affect the hemostatic system

A.V. Sulimov<sup>1,2</sup>, D.C. Kutov<sup>1,2</sup>, A.S. Tashchilova<sup>1,2</sup>, I.S. Ilin<sup>1,2</sup>, N.A. Podoplelova<sup>3,4</sup>, M.A. Panteleev<sup>3,4,5,6</sup>, I.V. Ledeneva<sup>7</sup>, Kh.S. Shikhaliev<sup>7</sup>, V.B. Sulimov<sup>3,4</sup>

<sup>1</sup> Research Computing Center, Lomonosov Moscow State University, Moscow

<sup>2</sup> Dimonta, Ltd, Moscow

<sup>3</sup> Dmitry Rogachev National Medical Research Center of Pediatric Hematology, Oncology,

Immunology Ministry of Healthcare of Russian Federation, Moscow

<sup>4</sup> Center for Theoretical Problems of Physicochemical Pharmacology, Moscow

<sup>5</sup> Lomonosov Moscow State University, Faculty of Physics, Moscow

<sup>6</sup> Moscow Institute of Physics and Technology (State University), Moscow

<sup>7</sup> Voronezh State University, Faculty of Chemistry, Voronezh

**Correspondence:**  
 Mikhail A. Panteleev, PhD, DSc,  
 Head of the Laboratory of cellular  
 hemostasis and thrombosis of Dmitry  
 Rogachev National Medical Research  
 Center of Pediatric Hematology,  
 Oncology, Immunology Ministry of  
 Healthcare of Russian Federation.  
 Address: Russia, 117997, Moscow,  
 Samory Machelst st., 1  
 Tel.: +7 (495) 287-6570  
 E-mail: mapanteleev@yandex.ru

The blood coagulation system plays an important role in health and disease. It is a complex network of proteolytic reactions that is activated during injuries and controls the formation of a fibrin clot. Although new components and reactions have not been discovered for thirty years, during this time there has been a revolution in understanding of how this system works and what enzymes are the optimal targets for the therapy. At the same time, new methods of drug development, first of all, computer docking, which are ideally suited for the discovery of inhibitors of blood clotting enzymes, have appeared. In this review, an attempt has been made to correlate the lines of development of new ideas about the mechanisms of coagulation, new methods of searching for drugs and their combination, thanks to which now there are more and more potentially interesting molecules that can change the face of the anticoagulant therapy in the near future. In the review, molecular modeling methods, primarily docking, which are increasingly used at the initial stage of developing new drugs, the role of docking at the initial stage of developing new inhibitors are briefly considered and the structure of the active centers of factors Xa and XIa, which determines their interaction with inhibitors, are discussed in detail.

**Key words:** blood coagulation, thrombosis, anticoagulants, blood coagulation factor inhibitors, computer design of inhibitors, molecular docking

Sulimov A.V., et al. Pediatric Hematology/Oncology and Immunopathology, 2019; 18 (4): 136–152.  
 DOI: 10.24287/1726-1708-2019-18-4-136-152

**М**етоды молекулярного моделирования, прежде всего докинг, все шире применяют на начальной стадии разработки новых лекарств, в том числе новых ингибиторов факторов свертываемости. Эти методы, а также строение активных центров факторов Ха и XIa, определяющее характер взаимодействий ингибиторов с этими белками, детально рассмотрены в этой статье.

### Докинг и другие методы компьютерного моделирования лекарственных препаратов

Рациональная разработка лекарств в основном основана на следующей парадигме: молекула лекарства должна избирательно связаться с определенным местом макромолекулы, участвующей в развитии заболевания, тем самым воздействуя на течение болезни. Чаще всего такими макромолекулами-мишенями являются белки, с их активными центрами и должны связываться биологически активные молекулы. В дальнейшем, для краткости, мы будем называть такие биологически активные молекулы ингибиторами, хотя это могут быть и агонисты, антагонисты и модуляторы рецепторов белковой природы, – важно, чтобы связывание этих молекул с белком-мишенью приводило к биологическому эффекту.

Компьютерное молекулярное моделирование начинает играть важную роль на начальной стадии разработки лекарств, и оно основано на информации о структуре белков-мишеней, которая аккумулируется в *Protein Data Bank* (PDB) [1]. Среди методов молекулярного моделирования, применяемых для разработки лекарств, центральную роль играет докинг [2, 3], с помощью которого осуществляется позиционирование лиганда (молекулы, которая может связаться с белком) в белке-мишени и дается оценка энергии связывания лиганда с белком, определяющая биологическую активность лиганда [4, 5]. Существует несколько десятков программ докинга, а обзоры по программам докинга и результатам соревнований по докингу появляются практически ежегодно [6, 7]. Последние обзоры представлены в работах [8, 9], причем в обзоре [9] фокус сделан на описании характеристик программ, влияющих на точность докинга: применяемые силовые поля, методы вычисления энергии, модели учета растворителя, алгоритмы поиска наилучшей позы лиганда, глобальной и локальной оптимизации, учет гибкости лиганда, подвижности атомов в активном центре белка и упрощений, сделанных для ускорения докинга.

Программы докинга сравнительно неплохо справляются с позиционированием лигандов в белках, но их общий недостаток – неудовлетворительная точность оценки свободной энергии связывания. Последняя и определяет в конечном счете надежность предсказания активности лиганда, на основе кото-

рой можно было бы отличать слабые ингибиторы от средних и средние – от сильных: чем выше свободная энергия связывания, тем меньшее количество ингибитора вызовет необходимый биологический эффект. Для надежности таких предсказаний точность оценки энергии связывания должна быть выше 1 ккал/моль (почти в два раза больше тепловой энергии при комнатной температуре – 0,6 ккал/моль) – этого предела пока не достигает ни одна программа докинга. Очевидно, что без хорошей точности позиционирования не может быть достигнута высокая точность оценки энергии связывания, но важно и быстрое действие, чтобы выполнять докинг многих тысяч молекул.

Работа многих программ докинга (но не всех) основана на парадигме докинга, которая предполагает, что лиганд связывается с белком вблизи глобального минимума энергии системы «белок–лиганд». Поскольку единственная экспериментальная информация о расположении лиганда в белке заключена в кристаллической структуре комплекса «белок–лиганд», представленной, например, в PDB, то парадигму докинга можно перефразировать: глобальный минимум энергии системы «белок–лиганд» должен соответствовать позе лиганда в белке вблизи положения этого лиганда, закристаллизованного вместе с белком в соответствующем комплексе, представленном в PDB. Конечно, играют роль и тепловые флуктуации, забрасывающие систему в локальные энергетические минимумы, отстоящие по энергии недалеко (на несколько *kT*) от глобального минимума, – они тоже обеспечивают вклад в энергию связывания.

На поиске глобального минимума энергии построены по крайней мере несколько следующих популярных программ: SOL [10–12], AutoDock [13–17], AutoDock Vina [18], ICM [19–21], DOCK [22–24], GOLD [25, 26] и Glide [27–29]. Все эти и другие программы докинга для описания молекулярного комплекса «белок–лиганд» используют те или иные силовые поля – наборы классических потенциалов, описывающие внутри- и межмолекулярные взаимодействия атомов: в программе ICM используются силовые поля ICMFF [30], MMFF94 [31] и ECEPP [32]; в программе SOL – силовое поле MMFF94 [31]; в программе DOCK – потенциалы ван-дер-ваальсова и кулоновского взаимодействия из силового поля AMBER [33], в программе Glide – силовое поле OPLS-AA [34] и его последние модификации [35]. В программе AutoDock используется наполовину эмпирическое силовое поле [15], параметры которого подгонялись, чтобы наилучшим образом воспроизвести экспериментально измеренные энергии связывания комплексов «белок–лиганд». В программе GOLD использованы различные вклады в оценку внутримолекулярных и межмолекулярных взаимодействий [9], которые не сводятся к какому-либо одному силовому полю. Для

оценки энергии в программе AutoDock Vina с целью ускорения докинга применяется очень упрощенная функция, в которую не включено даже кулоновское взаимодействие. При наличии ионов металлов в активном центре белка задача описания взаимодействия лиганда с белком в рамках классических потенциалов какого-либо силового поля усложняется, так как при этом могут образовываться координационные или хелатные связи, имеющие многоцентровый характер, – описание таких связей может быть адекватным только в рамках квантово-химических методов [36]. Тем не менее в некоторых программах докинга введены специальные потенциалы для описания таких взаимодействий: например, в программе GOLD в рамках скоринг-функции Goldscore [25] или в программе Glide в рамках Glidescore [29]. Для правильного описания водородных связей лиганда с белком также необходимы методы квантовой химии, но в большинстве силовых полей и программ докинга используют различные, чаще эмпирические, потенциалы для описания таких связей, а, например, в силовом поле MMFF94 водородные связи описаны просто модифицированным потенциалом ван-дер-ваальсова взаимодействия.

Учет растворителя важен для адекватного описания взаимодействия лиганда с белком, а значит, и в процессе поиска глобального минимума. Поскольку диэлектрическая проницаемость воды велика (78,5 при комнатной температуре), то учет диэлектрического континуума с такой диэлектрической проницаемостью, окружающего молекулу (белок, лиганд или их комплекс), приводит к значительному экранированию кулоновских взаимодействий атомов. При описании связывания лиганда с белком это дает так называемую энергию десольватации, которая определяется разностью энергии взаимодействия комплекса «белок–лиганд» с растворителем и энергий взаимодействия с растворителем белка и лиганда по отдельности. Во многих программах докинга, например, DOCK и SOL, такая континуальная модель растворителя используется, но чаще в упрощенной форме, как в программе AutoDock, где диэлектрическая проницаемость зависит от расстояния. Неполлярная составляющая энергии десольватации в большинстве случаев учитывается в виде простой линейной зависимости от площади поверхности, доступной растворителю. Растворитель важен не только своим диэлектрическим экранированием электростатических взаимодействий, но и участием отдельных молекул воды в образовании водородных связей между лигандом и белком [37]. Например, в программе Glide можно провести докинг отдельных молекул воды для каждого наиболее энергетически выгодного расположения лиганда в белке, найденного на предыдущем шаге докинга.

Для ускорения докинга иногда используют некоторые априорные соображения о том, как предпочтительнее должен быть расположен лиганд в белке, что может существенно снижать размерность и область пространства поиска глобального минимума энергии. Например, в программе GOLD используются точки предпочтительного сопряжения лиганда с белком (*fitting points*), отражающие возможность образования водородных связей между лигандом и белком, а в программе DOCK создаются точки (*matching points*) предпочтительного расположения тяжелых атомов (не атомов водорода) лиганда в активном центре белка, обеспечивающих сопряжение ван-дер-ваальсовых сфер атомов [9]. Такие трюки, конечно, существенно ускоряют докинг, но сказываются на его точности.

Еще одно важное упрощение, существенно ускоряющее докинг и применяемое в том или другом виде во многих программах, в частности, в AutoDock [13–17], ICM [19–21], DOCK [22–24], Glide [27–29] и SOL [10–12], – использование сетки заранее рассчитанных потенциалов взаимодействия пробных атомов лиганда с белком. Сетка охватывает область докинга, и в ее узлах содержатся потенциалы взаимодействия возможных (пробных) атомов лиганда со всеми атомами белка-мишени. Тогда при глобальной оптимизации энергия лиганда в поле белка вычисляется как сумма энергий сеточных потенциалов по всем атомам лиганда в любом его положении в области докинга. Сеточный потенциал в точке расположения атома лиганда вычисляют путем интерполяции значений этих потенциалов в ближайших к данному атому узлах сетки. При этом вся масса трудоемких вычислений энергий взаимодействия каждого атома лиганда (несколько десятков атомов) со всеми атомами белка (несколько тысяч атомов) переходит на этап, предваряющий поиск глобального минимума. Однако при этом существенно ограничены возможности докинга при учете подвижности атомов белка.

Важную роль при докинге играет алгоритм поиска глобального минимума энергии, поскольку его надо искать на многомерной и топологически сложной энергетической поверхности. Например, даже при докинге в жесткий белок гибкого лиганда с пятью внутренними вращательными степенями свободы (торсионными), в котором длины валентных связей и валентные углы фиксированы (а такое приближение используют во многих программах докинга), глобальный минимум надо искать на поверхности, по крайней мере, с 11 измерениями – 6 степеней свободы при трансляциях и вращениях лиганда как целого и 5 торсионных. Исчерпывающий перебор такого многомерного пространства невозможен; обычно используют какие-либо эвристические методы поиска глобального минимума, среди которых хорошо заре-

комендовал себя генетический алгоритм (например, [4]), применяемый в программах AutoDock, GOLD, Glide и SOL.

Для примера кратко опишем организацию работы и алгоритм докинга, реализованные в программе SOL [4, 10–12], которые эффективно применяли для разработки ингибиторов факторов свертываемости крови – тромбина [38] и фактора Ха [39, 40]. Сетка потенциалов строится с помощью модуля SOLGRID на основе силового поля MMFF94. На вход этого модуля подается молекулярный файл белка, содержащий координаты всех его атомов. Заметим, что в структурах комплексов, хранящихся в PDB, как правило, отсутствуют координаты атомов водорода, перед докингом их надо добавить в структуры белка и лиганда. Сделать это для белка вручную практически невозможно, для этого используют те или иные программы, которые, вообще-то, дают различное расположение атомов водорода в пространстве, что существенно влияет на результаты докинга, как показано в работе [41].

Создаваемая модулем SOLGRID сетка потенциалов включает целый набор сеток различных взаимодействий – ван-дер-ваальсово, кулоновское и потенциалы энергии десольватации для разных типов атомов лиганда в соответствии с типизацией атомов в рамках силового поля MMFF94, причем потенциалы ван-дер-ваальсова взаимодействия несколько смягчаются, то есть искусственно становятся шире на величину в несколько десятых долей ангстрема (по умолчанию это 0,4 Å). Этот прием несколько облегчает докинг и имитирует небольшую подвижность атомов белка. Узлы сетки равномерно заполняют куб (область докинга) – 101 узел по каждому из трех ортогональных направлений, размеры которого можно задавать, но обычно используют куб с размером ребра 22 Å; для большинства белков-мишеней и низкомолекулярных лигандов такой куб с запасом охватывает весь активный центр белка и предоставляет лиганду значительную свободу перемещений и изменений конфигурации при докинге. В результате создается бинарный файл размером 200 мегабайт, содержащий всю сетку потенциалов.

Далее проводится докинг с помощью модуля SOL, на вход которого подаются файл управляющих параметров, бинарный файл сетки потенциалов и молекулярный файл с координатами атомов лиганда. В этом модуле выполняется генетический алгоритм поиска глобального минимума энергии лиганда в поле белка. В данном случае целевая функция глобальной оптимизации – это сумма энергии внутримолекулярных взаимодействий лиганда, вычисляемая в силовом поле MMFF94, и энергии лиганда в поле белка. Тем самым при сравнении различных поз лиганда учитывается энергия его внутренних напряжений:

чем выше энергия внутримолекулярных взаимодействий в лиганде, тем энергетически менее выгодна такая поза лиганда. Абсолютные величины энергии внутренних напряжений лиганда (относительно энергии глобального минимума свободного лиганда) могут быть достаточно велики – до нескольких десятков ккал/моль.

В файле параметров задается большинство параметров, регулирующих работу генетического алгоритма, который имитирует процесс эволюции популяции особей с учетом изменения этих особей путем генетических механизмов (мутаций, кроссинговера), в результате которого выживают сильнейшие особи. Особь – это положение заданной конформации лиганда в белке; более сильной считается особь, имеющая более отрицательную целевую энергетическую функцию: энергия лиганда в поле белка плюс энергия внутренних напряжений лиганда. Генерация начальной популяции особей (например, 30 000 особей) происходит случайно, затем из наиболее «сильных» особей формируется новое поколение путем прямого переноса и кроссинговера, а также мутаций, причем и кроссинговер, и мутации – тоже случайные процессы. При отборе особей в новое поколение предусмотрен нишинг, позволяющий поддерживать разнообразие особей: при отборе данной особи в новое поколение предусмотрен значительный штраф по энергии, если особь, близкая к данной, уже имеется. С другой стороны, предусмотрен механизм сохранения лучших особей, имеющих наиболее отрицательную энергию: несколько таких особей передаются в новое поколение без всяких изменений. Эволюция начальной популяции особей за заданное количество поколений (например, 1000) в конечном счете в последнем поколении приводит к особи с наиболее отрицательным значением энергии лиганда. Поза этого лиганда и есть решение задачи глобальной оптимизации, и в соответствии с парадигмой докинга это и есть наилучшее положение лиганда в белке.

Модули SOLGRID и SOL могут работать и в многопроцессорном режиме, что позволяет существенно ускорить и создание бинарного файла с сеткой потенциалов, и докинг одного лиганда.

Для большей уверенности в том, что найден именно глобальный минимум, описанный выше алгоритм повторяется независимо несколько десятков раз; полученные несколько десятков решений (поз лиганда) разбивают на кластеры по близости расположения друг к другу найденных лучших поз лиганда. Разбиение проводится на основании значений среднеквадратичного отклонения атомов лиганда друг от друга (RMSD) – в один кластер попадают позы лиганда, имеющие между собой  $RMSD < 1 \text{ \AA}$ . Если, например, при 50 независимых запусках алгоритма было получено 50 кластеров, по одной позе лиганда в ка-

ждом, то с большой уверенностью можно сказать, что лучшая по энергии поза (в кластере № 1 в программе SOL) не является глобальным минимумом, так как с большой вероятностью еще один независимый запуск алгоритма может привести к позе лиганда с еще более низкой энергией. С другой стороны, если при всех 50 независимых запусках генетического алгоритма будет найдено примерно одно и то же лучшее положение лиганда, так что все 50 лучших поз лиганда соберутся в один кластер, то уверенность в нахождении глобального минимума энергии будет высокой. Чаще всего реализуются промежуточные случаи, и основной критерий надежности докинга – это высокая населенность кластера № 1 и небольшое число кластеров при заданном числе независимых запусков генетического алгоритма [4]. Исследование докинга в зависимости от параметров генетического алгоритма приведено в работе [42], где показано, что наиболее важные параметры – число поколений и размер популяции. В случае трудностей докинга со стандартными значениями этих параметров (размер популяции – 30 000, число поколений – 500) увеличение числа поколений и особенно размера популяции часто приводит к успеху докинга. Итак, лучшее положение лиганда в белке соответствует наиболее отрицательной энергии лиганда в поле белка, полученной при нескольких независимых запусках генетического алгоритма. Эта поза и используется при оценке скорости – свободной энергии связывания лиганда с белком. При этом в скор, помимо энергии лиганда в поле белка, включается энтропийная составляющая, которая пропорциональна числу торсионных лиганда, – это простейшая оценка потери энтропии при замораживании торсионных степеней свободы лиганда при его связывании с белком. Такая простейшая оценка учета энтропийного фактора при связывании лиганда с белком имеет место во многих программах докинга, например, в программах AutoDock, ICM и GOLD.

В программе ICM использован другой алгоритм глобальной оптимизации – так называемый метод Монте-Карло со смещенной вероятностью ВРМС (*Biased Probability Monte Carlo*) [43]. В этом методе реализуются случайные изменения торсионных лиганда и его вращательных и трансляционных степеней свободы, которые не выводят лиганд из заданной области докинга с последующей локальной оптимизацией энергии и принятием или отбрасыванием получившихся минимумов в соответствии с критерием Метрополиса [44]. Недавно разработанный метод глобальной оптимизации на основе разложения многомерных массивов в тензорные произведения (*Tensor Train* – TT) применен в программе TTDock [45]. Показано, что этот метод глобальной оптимизации может составить серьезную конкуренцию генетическому алгоритму.

В связи с увеличением размерности пространства при росте числа торсионных лиганда программы докинга, работа которых основана на поиске глобального минимума энергии системы «белок–лиганд», имеются значительные трудности при докинге лигандов с числом торсионных более 15. Например, в обзоре [8] на основе сравнительного исследования [46] отмечено, что программы докинга Glide, GOLD и ICM хорошо выполняют позиционирование только для лигандов, имеющих 10 и менее внутренних вращательных степеней свободы. С этой же причиной связаны и трудности учета при докинге подвижности атомов белка-мишени; наиболее распространенный метод решения этой проблемы – докинг гибких лигандов в ансамбль различных конформаций белка – по отдельности для каждого члена этого ансамбля.

Помимо докинга лиганда путем его позиционирования как единой целой гибкой молекулы в активном центре белка-мишени, существует другой метод, реализованный во многих программах докинга, – метод докинга *De Novo*. При этом лиганд разбивают на отдельные жесткие молекулярные группы (фрагменты), для каждого фрагмента выполняют докинг, а затем фрагменты, находящиеся в наиболее энергетически выгодных положениях в активном центре белка, вновь объединяют в исходный лиганд. Это основной алгоритм докинга в программе DOCK [22]; он реализован также в программе докинга SurfFlex-Dock [47].

В основе оценки качества программ докинга и их сравнения между собой лежат два критерия, определяющих точность докинга: критерий правильности позиционирования лиганда в белке-мишени и критерий правильности оценки энергии связывания. Первый из них определяет, насколько точно программа докинга может позиционировать лиганд в белке по отношению к его положению в кристалле, где белок и связанный с ним лиганд-ингибитор закристаллизованы вместе. Эти закристаллизованные положения лигандов в соответствующих белках (иногда их для краткости называют нативными положениями, а лиганды, закристаллизованные с белком, – нативными лигандами) представлены в PDB. Докинг обычно считается успешным, если среднеквадратичное отклонение (RMSD) соответствующих атомов лиганда в его нативном положении и в положении этого же лиганда, определенном программой докинга, менее 2 Å [4, 10].

Правильность оценки энергии связывания (скор или скоринг-функции) лигандов с заданным белком-мишенью показывают кривые обогащения, которые строятся на основе ранжирования большого множества неактивных лигандов, содержащего некоторое количество заведомо известных ингибиторов данного белка. Ранжирование по скору должно выде-

лить известные ингибиторы на фоне большого числа неактивных лигандов – для этого строятся соответствующие кривые обогащения [4, 10], или ROC-кривые [12]. В работе [7] представлена сводная таблица точности позиционирования и надежности выявления ингибиторов на основе оценки энергии связывания для десяти популярных программ: позиционирование лигандов с точностью  $RMSD < 2 \text{ \AA}$  удается в основном для 70–80% случаев (91% – для программы ICM), а величина площади под ROC-кривой:  $AUC \approx 0,60\text{--}0,78$ , что указывает на невысокую надежность выявления ингибиторов на основе сора (оценки энергии связывания «белок–лиганд») в программах докинга. Хотя эти данные были получены в 2012 году и с тех пор некоторые программы докинга улучшались, при практическом использовании любых программ докинга для разработки новых ингибиторов надо учитывать, что не любые лиганды хорошо позиционируются в белках-мишенях, и отбор ингибиторов среди большого количества неактивных соединений, основанный на оценке энергии связывания (скор), не всегда приводит к успеху. Таким образом, начиная работать с новой мишенью, надо прежде всего убедиться, что выбранная программа докинга с хорошей точностью позиционирует нативные лиганды вблизи их закристаллизованного с данным белком-мишенью положения и сравнительно надежно ( $AUC \sim 0,90$ ) выявляет известные ингибиторы среди неактивных соединений. Важно отметить, что в большинстве программ докинга в выражении для энергии применяются подгоночные коэффициенты, которые подбираются так, чтобы получить хорошую точность позиционирования лиганда в белке и хорошую точность оценки энергии связывания «белок–лиганд» на том или ином обучающем наборе комплексов. Проблема заключается в том, что для других лигандов и других белков-мишеней надежность предсказания позы лиганда и энергии связывания может неконтролируемо ухудшаться, и в рамках такого подхода уже трудно предложить какие-либо регулярные шаги по улучшению точности и позиционирования лиганда, а также оценки энергии его связывания с белком.

Быстродействие программ докинга сильно отличается: время докинга одного лиганда на одном вычислительном узле варьирует от нескольких часов для программы докинга SOL до 10 секунд для программы ICM. Конечно, высокое быстродействие – не всегда причина низкой точности докинга, и каждая программа имеет свои достоинства и недостатки, которые по-разному проявляются в тех или иных соревнованиях по докингу, но каждое такое соревнование представляет собой целое исследование, требующее дополнительных ресурсов и терпения [12]. У некоторых программ имеются разные по точности варианты расчета: например, у программы докинга Glide есть

стандартный вариант (SP – *Standard Precision*) и вариант повышенной точности (XP – *Extra Precision*), сильно отличающиеся по быстродействию, – первый предназначен для виртуального скрининга больших баз данных молекул, а второй, существенно более медленный, – для более точных расчетов небольшого числа молекул.

В связи с попытками разобраться в причинах, ограничивающих точность докинга [5], недавно были разработаны несколько суперкомпьютерных параллельных (многопроцессорных) программ докинга – FLM [48–50], SOL-T [51] и SOL-P [52, 53], в которых не используется сетка заранее рассчитанных потенциалов взаимодействия атомов лиганда с белком, и при выполнении докинга, то есть в процессе глобальной оптимизации, энергия комплекса «белок–лиганд» для любой его конфигурации вычисляется без всяких упрощений и ограничений – прямо в рамках заданного силового поля. Эти программы названы программами прямого (или бессеточного) обобщенного докинга, причем слово «обобщенный» отражает тот факт, что цель выполнения этих программ – не просто нахождение глобального минимума энергии, как это делает большинство описанных выше программ докинга, а нахождение всего спектра низкоэнергетических минимумов комплекса «белок–лиганд», включая, конечно, и глобальный минимум. Программы FLM [48–50], SOL-T [51] и SOL-P [52, 53] требуют уже сравнительно больших вычислительных ресурсов (сотен вычислительных ядер), существенно превосходящих вычислительные ресурсы, необходимые для обычных или классических программ докинга типа AutoDock, ICM, SOL и др. Программы FLM и SOL-T осуществляют докинг гибкого лиганда в жесткий белок, а программа SOL-P может проводить докинг гибкого лиганда в белок с выделенными подвижными атомами. Все три программы работают в рамках силового поля MMFF94 без учета растворителя, а программа FLM может вести докинг с учетом растворителя в континуальной модели. Алгоритм программы FLM прост: проводится параллельный массивный поиск уникальных низкоэнергетических минимумов из случайных положений различных конформаций лиганда в области докинга; локальные оптимизации осуществляются при варьировании декартовых координат всех атомов лиганда. В программах докинга SOL-T и SOL-P использован ТТ-алгоритм глобальной оптимизации, основанный на недавно разработанном методе глобальной оптимизации, который опирается на разложение многомерных массивов в тензорные произведения. Константа связывания (или константа диссоциации) лиганда с белком определяется через свободную энергию связывания, свободную энергию Гиббса, которая вычисляется как разность свободной энергии комплекса «белок–лиганд» и свободных

энергий несвязанных компонентов – белка и лиганда по отдельности. Свободную энергию каждой из этих трех молекулярных систем вычисляют через соответствующие конфигурационные интегралы (статистические суммы) [5], которые можно рассчитать методами молекулярной динамики, используя полученные траектории молекулярных систем. Однако конфигурационные интегралы можно вычислить и на основе спектра низкоэнергетических минимумов, найденного с помощью обобщенных программ докинга. В термодинамическом равновесии молекулярная система (белок, лиганд или их комплекс) находится преимущественно в своих низкоэнергетических минимумах. Если эти минимумы разделены достаточно высокими барьерами, то конфигурационный интеграл такой системы сведется к сумме конфигурационных интегралов по отдельным низкоэнергетическим минимумам, которые можно вычислить в том или ином приближении – например, в мульти-гармоническом приближении, когда интегралы вычисляются аналитически в приближении многомерного гармонического осциллятора [48].

Использование программы FLM привело к формулировке идеи квазидокинга, который осуществляется в два шага. На первом шаге с помощью программы FLM определяется достаточно широкий спектр низкоэнергетических минимумов и соответствующих положений лиганда в белке; второй этап – пересчет энергий комплекса для этих положений лиганда с другим силовым полем или даже квантово-химическим методом. С помощью такого подхода было показано следующее: во-первых, выявлено, что позиционирование лиганда гораздо лучше, если учитывать влияние растворителя [48]; во-вторых, такой подход позволяет сравнивать точность докинга при использовании различных методов расчета энергии комплексов «белок–лиганд» [54, 55]; в-третьих, было показано [54, 55], что наилучшая точность позиционирования лигандов при докинге достижима при использовании сравнительно недавно разработанного квантово-химического полуэмпирического метода PM7 [56] с учетом растворителя в модели COSMO [57]. Таким образом, методология квазидокинга фактически открывает возможность проводить приближенный докинг с помощью квантово-химических полуэмпирических методов: на первом шаге квазидокинга для нахождения нескольких тысяч низкоэнергетических минимумов программе FLM, например, необходимо выполнить несколько сотен тысяч локальных оптимизаций, а на втором шаге только для найденных нескольких тысяч положений лиганда в низкоэнергетических минимумах надо выполнить пересчет энергий с помощью квантово-химического полуэмпирического метода, причем без дополнительной локальной оптимизации [54, 55]. Для более точных расче-

тов свободной энергии связывания «белок–лиганд» используют методы молекулярной динамики (МД) [58, 59], но такие расчеты можно проводить только для небольшого количества молекул, для скрининга больших баз молекул альтернативы докингу нет. Молекулярная динамика используется для моделирования процесса связывания лиганда с белком при расчетах абсолютных значений энергии связывания или для описания превращения одних лигандов в другие при расчетах относительных энергий связывания. При этом находятся траектории движения молекулярной системы, и по ним проводится соответствующее усреднение – фактически вычисляются конфигурационные интегралы. При таком подходе на точность расчетов влияют не только те же проблемы, что и при докинге (качество силового поля, учет влияния растворителя), но и множество других приемов, связанных, например, с превращением одних молекул в другие или с искусственным понижением энергетических барьеров для ускорения связывания. Несмотря на то что появляется все больше публикаций об аккуратных расчетах свободной энергии связывания с помощью методов МД, применение их для каждого нового проекта достаточно трудно даже для экспертов, требуются большие вычислительные ресурсы, более аккуратные атомистические модели и более точные программные реализации. Проведение МД-расчетов комплексов «белок–лиганд» с помощью квантово-химических методов, хотя бы полуэмпирических, пока не реализовано в связи с очень большими вычислительными затратами, но это, по-видимому, дело ближайшего будущего.

### **Структура активных центров факторов свертывания $X_a$ и $X_{Ia}$**

Рассмотрим более детально строение активных центров многообещающих мишеней в системе свертывания крови – факторов  $X_a$  и  $X_{Ia}$ . Структура этих центров во многом определяет особенности взаимодействия с ними ингибиторов, специфичность этих взаимодействий и структуру наиболее перспективных и активных ингибиторов. Докинг позволяет позиционировать лиганд-молекулу, рассматриваемую в качестве кандидата в ингибиторы, в активном центре белка-мишени, оценить ее характерные взаимодействия с белком и продумать стратегию модификации такой молекулы для увеличения ее ингибирующей активности. В литературе регулярно появляются публикации (например, [60]), которые показывают, что докинг стал действенным инструментом на начальной стадии разработки новых ингибиторов для заданных белков-мишеней.

**Ингибиторы фактора  $X_a$ .** Существует более десяти химических классов прямых ингибиторов фактора  $X_a$ ; идентифицировано более 20 различных

структурных мотивов как P1, так и P4, встречающихся в составе данных ингибиторов, детально описанных в обзорах [61, 62]. Кроме того, пять соединений с избирательной активностью против фактора Ха используются в клинической практике для лечения патологий, связанных с повышенным тромбообразованием. Примеры наиболее интересных и детально изученных ингибиторов, для которых, по крайней мере, было начато клиническое тестирование, представлены в *таблице*, дана подробная информация об их активности и стадии разработки, на которой находится конкретное соединение.

С химической точки зрения, все существующие на сегодняшний день ингибиторы фактора Ха, исключая антагонисты витамина К, действующие непрямым и неселективным образом на данный фактор свертывания, можно условно разделить на три группы.

*Первая группа:* ингибиторы полисахаридной природы, опосредующие инактивацию фактора Ха с участием антитромбина III (например, фондапаринукс и гепарин). Высокая молекулярная масса этих ингибиторов исключает возможность перорального способа их приема.

*Вторая группа:* основные ингибиторы – соединения с активностью против фактора Ха (например, соединения на основе амидина), которые содержат основные группы, заряжающиеся положительно при

нейтральном pH. Заряженные молекулы этих ингибиторов обладают низкой липофильностью, что сказывается на их биодоступности.

*Третья группа:* слабо основные и неосновные ингибиторы – все остальные соединения с ингибирующей активностью в отношении фактора Ха, а также все пероральные ингибиторы фактора Ха, одобренные к медицинскому использованию.

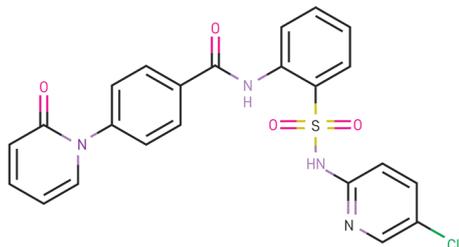
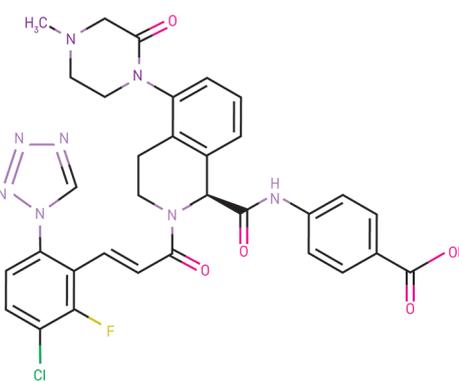
Такая классификация отражает исторические аспекты разработки стратегий ингибирования фактора Ха. Первые агенты, селективно влияющие на активность фактора Ха, были получены за счет ферментативного расщепления гепарина и представляли собой небольшие полисахариды [63]. Механизм действия этих веществ заключается в их способности связываться с антитромбином III и изменять конформацию этого белка таким образом, что аффинитет антитромбина III к фактору Ха многократно увеличивается. Антитромбин, в свою очередь, с усиленным сродством к фактору Ха связывается в окрестности активного центра этого белка, блокируя его протромбиназную функцию. Таким образом, полисахаридные ингибиторы фактора Ха не являются прямыми ингибиторами данной мишени, однако благодаря клинической эффективности их до сих пор используют в медицине, например, при остром тромбозе глубоких вен.

Таблица

**Примеры ингибиторов фактора Ха и XIa** (представленных на рисунках 1 и 2)

Table

Inhibitors of factor Xa and XIa inhibitors were shown in Fig. 1 and Fig. 2, respectively

№	Название Name	Структура Molecule structure	Ингибирующая активность Inhibitory activity	Селективность Inhibitor selectivity	Подход к разработке Development approach
1	Compound 14		$K_i = 5,5 \text{ nM}$	Не исследовалась Not investigated	Поиск SAR на основе экспериментального скрининга Search SAR based on Experimental Screening
2	BMS-962212		$K_i = 0,7 \text{ nM}$	Не влияет на активность тромбина It does not affect thrombin activity	Дизайн виртуальной библиотеки на основе фрагментов. Консенсус докинг для оценки связывания Library design for fragment based screening. Consensus docking evaluation for binding prediction

Разработку обратимых низкомолекулярных ингибиторов сильно ускорило появление в 1994 году первых кристаллических структур фактора Ха [61]. Первые стратегии прямого ингибирования фактора Ха с помощью низкомолекулярных соединений базировались на использовании производных амидина, положительно заряженных при  $pH$  7,4 и взаимодействующих с отрицательно заряженным остатком Asp189 кармана S1. Однако из-за низкой биодоступности исследователям пришлось отказаться от использования таких сильноосновных функциональных групп, как амидин, и сфокусироваться на слабоосновных и неосновных P1-мотивах. Новые нейтральные при физиологическом значении  $pH$  P1-мотивы разрабатывались с учетом их  $\pi$ -взаимодействия с остатком Tyr228 кармана S1. В большинстве прямых ингибиторов фактора Ха это взаимодействие достигается за счет использования галоген- или метокси-монозамещенных ароматических систем в качестве P1-мотива; P4-мотивы большинства ингибиторов также представлены моно- или ди-замещенными аромати-

ческими системами (чаще производными бензола), которые в некоторых случаях связаны с неароматическими гетероциклами. Однако есть примеры ингибиторов, у которых P4-мотив не содержит какой-либо  $\pi$ -системы, например, соединение GW813893. Как отмечено выше, существует большое разнообразие скаффолдов, которые могут быть использованы вместе для объединения P1- и P4-мотивов с целью успешного ингибирования фактора Ха. Важная особенность любого из ранее обнаруженных скаффолдов – их способность придавать молекуле ингибитора V- или L-образную форму, что при связывании способствует такой пространственной ориентации, при которой ингибитор занимает оба кармана активного центра фактора Ха. Кроме того, часто встречаемые скаффолды ингибиторов фактора Ха содержат доноры и/или акцепторы водородной связи для взаимодействия с остатками Gly216 и Gly218.

Рассмотрим структуру активного центра фактора Ха и его характерные свойства, определяющие характер взаимодействий с лигандами и структуру

#### Рисунок 1

Активный центр фактора Ха (PDB ID: 3CEN, <https://www.rcsb.org/structure/3CEN>). S1 и S4 обозначают карман S1 и карман S4. Соответствующие структурные мотивы закристаллизованного ингибитора, связывающиеся внутри этих карманов, обозначены P1 и P4. Поверхность белка показана как доступная растворителю, а цвет поверхности соответствует находящимся вблизи нее типам атомов белка: белый цвет соответствует атомам водорода, синий – азота, красный – кислорода, серый – углерода, желтый – атомам серы

#### Figure 1

Factor Xa active center (PDB ID: 3CEN, <https://www.rcsb.org/structure/3CEN>). Letters S1 and S4 denote pocket S1 and pocket S4, respectively. The structural motifs of the crystallized inhibitor binding within these pockets are shown as P1 and P4. The surface of the protein is shown as a surface accessible to the solvent, and the colors of the surface correspond to the types of protein atoms located close to it: white corresponds to hydrogen atoms, blue to nitrogen atoms, red to oxygen, gray to carbon, yellow to sulfur

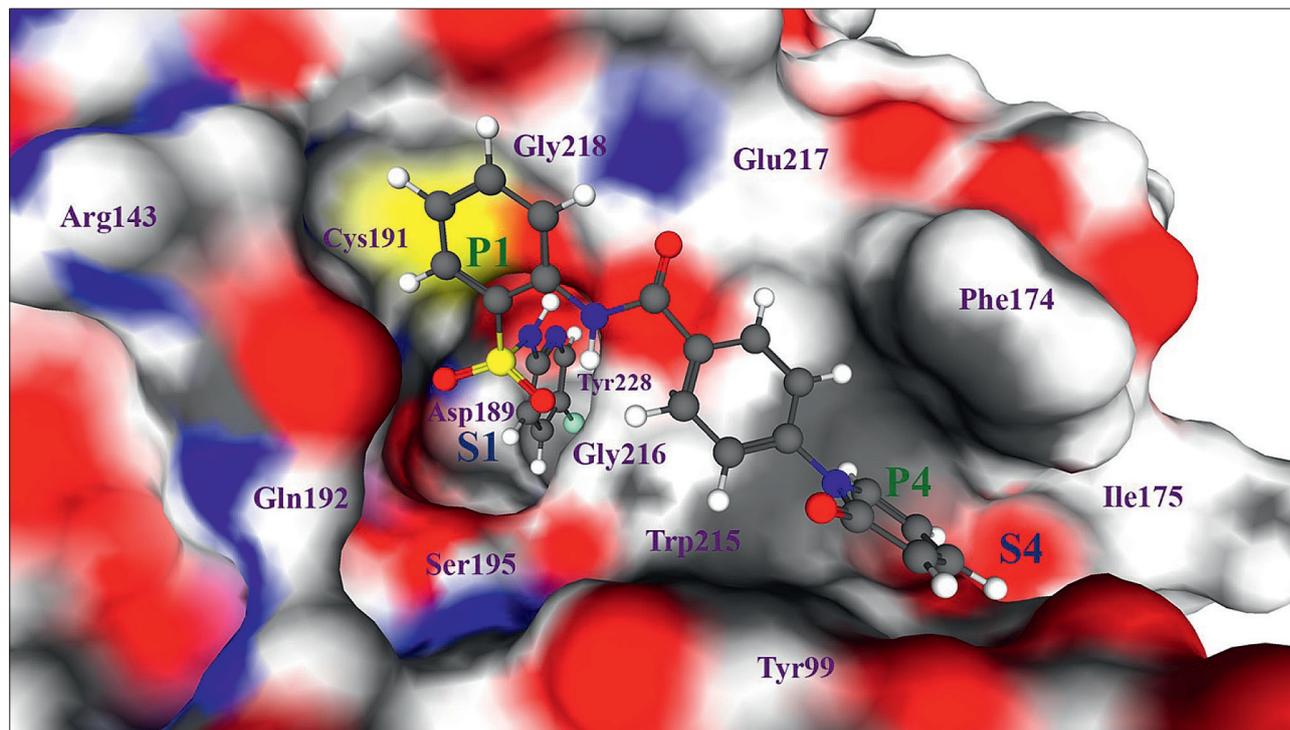


Рисунок сделан с помощью молекулярного редактора MolRed [65, 66]; добавление атомов водорода к белковой молекуле выполнено в программе Aplite [67]; протонирование молекулы лиганда – в программе Avogadro [68].

The picture was made using the molecular editor MolRed [65], [66]. Hydrogen atoms were added to the protein molecule in the Aplite program [67], and the protonation of the ligand molecule was performed in the Avogadro program [68].

ингибиторов этого фактора. Фактор Ха представляет собой витамин К-зависимую сериновую протеазу, состоящую из двух цепей – тяжелой и легкой, связанных между собой дисульфидными связями. Тяжелая цепь состоит из 303 аминокислотных остатков, а легкая – из 139. Каталитическая триада фактора Ха состоит из Ser195, His57, Asp102 и образует структуру наподобие  $\beta$ -бочки, схожую со структурой трипси-на [64]. На *рисунке 1* представлено строение активного центра этого белка, выделены аминокислотные остатки, образующие карманы связывания, и показано «типичное» кристаллическое положение ингибиторов фактора Ха на примере антраниламидного ингибитора, двумерная структура которого показана в *таблице*.

Карман S1 представляет собой глубокую полость, образованную Tyr228, Asp189, Gly216 (см. *рисунок 1*). С точки зрения дизайна ингибиторов, этот карман обладает высоким сродством к положительно заряженным функциональным группам, например, к бензамидину, а также ароматическим группам за счет анион- $\pi$  взаимодействия с участием Asp189 и за счет взаимодействия по типу  $\pi$ -стэкинга за счет наличия гидроксифенильной группы Tyr228 на дне этого кармана. Кроме того, остаток Gly216 и находящийся рядом с карманом S1 остаток Gly218 – ключевые в формировании водородных связей для многих известных ингибиторов фактора Ха [61]. Карман S4 образован аминокислотными остатками, содержащими в качестве боковых групп ароматические кольца: Phe174, Trp215, Tyr99, что создает выраженную гидрофобность этого кармана. Как следствие, в этом кармане хорошо связываются ароматические и незаряженные структурные мотивы ингибиторов, например, пиридиновый или тетрагидропирановый фрагменты. Центральная часть ингибитора, соединяющая P1 и P4 мотивы, обычно также содержит ароматическое кольцо со стороны P4 мотива для взаимодействия с индольным фрагментом Trp215 и акцепторы/доноры водородной связи ближе к P1 мотиву для формирования водородных связей с Gly216 и/или Gly218 [61]. Интересно, что некоторые вычислительные методы (например, CoMFA анализ [69]) выделяют карман S2 внутри активного центра фактора Ха на основе строения других сериновых протеаз и сравнения аминокислотных последовательностей. Однако фактически доступ к этому мнимому карману S2 заблокирован Tyr99, а потому ни субстрат, ни ингибиторы фактора Ха не связываются с этой теоретически найденной областью.

**Ингибиторы фактора XIa.** Предполагается, что ингибиторы активированного фактора XI (FXIa) будут сочетать антикоагулянтный и профибринолитический эффекты с низким риском кровотечения. Низкомолекулярные ингибиторы FXIa не прошли еще

всех необходимых доклинических и клинических испытаний и не получили разрешение для использования при лечении людей, однако разработки ингибиторов FXIa ведутся. Рассмотрим структуру активного центра фактора XIa, его свойства, определяющие характер взаимодействий с лигандами, и структуру ингибиторов этого фактора.

По данным PDB, существует более 70 кристаллических структур FXIa, большинство из них содержит каталитический домен или легкую цепь. Лиганды каталитического домена – это пептиды, ковалентные и нековалентные соединения и др. Структура FXIa состоит из двух  $\beta$ -листов, двух спиральных областей, нескольких петель и сходна с другими трипсиноподобными сериновыми протеазами [70]. Единственный глубокий карман в активном сайте FXIa – это S1 (см. *рисунок 2*). В основании кармана находится Asp189; сам карман представлен  $\beta$ -цепями и петлями, которые заканчиваются дисульфидным мостиком между Cys191-Cys219 в верхней части кармана. Карман S2 FXIa расположен рядом с His57 и ограничен Tyr58B; карманы S3 и S4 очень малы. Карман S1' – это область, противоположная каталитической триаде Asp102, His57 и Ser195, рядом с другим дисульфидным мостиком, образованным Cys40-Cys58. Карман S2' содержит  $\beta$ -цепь, в которую входят остатки Arg39, His40 и Leu41, а также остатки Ile151 и Tyr143, где часто наблюдаются полярные взаимодействия [70].

В основном кристаллические структуры комплексов трипсиноподобных сериновых протеаз показывают, что лиганды фиксируются в глубоком кармане S1 и распространяются в разные карманы в зависимости от сериновой протеазы (см. *рисунок 2*). Комплексы FXIa содержат лиганды, которые всегда связаны в кармане S1; однако ингибиторы FXIa связываются в разных карманах в зависимости от хемотипа, простираясь от кармана S1 до его ближайших окрестностей, включая карманы S1', S2 и S2'. На *рисунке 2* для примера показано положение прямого обратимого ингибитора BMS-962212 [71] (см. *таблицу*) фактора свертываемости FXIa, находящегося в фазе I клинических испытаний и имеющего в экспериментах *in vitro*  $K_i = 0,7$  нМ. Поскольку в PDB отсутствует структура фактора XIa, закристаллизованная вместе с этим ингибитором, то положение этого лиганда получено с помощью докинга программой SOL в структуру белка FXIa из PDB с идентификационным номером 4CRC, в которую недостающие атомы водорода были добавлены с помощью программы APlite. Положение молекулы BMS-962212 в активном центре охарактеризовано относительно расположения карманов активного сайта. Многие существующие низкомолекулярные ингибиторы FXIa также занимают три кармана – S1, S1', S2', показанные на *рисунке 2*. Карман S1 пред-

ставляет собой четко определенную полость; карманы S1', S2' – открытые поверхности. На *рисунке 2* показаны эти карманы и аминокислотные остатки, которые их образуют.

Лиганды во всех исследованных экспериментальных структурах из PDB заполняют именно эти карманы. Исследование структур из PDB показало, что практически все лиганды состоят из трех фрагментов (например, BMS-962212 [71], *рисунк 2*). Лиганды можно условно разделить на «левый» конец, содержащий преимущественно ароматические коль-

ца с хлорным заместителем, центральный фенолаланиноподобный фрагмент и «правый» конец, также с ароматическими кольцами и атомами кислорода в качестве заместителей. В эксперименте левый конец всегда занимает гидрофобный карман S1. Центральный фенолаланиноподобный фрагмент, присутствующий в большинстве экспериментальных ингибиторов и связанный с остальными фрагментами лиганда пептидными мостиками, заполняет гидрофобный карман S1'; правый конец помещается в карманы S2', образуя водородные связи с белком. В эксперимен-

### Рисунок 2

Положение молекулы ингибитора BMS-962212 [71] в активном центре фактора свертываемости крови XIa; показаны карманы активного центра белка S1, S1' и S2', занятые соответствующими группами P1, P1' и P2' лиганда, а также аминокислотные остатки, составляющие эти карманы. Поверхность белка показана как доступная растворителю; цвет поверхности соответствует находящимся вблизи нее типам атомов белка: белый цвет соответствует атомам водорода, синий – азота, красный – кислорода, серый – углерода, желтый – атомам серы

#### Figure 2

The position of the molecule of the inhibitor BMS-962212 [71] in the active center of coagulation factor XIa. The pockets of the active center of the S1, S1' and S2' protein occupied by the P1, P1' and P2' ligand groups are shown, as well as the amino acid residues that make up these pockets. The surface of the protein is shown as a surface accessible to the solvent, and the colors of the surface show the types of protein atoms located near it: white corresponds to hydrogen atoms, blue to nitrogen, red to oxygen, gray to carbon, yellow to sulfur

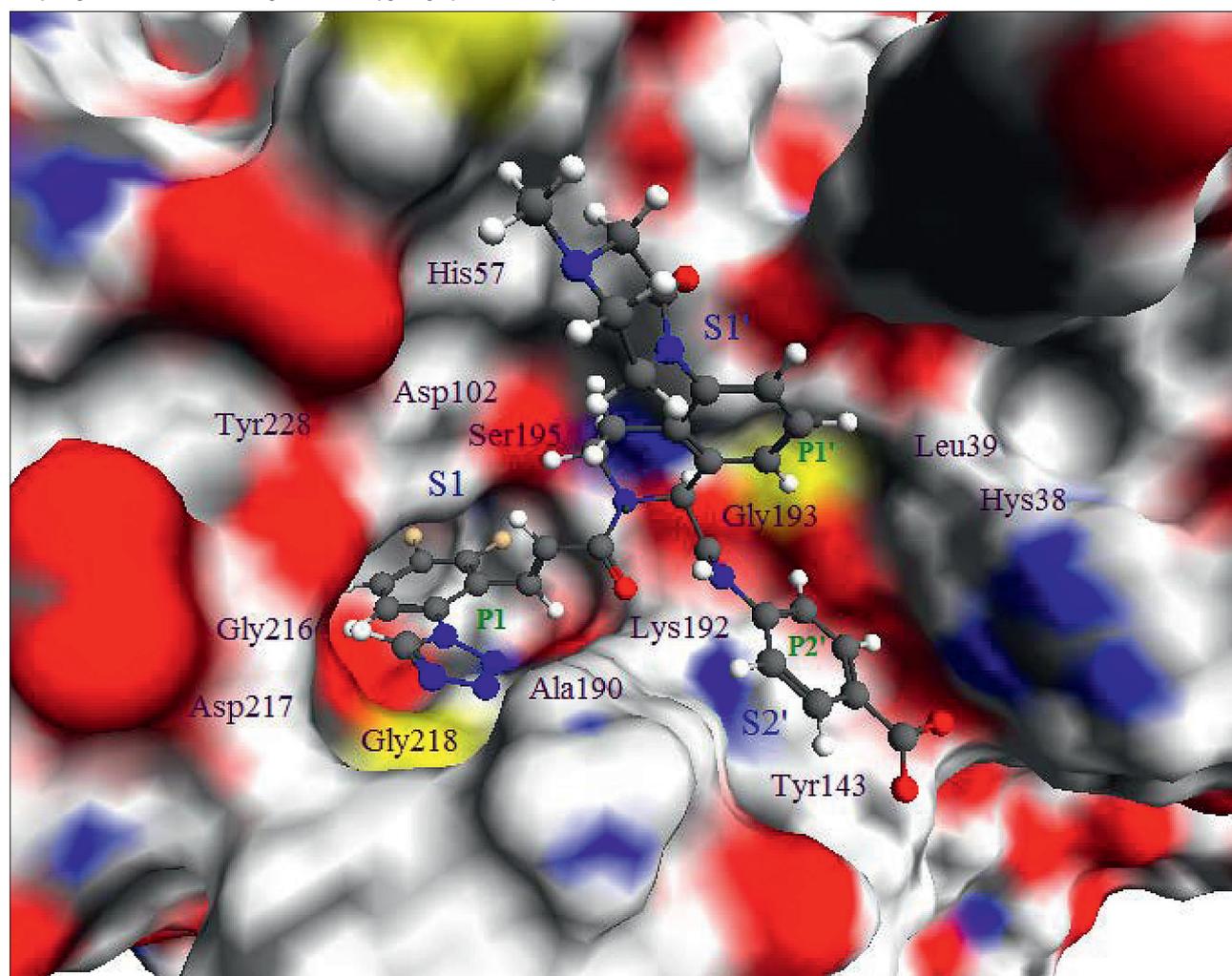


Рисунок сделан с помощью молекулярного редактора MolRed [65, 66]; добавление атомов водорода к белковой молекуле выполнено в программе Aplane [67], протонирование молекулы лиганда – в программе Avogadro [68].

The picture was made using the molecular editor MolRed [65], [66]. Hydrogen atoms were added to the protein molecule in the Aplane program [67], and the protonation of the ligand molecule was performed in the Avogadro program [68].

тальных структурах, помимо нейтральных концевых фрагментов ингибиторов, концевые заряженные фрагменты попадают также в карман S1, – например, с группами COO<sup>-</sup>, NH<sub>4</sub><sup>+</sup>. Соответственно, при процедуре докинга лиганды с растворителем должны быть расположены так же. Геометрия активного центра фактора XIa показана на *рисунке 2*.

Итак, оставляя в стороне ковалентные ингибиторы, необходимость и безопасность использования которых не до конца понятны, можно сказать, что представленные низкомолекулярные соединения находятся на стадии улучшения пероральной биодоступности или доклинических испытаний на животных, за исключением ингибитора FXIa – BMS-962212 [71], который прошел первую фазу клинических исследований.

С учетом рассмотренных выше аспектов строения активных центров факторов Ха и XIa можно выделить основные различия в подходах разработки ингибиторов этих белков. Самое очевидное различие состоит в том, что активный центр фактора XIa имеет большее число карманов, чем протеазный центр фактора Ха, – это, как минимум, предполагает более сложное строение ингибиторов фактора XIa по сравнению с ингибиторами фактора Ха (три структурных мотива вместо двух). Пространственное расположение трех карманов в активном центре фактора XIa требует использования Y-образной формы разрабатываемого ингибитора, в то время как эффективное ингибирование фактора Ха может быть достигнуто при использовании более линейных вариантов низкомолекулярных соединений, способных формировать L-подобные формы. Однако сложность и разветвленность типичного ингибитора фактора XIa гарантирует большую его специфичность в отношении типичных сериновых протеаз, что в целом облегчает процесс оптимизации лидерных молекул с целью увеличения селективности в ходе разработки ингибиторов данного белка. Помимо числа карманов, важное отличие между активными центрами фактора Ха и фактора XIa – различие в сродстве к определенным фрагментам ингибиторов. Так, например, в активном центре фактора XIa имеется положительно заряженный остаток Arg37, что позволяет использовать отрицательно заряженные группы в составе P2'-мотива для увеличения сродства фактора XIa к ингибитору. Напротив, поверхность активного центра фактора Ха за пределами кармана S1, с которой может связываться низкомолекулярный лиганд, лишена заряженных остатков и сформирована по большей части кластером гидрофобных аминокислот (Phe174, Trp215, Tyr99), что предполагает использование слабополярных групп и ароматических фрагментов при дизайне ингибиторов фактора Ха. Одним из следствий наличия более протяженного активного центра

является тот факт, что фактор XIa способен образовывать большее число водородных связей с низкомолекулярным лигандом, чем фактор Ха.

### **Применение докинга на начальной стадии разработки ингибиторов**

Как правило, наиболее эффективно докинг применяется на начальной стадии разработки новых ингибиторов для заданных белков-мишеней. Однако прежде чем приступить к докингу и поиску с его помощью новых биологически активных молекул, необходимо провести подготовительную работу. Во-первых, надо найти в PDB структуру необходимого белка-мишени, желательно вместе с закристаллизованным с ним ингибитором, который иногда называют нативным. Прежде всего он нужен для определения места связывания ингибиторов с активным центром белка (области докинга), а также для валидации работоспособности программы докинга на данном белке и получения референсных значений энергий связывания «белок–лиганд». Последнее возможно только в том случае, если для данного ингибитора экспериментально измерена либо константа связывания (диссоциации), которая напрямую соотносится со свободной энергией связывания (свободная энергия Гиббса), либо в присутствии субстрата измерена константа ингибирования или, по крайней мере, IC<sub>50</sub> – концентрация ингибитора, при которой активность белка подавляется на 50%.

Необходимо, чтобы выбранная структура комплекса имела высокое разрешение и не содержала пропущенных аминокислотных остатков и атомов; если же разрывы в полипептидной цепи белка есть, и/или в отдельных аминокислотных остатках пропущены некоторые атомы, то надо постараться найти в PDB такую структуру, у которой в активном центре мишени этих дефектов нет. Если таких совершенных структур в PDB нет, то применяют методы и программы, позволяющие «заштопать» в белке такие «дыры». Можно вручную расставить недостающие атомы или аминокислотные остатки и хотя бы локально оптимизировать их в том или ином силовом поле. Сделать это можно и методами квантовой химии, по крайней мере, с помощью полуэмпирических методов PM6-D3H4H с поправками на дисперсию и водородные и галогенные связи или метода PM7, в котором эти поправки внесены еще на стадии параметризации. Высоким считается разрешение структур PDB более 2–2,5 Å; (это некий интегральный показатель по всему комплексу); при более низком разрешении (> 3 Å) отдельные атомы структуры уже плохо различимы.

Далее следует рутинный шаг – добавление атомов водорода в белок и лиганд, который с ним закристаллизован, так как в большинстве структур PDB

атомы водорода не различимы – они слишком легкие и быстрые для того, чтобы их мог обнаружить рентгеноструктурный анализ. Разные группы используют для этих целей различные программы, в функции которых входит не только простое добавление атомов водорода к соответствующим атомам белка и отдельно – для лиганда, а также оптимизация их положений, но и определение зарядовых состояний отдельных молекулярных групп и аминокислотных остатков при заданном *pH*: изменяя количество атомов водорода, связанных с тем или другим атомом, можно менять зарядовое состояние соответствующей молекулярной группы [4]. Отметим одну особенность: не только зарядовые состояния отдельных аминокислотных остатков, но и положение атомов водорода в белке сильно зависят от используемых для этого программ [41]. Среди шести использованных для этой цели программ – *Aplite*, *MacroModel*, *Chimera 2* и *MOPAC*, *Chimera 1* и *Autodock* – первые четыре протонируют аминокислотные остатки одинаково практически для всех 16 белков исследованных комплексов, а каждая из двух последних программ создает совершенно другие зарядовые состояния аминокислотных остатков белков. Но даже если зарядовые состояния аминокислотных остатков, сделанные разными программами, идентичны, энергии связывания «белок–лиганд», рассчитанные одинаковым способом, могут сильно отличаться (до 10–20 ккал/моль) из-за разного положения в белке добавленных разными программами атомов водорода. Это неудивительно, поскольку среднеквадратичное отклонение положений атомов водорода, добавленных к заданному белку разными программами, равно 0,7 Å, а положения примерно четверти из них могут различаться на 1,1–1,8 Å [41]. Это, конечно, влияет на результаты докинга как по позиционированию, так и по оценке энергии связывания «белок–лиганд» [41]. Эту неопределенность результатов при моделировании, в том числе при докинге, надо учитывать в реальных разработках.

Итак, модель белка со всеми атомами готова, и ее надо протестировать. Для этого обычно используют лиганды, закристаллизованные в PDB вместе с белком-мишенью, – фактически, это ингибиторы данного белка. В ходе тестирования проводится докинг таких лигандов, определяется лучшее положение лиганда в белке по результатам докинга, сравнивается найденное при докинге положение лиганда с его закристаллизованным положением в белке. Если среднеквадратичное отклонение по всем соответствующим атомам лиганда в этих двух положениях менее 2 Å, то считается, что докинг успешно справился с позиционированием данного нативного лиганда. При этом происходит проверка не только модели белка, но и способности соответствующей программы докинга правильно позиционировать ингибиторы

в активном центре данного белка. Проверку адекватности оценки энергии связывания «белок–лиганд» при докинге проводят другим способом: составляют набор лигандов, среди которых должны быть неактивные для данного белка-мишени соединения – их может быть от нескольких десятков до нескольких тысяч. К ним добавляют некоторое (небольшое) количество соединений – заведомо известных ингибиторов данного белка. Далее проводится докинг этого набора соединений, и они ранжируются по оценке энергии связывания, выдаваемой программой докинга. На основании этого ранжирования строится либо кривая обогащения, либо ROC-кривая, и по величине площади под этой кривой оценивают, насколько эффективно рассчитанная энергия связывания разделяет активные соединения от неактивных [4]. Если площадь под кривой обогащения более 0,9, это означает, что с помощью данной модели белка и данной программы докинга можно проводить виртуальный скрининг (перебор) баз данных готовых соединений и на основе выдаваемой программой докинга оценки энергии связывания выявлять молекулы-ингибиторы среди большой массы неактивных соединений.

На первом этапе реальной разработки новых ингибиторов сначала проводят виртуальный скрининг тех или других баз данных готовых соединений. Таких баз данных много – они есть практически у каждой группы активно работающих химиков-синтетиков, тем более у фармкомпаний; такие базы содержат от нескольких сотен до многих тысяч соединений, которые «лежат на полке», готовые к употреблению. Виртуальный скрининг таких баз заключается в проведении докинга каждой молекулы из базы в активный центр белка-мишени и в оценке энергии связывания этой молекулы с белком. По проведенной оценке энергии связывания «белок–лиганд» выявляют наиболее перспективные в качестве ингибиторов молекулы; их заказывают (их не надо синтезировать, либо их синтез известен!) и проводят с ними первые эксперименты на тест-системах, в которых присутствуют только белок-мишень, его субстрат и добавлен раствор ингибитора. Этот процесс виртуального скрининга различных баз данных готовых соединений продолжается до тех пор, пока не будет накоплено хотя бы несколько соединений, которые в экспериментах показали ингибирующую активность. Желательно, чтобы такие соединения принадлежали к различным химическим классам. Этот этап работы с готовыми соединениями нужен для того, чтобы нащупать характерные молекулярные мотивы лигандов, обеспечивающие их сравнительно высокую энергию связывания с белком-мишенью. Понимание характера взаимодействий найденных таким способом новых, хотя, может быть, и не очень эффективных ингибиторов, дает основу для целенаправленно-

го дизайна новых молекул, иногда построения целых виртуальных библиотек новых соединений, их докинга и выявления наиболее перспективных кандидатов для синтеза. На этом новом этапе нужно тесное сотрудничество с химиками-синтетиками – только они могут синтезировать реально новые соединения, которые затем проходят экспериментальную проверку их ингибирующей активности на тест-системах. Такой итерационный процесс на пути к открытию новых ингибиторов для заданной мишени неизбежен, его эффективность определяется в первую очередь точностью предсказания на основе докинга энергии связывания «белок–лиганд», а также интуицией и умением химиков-синтетиков синтезировать новые химические соединения.

## ЗАКЛЮЧЕНИЕ

В этой работе мы кратко представили общую картину начальной стадии рациональной разработки новых лекарств, воздействующих на систему свертывания крови; показали, как знания, полученные при фундаментальных исследованиях этой системы, представляющей собой сложно организованную сеть ферментативных реакций, приводят к выбору наиболее актуальных факторов свертываемости крови – мишеней для воздействия лекарств, основа которых – ингибиторы этих ферментов. Сегодня самая актуальная мишень – это в первую очередь фактор свертываемости крови  $XIa$  и фактор  $Xa$ . Рентгеноструктурный анализ больших кристаллов этих ферментов, закристаллизованных вместе со своими ингибиторами, дает основу для проведения молекулярного моделирования и выявления с помощью программ докинга из больших баз данных готовых химических соединений молекул – первых кандидатов для экспериментальной проверки их ингибирующей активности заданных белков-мишеней. Первые найденные таким образом ингибиторы с подтвержденной в эксперименте активностью становятся основой для дизайна новых химических соединений, оценки ингибирующей способности последних с помощью методов молекулярного моделирования, дальнейшего синтеза лучших по расчетам кандидатов и экспериментального подтверждения их ингибирующей активности.

Поэтапный процесс дизайна, докинга, синтеза и экспериментальной проверки активности новых молекул по ингибированию белков-мишеней за несколько итераций приводит к успеху – открытию новых соединений ингибиторов, настолько активных, что они могут стать основой новых лекарств. Этот успех в большинстве случаев зависит от слаженной работы медиков и молекулярных биологов, биофизиков и биохимиков, кристаллографов, химиков-синтетиков, молекулярных дизайнеров, вычис-

лительных математиков и программистов, системных администраторов компьютеров и суперкомпьютеров. Для повышения эффективности разработки новых ингибиторов применение молекулярного моделирования может сыграть еще более важную роль, если точность предсказания расчетов свободной энергии связывания «белок–лиганд», их константы связывания будет существенно повышена. Для этого есть необходимая база – уже существующие программы докинга [60], наработки по созданию нового поколения программ докинга [5], в которых отброшено большинство приближений, ограничивающих точность докинга и используемых в классических программах докинга, созданных 10–20 лет назад и ориентированных в основном на однопроцессорные вычисления на обычных рабочих станциях или ноутбуках.

Новое поколение программ докинга ориентировано на суперкомпьютерные многопроцессорные вычисления даже для докинга одного лиганда и использует не только силовые поля на основе классических потенциалов взаимодействия атомов, но и включает в методику расчеты с помощью квантово-химических полуэмпирических методов, что позволяет существенно улучшить, по крайней мере, точность позиционирования лигандов [72, 73]. Наконец, разработаны новые алгоритмы глобальной оптимизации на основе последних достижений тензорного анализа, позволяющие осуществлять докинг гибких лигандов одновременно с учетом подвижности атомов белка [74]. Есть основание оптимистично смотреть в будущее и ожидать в ближайшие годы существенного повышения точности докинга, а значит, и существенного повышения эффективности применения молекулярного моделирования для разработки лекарств.

### ИСТОЧНИК ФИНАНСИРОВАНИЯ

Работа выполнена при финансовой поддержке Российского научного фонда (соглашение № 15-11-00025-П).

### КОНФЛИКТ ИНТЕРЕСОВ

Авторы статьи подтвердили отсутствие конфликта интересов, о котором необходимо сообщить.

### ORCID

**Sulimov A.V.** ORCID: <https://orcid.org/0000-0002-8767-642X>

**Tashchilova A.S.** ORCID: <https://orcid.org/0000-0001-8701-5472>

**Ilin I.S.** ORCID: <https://orcid.org/0000-0002-3599-8718>

**Podoplelova N.A.** ORCID: <https://orcid.org/0000-0002-3612-393X>

**Panteleev M.A.** ORCID: <http://orcid.org/0000-0002-8128-7757>

**Ledeneva I.V.** ORCID: <https://orcid.org/0000-0002-4277-5838>

**Shikhaliev Kh.S.** ORCID: <https://orcid.org/0000-0002-6576-0305>

**Sulimov V.B.** ORCID: <https://orcid.org/0000-0002-7102-6107>

## Литература

- Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., et al. The Protein Data Bank. *Nucleic Acids Res* 2000; 28: 235–42.
- Садовничий В.А., Сулимов В.Б. Суперкомпьютерные технологии в медицине. Суперкомпьютерные технологии в науке, образовании и промышленности. Москва: Издательство Московского университета, 2009; 1: 16–23.
- Gupta M., Sharma R., Kumar A. Docking techniques in pharmacology: How much promising? *Comput Biol Chem* 2018; 76: 210–17.
- Сулимов В.Б., Сулимов А.В. Докинг: молекулярное моделирование для разработки лекарств. – М.: ООО "Интелл", 2017.
- Сулимов А.В., Кутов Д.К., Каткова Е.В., Кондакова О.А., Сулимов В.Б. Поиск подходов к улучшению точности расчетов энергии связывания белок-лиганд с помощью докинга. *Известия Академии наук. Серия химическая*. 2017: 1913–24.
- Chen Y.C. Beware of docking! *Trends Pharmacol Sci*. 2015;36: 78–95.
- Yuriev E., Holien J., Ramsland P.A. Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. *J. Mol. Recognit*. 2015;28: 581–604.
- Pagadala N.S., Syed K., Tuszynski J. Software for molecular docking: a review. *Biophys Rev* 2017; 9: 91–102.
- Sulimov V.B., Kutov D.C., Sulimov A.V. *Advances in Docking*. *Curr Med Chem* 2019; 26: 1–25.
- Романов А.Н., Кондакова О.А., Григорьев Ф.В., Сулимов А.В., Луцкекина С.В., Мартынов Я.Б., et al. Компьютерная разработка лекарств: программа докинга SOL. *Вычислительные методы и программирование*, 2008; 9: 64–84.
- Оферкин И.В., Сулимов А.В., Кондакова О.А., Сулимов В.Б. Реализация поддержки параллельных вычислений в программах докинга SOLGRID и SOL. *Вычислительные методы и программирование*, 2011; 12: 9–23.
- Sulimov A.V., Kutov D.C., Oferkin I.V., Katkova E.V., Sulimov V.B. Application of the docking program SOL for CSAR benchmark. *J Chem Inf Model* 2013; 53: 1946–56.
- Forli S., Huey R., Pique M.E., Sanner M., Goodsell D.S., Olson A.J. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc* 2016; 11: 905–19.
- Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S., et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 2009; 30: 2785–91.
- Huey R., Morris G.M., Olson A.J., Goodsell D.S. A semiempirical free energy force field with charge-based desolvation. *J Comput* 2007; 28 (6): 1145–52.
- Osterberg F., Morris G.M., Sanner M.F., Olson A.J., Goodsell D.S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* 2002; 46: 34–40.
- Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K., et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function *Journal of Computational Chemistry* Volume 19, Issue 14. *J Comput Chem* 1998; 19: 1639–62.
- Trott O., Olson A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010; 31: 455–61.
- Neves M.A., Totrov M., Abagyan R. Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J Comput Aided Mol Des* 2012; 26: 675–86.
- Totrov M., Abagyan R. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins* 1997; 29: 215–20.
- Abagyan R., Totrov M., Kuznetsov D. ICM - a new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* 1994; 15: 488–506.
- Allen W.J., Balias T.E., Mukherjee S., Brozell S.R., Moustakas D.T., Lang P.T., et al. DOCK 6: Impact of new features and current docking performance. *J Comput Chem* 2015; 36: 1132–56.
- Brozell S.R., Mukherjee S., Balias T.E., Roe D.R., Case D.A., Rizzo R.C. Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J Comput Aided Mol Des* 2012; 26: 749–73.
- Moustakas D.T., Scott C.H.P., Kuntz I.D. A practical guide to DOCK5. In: Alvarez J., Shoichet B.K., eds. *Virtual Screen. Drug Discov*. Taylor & Francis Group, LLC 2005: 303–26.
- Liebeschuetz J.W., Cole J.C., Korb O. Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *J Comput Aided Mol Des* 2012; 26: 737–48.
- Cole J.C., Nissink J.W.M., Taylor R. Protein-ligand docking and virtual screening with GOLD. In: Alvarez J., Shoichet B.K. eds. *Virtual Screen. Drug Discov*. Taylor & Francis Group, LLC 2005: 379–415.
- Friesner R.A., Murphy R.B., Repasky M.P., Frye L.L., Greenwood J.R., Halgren T.A., et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* 2006; 49: 6177–96.
- Friesner R.A., Banks J.L., Murphy R.B., Halgren T.A., Klicic J.J., Mainz D.T., et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004; 47: 1739–49.
- Halgren T.A., Murphy R.B., Friesner R.A., Beard H.S., Frye L.L., Pollard W.T., et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J Med Chem* 2004; 47: 1750–9.
- Arnautova Y.A., Abagyan R.A., Totrov M. Development of a new physics-based internal coordinate mechanics force field and its application to protein loop modeling. *Proteins* 2011; 79: 477–98.
- Halgren T.A. Merck molecular force field. *J Comput Chem* 1996; 17: 490–641.
- Arnautova Y.A., Jagielska A., Scheraga H.A. A new force field (ECEPP-05) for peptides, proteins, and organic molecules. *J Phys Chem B* 2006; 110: 5025–44.
- Wang J., Wolf R.M., Caldwell J.W., Kollman P.A., Case D.A. Development and

- testing of a general amber force field Journal of Computational Chemistry Volume 25, Issue 9. J Comput Chem 2004; 25: 1157–74.
34. Jorgensen W.L., Maxwell D.S., Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J Am Chem Soc 1996; 118: 11225–36.
  35. Harder E., Damm W., Maple J., Wu C., Reboul M., Xiang J.Y., et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput 2016; 12: 281–96.
  36. Grigoriev F.V., Golovacheva A.Y., Romanov A.N., Kondakova O.A., Sulimov A.V., Smolov M.A., et al. Stability of HIV-1 integrase–ligand complexes: the role of coordinating bonds. Struct Chem 2012; 23: 185–95.
  37. Nikitina E., Sulimov V., Grigoriev F., Kondakova O., Lushechina S. Mixed implicit/explicit solvation models in quantum mechanical calculations of binding enthalpy for protein–ligand complexes International Journal of Quantum Chemistry Volume 106, Issue 8. Int J Quantum Chem 2006; 106: 1943–63.
  38. Sinauridze E.I., Romanov A.N., Gribkova I.V., Kondakova O.A., Surov S.S., Gorbatenko A.S., et al. New synthetic thrombin inhibitors: molecular design and experimental verification. PLoS One 2011; 6: e19969.
  39. Sulimov V.B., Gribkova I.V., Kochugayeva M.P., Katkova E.V., Sulimov A.V., Kutov D.C., et al. Application of Molecular Modeling to Development of New Factor Xa Inhibitors. Biomed Res Int 2015; 2015: 120802.
  40. Медведева С.М., Потапов А.Ю., Грибкова И.В., Каткова Е.В., Сулимов В.Б., Шихалиев Х.С. Синтез, докинг и антикоагулянтная активность новых ингибиторов фактора ХА в ряду производных пирроло[3, 2, 1-*I*]хинолин-1,2-диона. Химико-фармацевтический журнал 2017; 51: 97–101.
  41. Kutov D.C., Katkova E.V., Kondakova O.A., Sulimov A.V., Sulimov V.B. Influence of the method of hydrogen atoms incorporation into the target protein on the protein–ligand binding energy. Bull. South Ural State Univ. Ser. Math. Model. Program Comput Softw 2017; 10: 94–107.
  42. Каткова Е.В. Исследования влияния параметров генетического алгоритма на эффективность докинга с помощью программы SOL. Вычислительные методы и программирование 2012; 13: 536–50.
  43. Abagyan R., Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol 1994; 235: 983–1002.
  44. Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E. Equation of State Calculations by Fast Computing Machines. J Chem Phys 1953; 21: 1087–92.
  45. Желтков Д.А., Оферкин И.В., Каткова Е.В., Сулимов А.В., Сулимов В.Б., Тыртышников Е.Е. TTDock: метод докинга на основе тензорных поездов. Вычислительные методы и программирование 2013; 14: 279–91.
  46. Perola E., Walters W.P., Charifson P.S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins 2004; 56: 235–49.
  47. Spitzer R., Jain A.N. Surflex-Dock: Docking benchmarks and real-world application. J Comput Aided Mol Des 2012; 26: 687–99.
  48. Oferkin I.V., Katkova E.V., Sulimov A.V., Kutov D.C., Sobolev S.I., Voevodin V.V., et al. Evaluation of Docking Target Functions by the Comprehensive Investigation of Protein–Ligand Energy Minima. Adv Bioinforma 2015; 2015: 126858.
  49. Sulimov A.V., Kutov D.C., Sulimov V.B. Parallel Supercomputer Docking Program of the New Generation: Finding Low Energy Minima Spectrum. In: Voevodin V, Sobolev S eds. 4th Russ. Supercomput. Days, RuSCDays. Moscow, Russia: Springer International Publishing 2018; 965: 314–30.
  50. Kutov D.C., Sulimov A.V., Sulimov V.B. Supercomputer Docking: Investigation of Low Energy Minima of Protein–Ligand Complexes. Supercomput Front Innov 2018; 5: 134–7.
  51. Oferkin I.V., Zheltkov D.A., Tyrtysnikov E.E., Sulimov A.V., Kutov D.C., Sulimov V.B. Evaluation of the docking algorithm based on Tensor Train global optimization. Bull. South Ural State Univ. Ser. Math. Model. Program Comput Softw 2015; 8: 83–99.
  52. Sulimov A.V., Zheltkov D.A., Oferkin I.V., Kutov D.C., Katkova E.V., Tyrtysnikov E.E., et al. Evaluation of the novel algorithm of flexible ligand docking with moveable target–protein atoms. Comput Struct Biotechnol J 2017; 15: 275–85.
  53. Sulimov A.V., Zheltkov D.A., Oferkin I.V., Kutov D.C., Katkova E.V., Tyrtysnikov E.E., et al. Tensor Train Global Optimization: Application to Docking in the Configuration Space with a Large Number of Dimensions. In: Voevodin V, Sobolev S eds. Supercomput. Third Russ. Supercomput. Days, RuSCDays 2017, Moscow, Russ. Sept. 25–26, 2017, Revis. Sel. Pap. Cham: Springer International Publishing 2017: 151–67.
  54. Sulimov A.V., Kutov D.C., Katkova E.V., Sulimov V.B. Combined Docking with Classical Force Field and Quantum Chemical Semiempirical Method PM7. Adv Bioinforma 2017; 2017: 7167691.
  55. Sulimov A.V., Kutov D.C., Katkova E.V., Ilin I.S., Sulimov V.B. New generation of docking programs: Supercomputer validation of force fields and quantum-chemical methods for docking. J Mol Graph Model 2017; 78: 139–47.
  56. Stewart J.J. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. J Mol Model 2013; 19: 1–32.
  57. Klamt A., Schuurmann G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 2 1993; 799–805.
  58. Brown S., Shirts M., Mobley D. Free-energy calculations in structure-based drug design. In: Kenneth M. Merz Charles H. Reynolds DR ed. Drug Des. Struct. Ligand-Based Approaches. Cambridge University Press 2010: 61–86.
  59. Klimovich P.V., Shirts M.R., Mobley D.L. Guidelines for the analysis of free energy calculations. J Comput Aided Mol Des 2015; 29: 397–411.

60. Sulimov V.B., Kutov D.C., Sulimov A.V. *Advances in Docking*. *Curr Med Chem* 2019; 26: 1–25.
61. Patel N.R., Patel D.V., Murumkar P.R., Yadav M.R. Contemporary developments in the discovery of selective factor Xa inhibitors: A review. *Eur J Med Chem* 2016; 121: 671–98.
62. Кабанкин А.С., Синауридзе Е.И., Липец Е.Н., Атауллаханов Ф.И. Компьютерный дизайн низкомолекулярных ингибиторов факторов системы свертывания крови. *Биохимия* 2019; 84: 191–211.
63. Bauer K.A., Hawkins D.W., Peters P.C., Petitou M., Herbert J.-M., van Boeckel C.A.A., et al. Fondaparinux, a synthetic pentasaccharide: the first in a new class of antithrombotic agents – the selective factor Xa inhibitors. *Cardiovasc Drug Rev* 2002; 20: 37–52.
64. Padmanabhan K., Padmanabhan K.P., Tulinsky A., Park C.H., Bode W., Huber R., et al. Structure of Human Des(1–45) Factor Xa at 2.2 Å Resolution. *J Mol Biol* 1993; 232: 947–66.
65. Жабин С.Н., Сулимов В.Б. Визуализация и редактирование молекул: программа MolRed. *Научная визуализация* 2009; 1: 108–14.
66. Жабин С.Н., Сулимов В.Б. Реализация интерактивности в молекулярном редакторе MOLRED. *Научная визуализация* 2010; 2: 59–81.
67. Sulimov A.V., Kutov D.C., Oferkin I.V., Katkova E.V., Sulimov V.B. Application of the docking program SOL for CSAR benchmark. *J Chem Inf Model* 2013; 53: 1946–56.
68. Hanwell M.D., Curtis D.E., Lonie D.C., Vandermeersch T., Zurek E., Hutchison G.R., et al. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform* 2012; 4: 14.
69. Böhm M., Stürzebecher J., Klebe G. No Title. *J Med Chem* 1999; 42: 458–77.
70. Quan M.L., Pinto D.J.P., Smallheer J.M., Ewing W.R., Rossi K.A., Luetzgen J.M., et al. Factor Xla Inhibitors as New Anticoagulants. *J Med Chem* 2018; 61: 7425–47.
71. Pinto D.J.P., Orwat M.J., Smith L.M., Quan M.L., Lam P.Y.S., Rossi K.A., et al. Discovery of a Parenteral Small Molecule Coagulation Factor Xla Inhibitor Clinical Candidate (BMS-962212). *J Med Chem* 2017; 60: 9703–23.
72. Sulimov A.V., Kutov D.C., Katkova E.V., Sulimov V.B. Combined Docking with Classical Force Field and Quantum Chemical Semiempirical Method PM7. *Adv Bioinforma* 2017; 2017: 7167691.
73. Sulimov A.V., Kutov D.C., Katkova E.V., Ilin I.S., Sulimov V.B. New generation of docking programs: Supercomputer validation of force fields and quantum-chemical methods for docking. *J Mol Graph Model* 2017; 78: 139–47.
74. Sulimov A.V., Zheltkov D.A., Oferkin I.V., Kutov D.C., Katkova E.V., Tyrtysnikov E.E., et al. Evaluation of the novel algorithm of flexible ligand docking with moveable target-protein atoms. *Comput Struct Biotechnol J* 2017; 15: 275–85.